Theor Appl Genet (1994) 88:220–230

C. S. Wang · D. Gianola · D. A. Sorensen · J. Jensen
A. Christensen · J. J. Rutledge

Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis

Received: 4 June 1993 / Accepted: 9 September 1993

Abstract A replicated selection experiment aimed at increasing litter size (total number of pigs born per litter) in Danish Landrace pigs was conducted from 1984 to 1991. The experiment included two selection and two control lines. In each generation, 30 and 14 first litters were produced in selection and control lines, respectively, and dams produced two litters. Each replicate, consisting of one selection and one control line, was founded from 60 families chosen randomly from the population at large. Family selection was practiced, and the criterion was the predicted breeding value for litter size computed using a repeatability animal model, and taking into account all available information. The data consisted of 947 records from 523 dams (424 dams had two litters) representing five cycles of selection of increased litter size. Data were analyzed from a Bayesian perspective, based on marginal posterior distributions of genetic parameters of interest. Marginalization was achieved using Gibbs sampling, with a single chain length of 1 205 000. After discarding the first 5 000 iterations, a sample was drawn every ten iterations, so 120 000 samples in total were saved. Densities were estimated and plotted, and summary statistics were computed from the estimated densities. The posterior means (± standard error) of heritability and repeatability were 0.22 ± 0.06 and 0.32 ± 0.05, respectively. These point estimates of genetic parameters were within the range of literature values, although on the high side. The posterior mean (± standard error) of genetic response to selection, defined as the difference between the mean breeding values of the selected lines and that of the base population, was 1.37 ± 0.43 pigs after five cycles of selection. The regression (through the origin) of breeding values in the selected lines on generation was 0.25 ± 0.08 pigs. Several informative priors constructed from information obtained with field data in this population were used to examine their influence on inferences. The priors were influential because of the relatively small scale of the experiment. An analysis excluding data from one of the control lines gave smaller genetic variance and heritability, and a smaller response to selection. However, it appears that selection for litter size is effective, but that the true rate of response is probably smaller than data from this experiment suggest.

Key words Response to selection · Landrace pigs · Litter size · Bayesian analysis · Gibbs sampler

Introduction

It is generally accepted that the economic efficiency of pig production can be enhanced by increasing reproductive efficiency and that litter size is its most important component (Bichard et al. 1983; Smith et al. 1983; Tess et al. 1983). Haley et al. (1988), in a review, reported a mean heritability for number born alive slightly under 0.09, with a mean repeatability of about 0.15. They concluded that for genetic evaluation, a repeatability model gives a rasonable approximation to the more complicated multiple-trait model, where litter sizes in different parities are viewed different traits. This is because genetic correlations between parities do not deviate much from unity.

Such estimates of genetic partameters are needed to assess response to selection for litter size when breeding values are predicted using several sources of information. Expected response can be as large as 0.50 pigs per litter and generation (Avalos and Smith 1987), or about

Communicated by E. J. Eisen

C. S. Wang[1] (✉) · D. Gianola · J. J. Rutledge
Department of Meat and Animal Science, University of Wisconsin-Madison, Madison, WI 53706-1284, USA

D. A. Sorensen · J. Jensen · A. Christensen
National Institute of Animal Science, Research Center Foulum, P.O. Box 39, DK-8830 Tjele, Denmark

---

Present address:
[1] Morrison Hall, Department of Animal Science, Cornell University, Ithaca, NY 14853-4801, USA

5% of the mean, which is high. With the exception of a selection experiment reported by Lamberson et al. (1991), and the hyperprolific shemes described by Legault (1985), Le Roy et al. (1987) and Sorensen and Vernersen (1991), selection experiments aimed at increasing litter size in pigs have not given statistically significant responses. The encouraging predictions based on the assumptions of additive inheritance and infinite population size do not seem to be born out in practice.

Several reasons can be advanced to expalin the apparent disagreement between expected and observed responses to selection for litter size in pigs, but there is little doubt that finite population size plays an important role. Firstly, because litter size is considered to be a trait of the dam, selection must operate between families. Secondly, use of family information accelerates the build up of inbreeding (Robertson 1961), and will also lead to smaller responses to selection than those expected assuming independence of predicted breeding values (Hill 1976). Given the type of family structure in pigs and the low heritability of the trait, these factors can lead to a loss of response, relative to that expected on the basis of infinite population theory, of up to 30% (Toro et al. 1988).

An additional problem associated with selection experiments is that of making exact inferences about selection response. Often, due to financial or space limitations, the scale of the experiment does not permit unambiguous inferences. Hence, a correct assessment of uncertainty about response is important. Classical analysis of selection experiments is based either on least-squares estimation, or on mixed linear model methods using animal models (Kennedy 1990). To implement the latter, variance ratios are needed and, because these are unknown (otherwise, the experiment would not be justified), likelihood-based estimates are often used in place of true parameter values. In either case, it is difficult to derive the exact sampling variance of the predictor of selection response, and one has to resort to appoximations. These approximations typically ignore that the data have been generated under selection, and that parameters have been estimated from the data. Such approximations probably understimate the true variance of selection response. Further, it is unlikely that the distribution of the estimator is normal, in which case a point estimate of varience gives an incomplete description of uncertainty. Variances of estimates of response to selection, approximated and applied in a normal theory framework, give the basis for deciding wheather changes due to selection have, or have not, been significant. The amount of information available in many selection experiments may be smaller than is needed for these approximations to work well. This could be crucial for litter size, as the number of records is proportional to the number of dams rather than to the number of pigs raised. An inappropriate assessment of uncertainty about response to selection may lead to false conclusions.

An alternative way of drawing inferences from selection experiments has been proposed by Sorensen et al. (1994). It is based on a Bayesian model where random samples from the requird posterior distributions are drawn by means of the Gibbs sampler, a numerical integration procedure. This method yields a full description of selection response through its marginal posterior distribution, where account is taken not only of the selection mechanism by which the data are generated, but also of the uncertainty about values of fixed effects and variance components. The purpose of this present paper is to report results of a Bayesian analysis of a five-generation selection experiment for an increased total number of pigs born per litter in the Danish Landrace breed. Where appropriate, some results from a classical analysis are included for comparison. The Bayesian analysis was carried out with four different sets of priors for the variance components, to illustrate how external information can be used to supplement that arising from the experiment, and also to influence conclusions, when the contribution of the experiment to inferences about the parameter is weak.

Material and methods

Experimental lines and data

The selection experiment was conducted at a pig research farm belonging to the National Institute of Animal Science, Foulum, Denmark. The experiment was started in 1984 and ended in 1991. The experimental design (Fig. 1) included two replicates with a 6-month lag between replicates 1 and 2.

The reference population was the Danish Landrace breed, with a histoy of selection for increased daily body weight gain, percent of meat and feed efficiency, but not directly for prolificacy. Each replicate consisted of approximately 60 (58 for one, 63 for the other) gilts which originated from a random sample taken from all certified breeding herds in Denmark. These foundation animals had no known genetic relationships. These gilts were mated to produce 60 litters. From these litters, 60 (44 females and 16 males) first generation animals (Fig. 1) were sampled randomly, one from each litter, and brought to the research farm. From these 60 sampled animals, 30 females and 10 males were randomly assigned to a selection line, and the rest (14 females and 6 males) served as a control line. The lines were closed from this point. Within each replicate, selection and control lines were contemporaneous.

Family selection was practiced using the criterion described below. Each female was allowed to have two litters. At parity one, all litters were standardized to two males and two females immediately after birth. Replacement animals were taken from first parity litters only. For the selection line, 30 replacement (selected) females were chosen from the 15 highest scoring families, two females per family, out of the 30, at each generation; 8–10 males were chosen from the 8–10 highest scoring families, one male per family. These numbers varied somewhat between generations. An effort was made to keep population size constant from generation to generation by using the next best animals if some of the originally selected ones were involuntarily culled. Matings of half or full-sibs were not made. In the control line, replacement animals were chosen at random from the 14 litters in each generation; the line was maintained such that the mean and variance of the inbreeding coefficient were equal to those of the selected line. The experiment was run with discrete (non-overlapping) generations. A total of five cycles of selection were practiced in each replicate. As shown in Fig. 1, generation 2 animals were the progeny obtained after the first cycle of selection.

The selection criterion consisted of breeding values predicted with BLUP, using records on the total number of pigs born per litter at the first and second parities. The model included parity number (1 or 2) and batch (generation × replicate) as fixed effects, and additive ge-

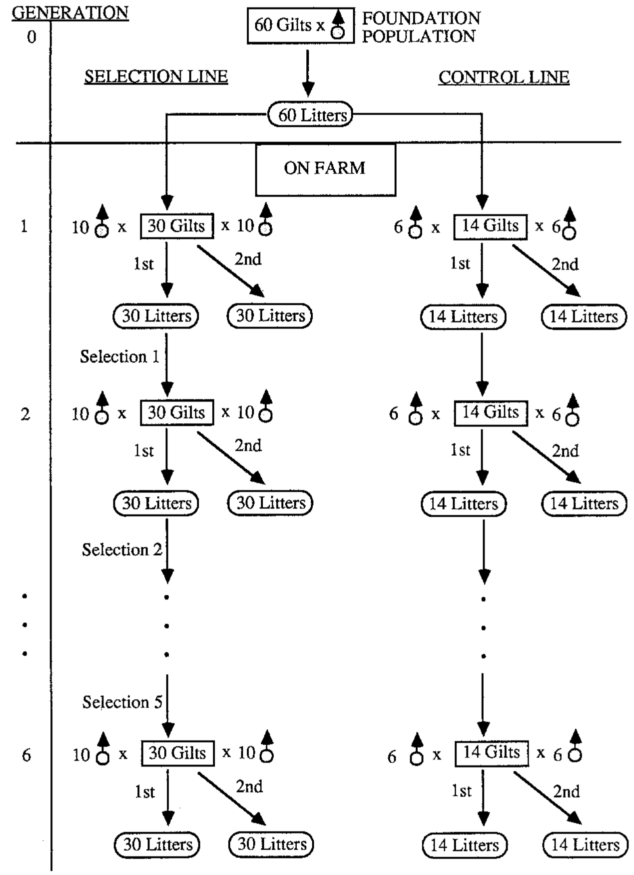


Fig. 1 Design of the selection experiment for one of the two replicates

netic values and permanent environmental effects as random terms. In order to compute breeding values, heritability and repeatability were assumed to be 0.10 and 0.15, respectively.

Within each replicate, selection and control animals were randomly allocated to pens. Sows were mated at the third post-pubertal estrus, at about 7 months of age, to produce first parity litters. Weaning was 5 weeks after farrowing, and sows were mated to produce the second parity litters 5 days after weaning.

Animals were fed a balanced diet based on barley, wheat and soybeans. From about 3 weeks of age and until they reached 25 kg, pigs were fed a starter diet with 21.7% crude protein. From 25 kg until slaughter, males were fed a diet with 19.5% crude protein. Gilts were fed a diet consisting of 18% crude protein from the time of first mating. The amount of feed per day varied according to the physiological stage of the animal.

The resulting data, including the two selection and the two control lines, consisted of 947 records fom 523 dams. Included were 424 dams who had two records (farrowings). The data represented seven generations (0–6), including the foundation population and the progeny resulting from five cycles of selection. The mean inbreeding increase after generation 1 was approximately 2.1% each generation in both selection and control lines, and the mean inbreeding levels at generation 6 reached 10.3% and 11.6% in control and selection lines, respectively. The overall mean ± SD litter size was 10.26 ± 2.67 pigs. Details about the data structure and summary statistics by generation and line are given in Tables A1 and A2 in Appendix 1.

Statistical methods

Data of the four lines were analyzed jointly using the mixed linear model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{c} + \mathbf{e} \tag{1}$$

where **y** is an n × 1 vector of observations on the total number of pigs born per litter; **X**, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are known incidence matrices related location parameters β, **u** and **c**, respectively, to **y**; β is a vector of fixed effects, including a batch (generation × replicate × parity) factor with 24 levels, and two contrasts (line mean difference) between selection and control lines, one for each of two replicates; **u** is a random vector of additive genetic effects; **c** is a random vector of permanent environmental effects, and **e** is a random vector of residuals. The distributional assumptions were:

$$\mathbf{u} \sim \mathrm{N}(0, \mathbf{A}\sigma_u^2), \mathbf{c} \sim \mathrm{N}(0, \mathbf{I}_c\sigma_c^2) \text{ and } \mathbf{e} \sim \mathrm{N}(0, \mathbf{I}_e\sigma_e^2), \tag{2}$$

respectively, where σ_u^2, σ_c^2 and σ_e^2 are variance components (scalars) and **A** is the numerator of Wright's relationship matrix; vectors **u**, **c** and **e** were assumed to be pairwise independent. The rank of **X** was $p = 26$, the order of **u** was $q_u = 838$ (number of animals in the pedigree), and the order of **c** was $q_c = 523$. The rank of the coefficient matrix of the mixed model equations was: $N = p + q_u + q_c = 1387$.

Data were analyzed from a Bayesian perspective using Gibbs sampling, and a standard analysis was carried out for comparison. The focus of the analysis was response to selection using differences in mean additive genetic value between generations, and the regression of additive genetic values on generation as end points. Posterior distributions for variance components and functions thereof were constructed as well.

Standard analysis

The analysis was carried out in three stages. First, REML (Patterson and Thomson 1971; Meyer 1988) estimates of the variance components were obtained. The estimated variance components were then used in Henderson's mixed model equations, to obtain "BLUE" and "BLUP" of all location parameters, including breeding values. Finally, response to selection was estimated by constructing appropriate functions of the predicted breeding values.

Bayesian analysis via Gibbs sampling

The model. Methods for making inferences in a Gaussian mixed linear model using the Gibbs sampler, with applications to animal breeding, are found in Wang et al. (1993, 1994). Sorensen et al. (1994) gave an extension to inferences about genetic response to selection. In all cases, inferences about an unknown parameter of interest are based on its marginal posterior distribution, and the marginalization of the joint posterior distribution is achived by means of the Gibbs sampler.

Below, we summarize pertinent results. Briefly, aside from the distributional assumption described above, a Bayesian analysis needs to assign priors to σ_u^2, σ_c^2 and σ_e^2, and to β. A flat prior was assigned to the latter, i.e.,

$$p(\beta) \propto \text{constant}, \tag{3}$$

and for the variance components, independent scaled inverted chi-square distributions were used:

$$p(\sigma_i^2 \mid v_i, s_i^2) \propto (\sigma_i^2)^{-v_i/2-1} \exp(-\tfrac{1}{2}v_i s_i^2/\sigma_i^2), \quad i = u, c \text{ and } e \tag{4}$$

where, v_i is a "degree of belief" parameter and s_i^2 can be interpreted as a prior value for the appropriate variance.

The joint posterior density of all unknowns (θ, **v** and σ_e^2) was then:

$$p(\theta, \mathbf{v}, \sigma_e^2 \mid \mathbf{y}, \mathbf{s}, v) \propto$$

$$(\sigma_e^2)^{-(n+v_e+2)/2} \exp\left\{-\frac{1}{2\sigma_e^2}[(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}_1\mathbf{u} - \mathbf{Z}_2\mathbf{c})' \cdot (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}_1\mathbf{u} - \mathbf{Z}_2\mathbf{c}) + v_e s_e^2]\right\}$$

$$\times (\sigma_u^2)^{-(q_u+v_u+2)/2} \exp\left\{-\frac{1}{2\sigma_u^2}[\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + v_u s_u^2]\right\}$$

$$\times (\sigma_c^2)^{-(q_c+v_c+2)/2} \exp\left\{-\frac{1}{2\sigma_c^2}[\mathbf{c}'\mathbf{c} + v_c s_c^2]\right\} \tag{5}$$

where $\theta' = (\beta, \mathbf{u}', \mathbf{c}')$, the collection of all location parameters, $\mathbf{v}' = (\sigma_u^2, \sigma_c^2)$, and $s' = (s_u^2, s_c^2, s_e^2)$ and $\nu' = (\nu_u, \nu_c, \nu_e)$, are the collections of all prior variances and prior degrees of belief parameters, respectively.

Inference about $\mathbf{u}$ is based on its marginal posterior distribution with density $p(\mathbf{u}|\mathbf{y})$, which is obtained by integrating successively all parameters other than $\mathbf{u}$ out of the joint posterior density (5). Response to selection was defined as a linear function of $\mathbf{u}$:

$$\mathbf{r} = \mathbf{K}\mathbf{u} \tag{6}$$

where $\mathbf{K}$ is an appropriately defined matrix and $\mathbf{r}$ can be a vector (or scalar) whose elements would be genetic means for each generation, or contrasts between these genetic means or, alternatively, regression coefficients representing linear and quadratic changes of genetic means with respect to time or some measure of selection pressure. Inference about $\mathbf{r}$ is based on its marginal posterior distribution with density $p(\mathbf{r}|\mathbf{y})$; this can be obtained from $p(\mathbf{u}|\mathbf{y})$ by the theory of transformation of random variables.

The Gibbs sampler. Let θ_{-i} be θ without its i-th component, and $\mathbf{v}_{-i}$ be $\mathbf{v}$ without its i-th component. The full conditional posterior distributions for the location parameters are normal (Wang et al. 1994):

$$\theta_i|\mathbf{y}, \theta_{-i}, \mathbf{v}, \sigma_e^2, \mathbf{s}, \nu \sim N(\tilde{\theta}_i, \tilde{v}_i), i = 1, 2, \ldots, N \tag{7}$$

where $\tilde{\theta}_i = (b_i - \sum_{j=1, j\neq i}^{N} w_{ij}\theta_j)/w_{ii}$ and $\tilde{v}_i = \sigma_e^2/w_{ii}$. Further, $w_{ij}(b_i)$ is the ij-th (i-th) element of the coefficient matrix (right-hand side) of Henderson's mixed model equations.

The conditional distribution of σ_e^2 is in the scaled inverted chi-square form, so:

$$\sigma_e^2|\mathbf{y}, \theta, \mathbf{v}, \mathbf{s}, \nu \sim \tilde{\nu}_e \tilde{s}_e^2 \chi_{\tilde{\nu}_e}^{-2} \tag{8}$$

with parameters $\tilde{\nu}_e = n + \nu_e$, and

$$\tilde{s}_e^2 = [(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}_1\mathbf{u} - \mathbf{Z}_2\mathbf{c})'(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}_1\mathbf{u} - \mathbf{Z}_2 c) + \nu_e s_e^2]/\tilde{\nu}_e.$$

Each full conditional density of the other two variance components (σ_i^2) also is in the scaled inverted chi-square form:

$$\sigma_i^2|\mathbf{y}, \theta, \mathbf{v}_{-i}, \sigma_e^2, \mathbf{s}, \nu \sim \tilde{\nu}_i s_i^2 \chi_{\tilde{\nu}_i}^{-2}, \quad i = u, c \tag{9}$$

with parameters $\tilde{\nu}_i = q_i + \nu_i$; $\tilde{s}_u^2 = (\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + \nu_u s_u^2)/\tilde{\nu}_u$ and $\tilde{s}_c^2 = (\mathbf{c}'\mathbf{c} + \nu_c s_c^2)/\tilde{\nu}_c$, respectively.

The above full conditional posterior distributions (7)–(9) are called the Gibbs sampler.

Specification of prior values for ν_i and s_i^2. Prior values for ν_i and s_i^2 ($i = u, c, e$) need to be specified. Four sets of prior values were employed in this study. The first and second sets (I and II) were specified such that $100[\nu_i/(\nu_i + n_i)] = 1$ and 10, respectively, where n_i is the "degree of belief" about the appropriate variance contributed by the data. In our case, $n_u = 523$ (number of animals with records), $n_c = 424$ (number of animals with repeated records) and $n_e = n = 947$ (the total number of records). The values used for s_i^2 were as in set III below. The third set of priors was specified using a method of moments fit based on an independent field data set on litter size of Danish Landrace pigs (Estany and Sorensen, in preparation). Details about the specification of values for set III of priors are in Appendix 2. The fourth set (IV) was $\nu_i = -2$ and $s_i^2 = 0 (i = u, c, e)$, corresponding to flat priors for the three varience components, instead of the informative inverted chi-squares used in (4). The four sets are summarized in Table 1. Note that the concept of percentage of prior degree of belief relative to that of data used in Table 1, and throughout the paper, is for reference only and does not mean that a certain percentage of information is from prior or data.

Running the Gibbs sampler. For each of the four prior sets, the Gibbs sampler was run with a single chain length of 1 205 000. The first 5 000 iterations were discarded, and samples were thereafter saved each ten iterations. In total, 120 000 Gibbs samples were saved. More details about implementation of the Gibbs sampling in Gaussian mixed linear models in an animal breeding context were given in Wang et al. (1993, 1994) and Sorensen et al. (1994). The specification used here was based on our experience by trial and error. For example, convergence checking was done under different specifications of starting values, chain length and the number of samples saved. In the final production run, several independent runs were made. If they produced similar results, convergence was assumed.

Table 1 Specification of priors for the variance components (s_i^2, prior varience; ν_i, prior degree of belief)

Variance	σ_u^2		σ_c^2		σ_e^2	
Prior[a]	s_u^2	ν_u	s_c^2	ν_c	s_e^2	ν_e
I (1%)	0.810	5	0.531	4	6.129	10
II (10%)	0.810	58	0.531	47	6.129	105
III	0.810	124	0.531	74	6.129	18 867
IV (flat)	0	−2	0	−2	0	−2

[a] Degree of belief for Prior I (1%) and II (10%) are specified such that prior degrees of belief are 1% (10%) relative to total degrees of belief (prior plus data). For σ_e^2, for example, the total number of observations was 947, yielding 10 as the prior degree of belief for Prior I10/(10 + 947) = 1%]. See Appendix 2 for specification of Prior III

A FORTRAN program was written to generate the samples, and subroutines form IMSL (IMSL Inc. 1989) were used for random number generation. All computations were in double precision.

Post Gibbs analysis: density estimation and inferences

Marginal densities of interest were estimated using two estimators: (1) a normal kernel density estimator using the samples of a parameter of interest, and (2) the average of conditional posterior densities. Suppose we want to estimate the density of x, $p(x)$; the normal kernel density estimator (Silverman 1986) is:

$$\hat{p}(x) = \frac{1}{mh}\sum_{i}^{m} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - x_i}{h}\right)^2\right] \tag{10}$$

where $\hat{p}(x)$ is the estimated density at x; x_i ($i = 1, 2, \ldots, m$) are the Gibbs samples from $p(x)$ and h is a fixed constant (called window width) given by the user. The estimator based on averaging m conditional densities (Gelfand and Simth 1990), assuming the conditional density, $p(x|y)$, is available, is

$$\hat{p}(x) = \frac{1}{m}\sum_{i=1}^{m} p(x|y_i) \tag{11}$$

where $y_i, i = 1, 2, \ldots, m$, are the samples from the marginal distribution of $p(y)$. Note that, contrary to (10), (11) does not require the samples x_i. Estimation of the density of a function of the original variables is accomplished by applying the theory of transformation of random variables without having to rerun the Gibbs sampling scheme, provided the needed samples are saved. Again, the density of a transformed random variable can be estimated in the above mentioned two ways. Suppose we want to make an inference about $z = f(x, y)$, a function of x and y. Let the transformed samples be $z_i = f(x_i, y_i), i = 1, 2, \ldots, m$. The marginal density of $z, p(z)$, can be estimated by applying (10) using z_i directly. Alternatively, the marginal density estimate of $p(z)$ is computed using (11) via either $p(z|y)$ or $p(z|x)$, where $p(z|y)$ [or $p(z|x)$] is the conditional density of z transformed from $p(x|y)$ [or $p(y|x)$] with a transformation $x \rightarrow z$ [or $y \rightarrow z$].

Inferences about variance components and functions thereof were as in Wang et al. (1993, 1994). Sorensen et al. (1994) applied the Gibbs sampling in the context of making inferences about genetic response using simulated data. Some relevant results of Sorensen et al. (1994) are summarized and extended below.

The conditional posterior distribution of a particular breeding value, from (6), is $u_i|\text{REST} \sim N(\tilde{\theta}_i, \tilde{v}_i)$, with density:

$$p(u_i|\text{REST}) = \frac{1}{\sqrt{2\pi\tilde{v}_i}} \exp\left[-\frac{1}{2\tilde{v}_i}(u_i - \tilde{\theta}_i)^2\right] \tag{12}$$

where REST is a short-hand for all the conditioning parameters. An estimator of the marginal distribution of u_i using (11) is

$$\hat{p}(u_i|\mathbf{y})=\frac{1}{m}\sum_{j=1}^{m}p(u_i|\mathrm{REST})=\frac{1}{m}\sum_{j=1}^{m}\frac{1}{\sqrt{2\pi\tilde{v}_{ij}}}\exp\left[-\frac{1}{2\tilde{v}_{ij}}(u_i-\tilde{\theta}_{ij})^2\right] \quad (13)$$

where $\tilde{\theta}_{ij}$ and $\tilde{v}_{ij}$ are the values of $\tilde{\theta}_i$ and $\tilde{v}_i$, respectively, for Gibbs sample j. As in Sorensen et al. (1994), we define genetic response as the linear function

$$\mathbf{r}=\mathbf{Ku} \quad (14)$$

where $\mathbf{K}$ is a known matrix. After transformation from $\mathbf{u}$ to $\mathbf{r}$, the marginal density of $\mathbf{r}$ is estimated by

$$\hat{p}(\mathbf{r}|\mathbf{y})=\frac{1}{m}\sum_{j=1}^{m}p(\mathbf{r}|\mathrm{REST})_j. \quad (15)$$

When $\mathbf{r}$ is a scalar, the above procedure is easy to apply, as one transforms directly from $p(u_i|\mathrm{REST})$ to $p(\mathbf{r}|\mathrm{REST})$. However, if $\mathbf{r}$ is not a scalar, the dimention of the conditional density used for transformation must be the same as that of $\mathbf{r}$. This can cause difficulties in two ways: first, we may need to sample from a multivarieate distribution so the needed samples and parameters from the appropriate multivariate conditional distribution must be saved; this requires a modification of our basic Gibbs algorithm. Secondly, the transformation from a conditional density must be one to one. This means that the $\mathbf{u}$ vector must be partitioned into two parts: a part related to the conditional density, and the other related to the rest. The corresponding transformation matrix $\mathbf{K}$ must be partitioned consistently; this is not easy to do in general. In this paper, we used the two density estimation methods, (10) and (11), when genetic response $\mathbf{r}$ was defined as a scalar function of $\mathbf{u}$. The normal kernel density estimator (11) was used through transformed samples, when $\mathbf{r}$ was not a scalar.

For convenience, the transformation matrix $\mathbf{K}$ can be defined in terms of another matrix ($\mathbf{Q}$) such that:

$$\mathbf{r}=\mathbf{Ku}=(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{u}. \quad (16)$$

If $\mathbf{r}$ is the mean additive genetic value of a particular generation (t), $\mathbf{Q}$ then is a vector of 1s of length n, the number of animals in that generation, and $\mathbf{u}$ contains elements from that generation only. Applying (16) gives

$$t=\mathbf{Ku}=(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{u}=\frac{1}{n}\sum_{i=1}^{n}u_i=\frac{1}{n}u_j+\frac{1}{n}\sum_{i=1,i\neq j}^{n}u_i. \quad (17)$$

Let the transformation be $u_j\rightarrow t$; since the conditional posterior distribution of u_j is normal, the conditional posterior distribution of t is also normal:

$$t|\mathrm{REST}\sim N\left[\frac{1}{n}\left(\tilde{\theta}_j+\sum_{i=1,i\neq j}^{n}u_i\right),\frac{1}{n^2}\tilde{v}_j\right] \quad (18)$$

where $\tilde{\theta}_j$ and $\tilde{v}_j$ are as in (12). Averaging m conditional densities such as (18) yields an estimate of the marginal density of t, $p(t|\mathbf{y})$. This can be done for each of the seven generations ($i=0,1,\ldots,6$), by appropriately defining $\mathbf{Q}$.

Consider now making interences about total response (TR) to selection, defined as the difference in mean additive genetic values between the last generation and the foundation stock:

$$\mathbf{r}=TR=\mathbf{Km}=t_6-t_0. \quad (19)$$

In this case, $\mathbf{Q}'=[-.5\ .5]$ and $\mathbf{t}'=[t_0\ t_6]$. Transforming from t_6 to TR, the conditional posterior density of TR can be seen to be normal with parameters:

$$TR|\mathrm{REST}\sim N(\widetilde{TR},\tilde{v}_{TR}) \quad (20)$$

where $\widetilde{TR}=\tilde{t}_6-t_0$; this is so because the conditional posterior density of t_6 is normal. The marginal density, $p(TR|\mathbf{y})$, is estimated by averaging m normal densities as in (20).

Another measure of response to selection is the slope δ (scalar) of the regression through the origin of additive genetic values on generations. Here

$$\mathbf{r}=\delta=(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{t}=\left(\sum_{j=0}^{5}t_jq_j+t_6q_6\right)\Big/\sum_{j=0}^{6}q_j^2 \quad (21)$$

where now $\mathbf{Q}'=\{q_j\}=[0,1,\ldots,6]$. Letting the transformation be $t_6\rightarrow\delta$, the conditional posterior disttribution of δ is, again, normal:

$$\delta|\mathrm{REST}\sim N(\tilde{\delta},\tilde{v}_\delta) \quad (22)$$

with parameters $\tilde{\delta}=(\sum_{j=0}^{5}t_jq_j+\tilde{t}_6q_6)/\sum_{j=0}^{6}q_j^2$ and $\tilde{v}_\delta=[q_6/\sum_{j=0}^{6}q_j^2]^2\tilde{v}_{t_6}$. The marginal density of $\delta|\mathbf{y}$ is estimated by averaging m normal densities as in (22).

In (21), $\mathbf{r}$ can be expanded to include an intercept term (δ_0), in which case the corresponding $\mathbf{Q}$ is

$$\mathbf{Q}'=\begin{bmatrix}1&1&1&1&1&1&1\\0&1&2&3&4&5&6\end{bmatrix} \quad (23)$$

resulting in

$$\mathbf{r}=\begin{bmatrix}\delta_0\\\delta_1\end{bmatrix}=(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{t} \quad (24)$$

where δ_1 is now the slope of the linear regression of mean additive genetic values on generations. The vector $\mathbf{r}$ in (24) can be further expanded to include a quadratic term such that:

$$\mathbf{r}=\begin{bmatrix}\delta_0\\\delta_1\\\delta_2\end{bmatrix}=(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{t} \quad (25)$$

where δ_2 is the appropriate regression coefficient, and

$$\mathbf{Q}'=\begin{bmatrix}1&1&1&1&1&1&1\\0&1&2&3&4&5&6\\0&1&4&9&16&25&36\end{bmatrix} \quad (26)$$

Marginal densities of (24) and (25) [e.g., $p(\delta_0|\mathbf{y})$ and $p(\delta_1|\mathbf{y})$ in (24)] can be estimated using the normal kernel density estimator (10) from the transformed samples. It is seen that $\mathbf{Q}$ can be modified as needed for other measures of response to selection. It is also noted that estimating densities of (22), (24) and (25) using (11) requires the storage of all appropriate conditional posterior means and variances, whereas only Gibbs samples are needed using (10).

Window width (h) used in (10) was the range of the effective domain of a parameter of interest, divided by 75. Marginal densities were plotted using pairs, $[x_i,\hat{p}(x_i|\mathbf{y})]$, where x is a parameter of interest, by dividing the effective domain of x into 100 evenly-spaced intervals. Summary statistics were computed from the estimated density of x, by further dividing the effective domain of x into 1 000 equally-spaced intervals, using cubic spline techniques (IMSL Inc. 1989).

Results

Estimated densities of variance components and functions thereof

In the discussion, we emphasize the analysis under flat priors for the variances (such that most of the information was from the experiment), and then examine the consequences of incorporating external information (Priors I–III). Estimated densities of variance components, total variance, variance ratio including heritability and repeatability under flat priors for variance components (Prior IV) are shown in Fig. 2. All marginal distributions were unimodal and nearly symmetrical about their modes, except for $p(\sigma_c^2|\mathbf{y})$ and $p(\sigma_c^2/\sigma_e^2|\mathbf{y})$; this suggests that the experiment contained relatively less information about σ_c^2 than that of two other variance components. Estimated posterior means (standard deviations, SD) of the marginal distributions for additive genetic variance, heritability and repeatability were 1.507 (0.463), 0.222 (0.062) and 0.323 (0.046), respective-

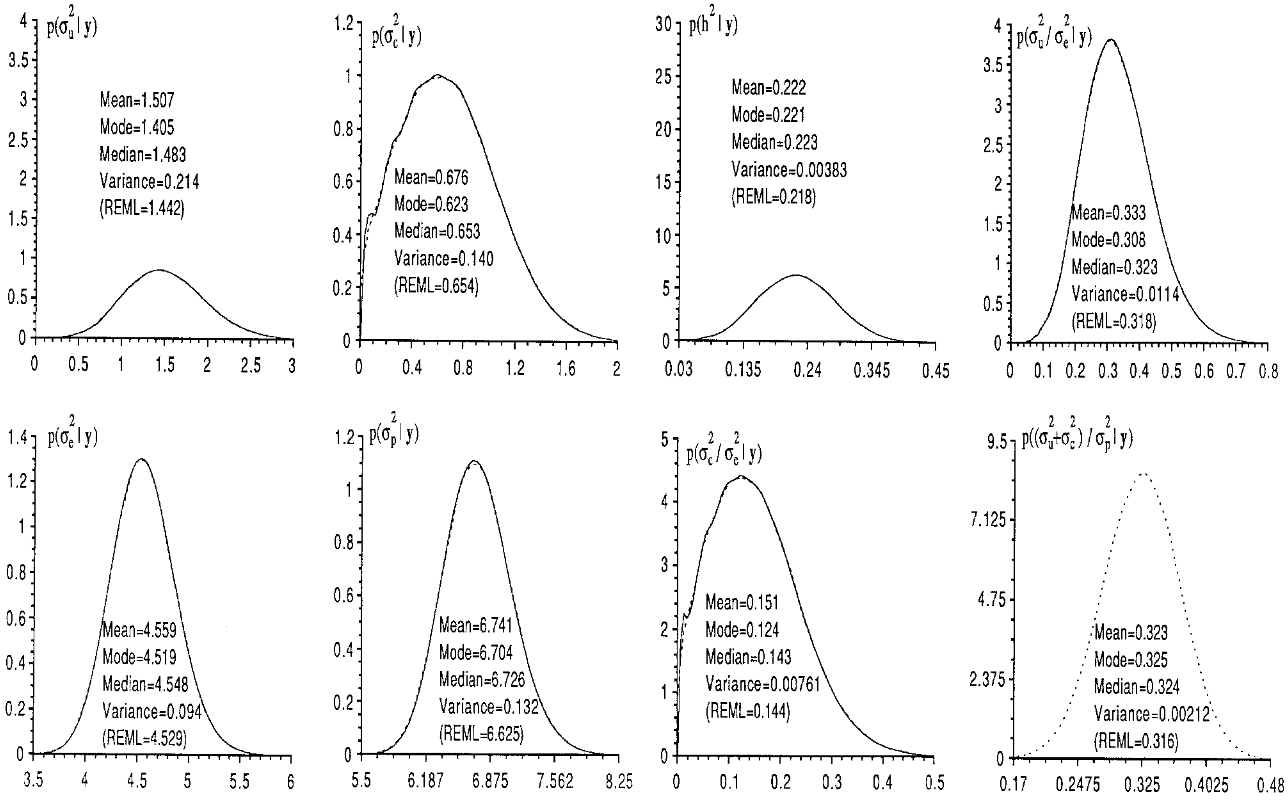


Fig. 2 Estimated marginal posterior distributions of varience components and functions thereof, under Prior IV (flat), along with estimated summary statistics. $\sigma_p^2 = \sigma_u^2 + \sigma_c^2 + \sigma_e^2$

ly. The posterior probabilities that the true parameters lie in the interval [mean − 2 SD, mean + 2 SD] were 0.956, 0.957 and 0.953, respectively. This also indicated that these distributions were nearly symmetric; in these cases, REML estimates were similar to the mean, mode and median. For parameters whose posterior distributions were skewed, REML was closer to the posterior median than to the mean or mode.

It is interesting to contrast the point estimates of dispersion parameters given in Fig. 2 with those resulted from incorporating external information (Table 2). In particular, consider additive genetic variance and heritability. Under prior III, where the hyperparameters derived from Estany and Sorensen (in preparation) were used, the posterior mean (SD) for σ_u^2 was 0.831 (0.010); for heritability, the corresponding estimate was 0.11 (0.01). Clearly, the inferences are much sharper under Prior III than under Prior IV; the external information for this population suggests that it is unlikely that heritability or genetic variance are as large as the limited information from the experiment would lead one to believe. Under Priors I and II, the information contained in Estany and Sorensen (in preparation) is less influential than the one contained in the experimental data. For example, the posterior mean (SD) of heritability was 0.201 (0.05) for Prior I (1% of weight to the external data in terms of degree of belief) and 0.148 (0.02) for Prior II (10% of weight to the external data in terms of degree of belief). This illustrates how the Bayesian analysis can be used to incorporate external evidence in a real animal breeding situation.

Table 2 Estimated summary statistics of marginal posterior distributions of eight selected parameters under three informative priors (I, II and III), as in Table 1. Response parameters are: $t_6 - t_0$ (total response); δ_0 and δ_1 denote intercept and slope, respectively, of the linear regression of mean additive genetic value on generation. REML and ST correspond to results from standard analyses. See text for other parameters

Item	Prior	Mean	Mode	Median	Variance	REML(ST)
σ_u^2	I	1.341	1.256	1.318	0.151	
	II	0.966	0.923	0.951	0.028	1.442
	III	0.831	0.824	0.823	0.010	
σ_c^2	I	0.653	0.477	0.610	0.086	
	II	0.584	0.546	0.571	0.015	0.654
	III	0.516	0.493	0.508	0.006	
σ_e^2	I	4.632	4.600	4.623	0.092	
	II	4.985	4.994	4.978	0.075	4.529
	III	6.087	6.086	6.086	0.004	
σ_p^2	I	6.626	6.608	6.629	0.116	
	II	6.535	6.510	6.528	0.077	6.625
	III	7.433	7.418	7.428	0.019	
h^2	I	0.201	0.200	0.200	0.003	
	II	0.148	0.142	0.146	0.001	0.218
	III	0.112	0.109	0.111	0.000	
$t_6 - t_0$	I	1.315	1.304	1.306	0.163	
	II	1.155	1.144	1.150	0.111	1.381
	III	1.011	1.004	1.010	0.099	
δ_0	I	−0.114	−0.107	−0.112	0.022	
	II	−0.010	−0.098	−0.010	0.016	−0.135
	III	−0.087	−0.085	−0.087	0.014	
δ_1	I	0.262	0.262	0.261	0.005	
	II	0.230	0.226	0.229	0.004	0.280
	III	0.201	0.201	0.201	0.003	

Estimated marginal densities of line mean differences

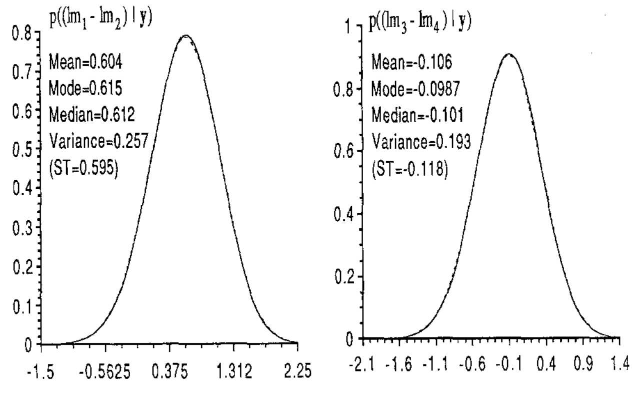
In theory, in a mixed model analysis, genetic differences due to selection should be picked up by the breeding values of animals in the appropriate lines. Hence, significant contrasts between selection and control lines in the fixed part of the model could suggest either randomization failure, or effects due to other pertinent factors not included in the model.

Posterior distributions of selection versus control contrasts for each of the two replicates under Prior IV, are plotted in Fig. 3. In one of the two replicates, there was some evidence of a difference in favor of the selection line, as the posterior mean of that contrast was 0.604. However, the posterior probability that the true difference resides in the interval: posterior mean ± 2 SD $(-0.410, 1.618)$ was 95.4%; this interval includes zero, so the hypothesis that the true difference is zero cannot be rejected. The posterior distribution of the other contrast (line 3 vs 4) was centered near zero. It should be noted that the classical analysis (ST) was in agreement with the Bayesian one, but it is impossible to assign a measure of uncertainty to this estimate, because of the unknown distribution of the classical estimator.

Genetic response to selection

All estimated posterior densities of measures of response to selection under flat priors for variance components (Fig. 4) were nearly symmetric, in spite of the fact that all variance components were unknown. The posterior mean of total response ($TR = t_6 - t_0$) was 1.367 pigs per litter. The estimated posterior means of the regression through the origin (δ), and of the slope (δ_1) of the regression of breeding value on generation were $0.245 (6 \times 0.245 = 1.472$ pigs in total) and $0.273 (-0.120 + 6 \times 0.273 = 1.518$ pigs in total) pigs per litter per generation, respectively. These three posterior distributions indicate that selection was effective; intervals based on the pesterior mean ± 2SD$[0.516, 2.217]$ for TR, $[0.083, 0.407]$ for δ and $[0.118, 0.428]$ for δ_1 had posterior coverage probabilities of 0.950, 0.954 and 0.954, respectively. The evidence from the experiment suggests strongly that selection for the total number of pigs born per litter based on an animal model with repeated records was effective in the Danish Landrace population. An increase of about 1.4 pigs per litter is about 13.6% of the mean of the base population (10.3 pigs); since the actual selection spanned a 6-year period, this implied an increase of about 2.3% per year.

Fig. 3 Estimated marginal posterior distributions of contrasts between selection and control lines under Prior IV (flat), along with estimated summary statistics. Selected lines are 1 and 3, and control lines are 2 and 4. STs are estimates from a classical analysis



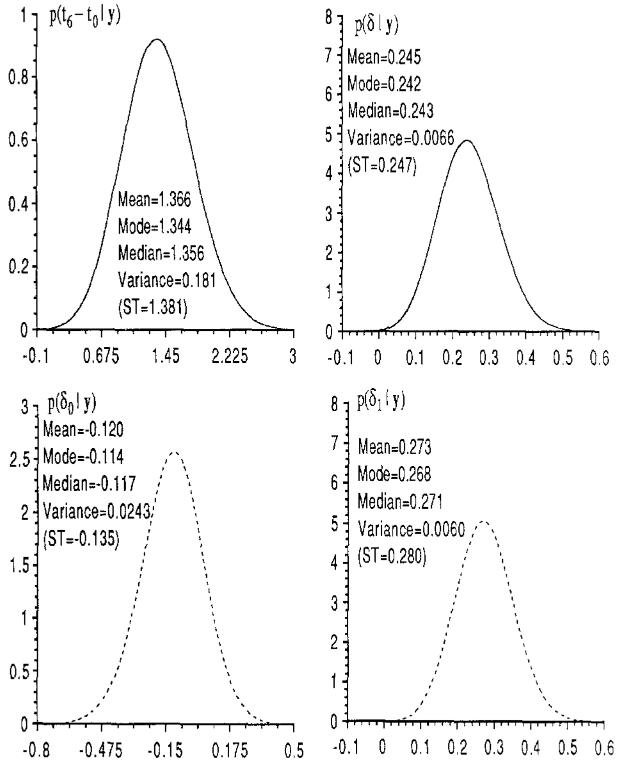


Fig. 4 Estimated marginal posterior distributions of measures of selection response: total response ($t_6 - t_0$), linear regression through origin (δ), and linear regression coefficients (δ_0 for intercept and δ_1 for slope) of breeding value on generation under Prior IV (flat), along with estimated summary statistics. STs are estimates from a classical analysis

The estimated means of the marginal distributions of generation means ($t_i, i = 0, 1, \ldots, 6$) under Priors I–IV, along with point estimates based on a classical analysis, are plotted in Fig. 5. The upper part of the graph pertains to selected lines; the bottom corresponds to control lines. The mean, mode and median of each of the marginal posterior distributions of generation means were close to each other, because these distributions were nearly symmetric. Also included in the figure are the predicted generation means based on:

$$\hat{t}_i = \hat{\delta}_0 + \hat{\delta}_1 i = -0.120 + 0.273\, i; \; i = 0, 1, 2, \ldots, 6 \qquad (27)$$

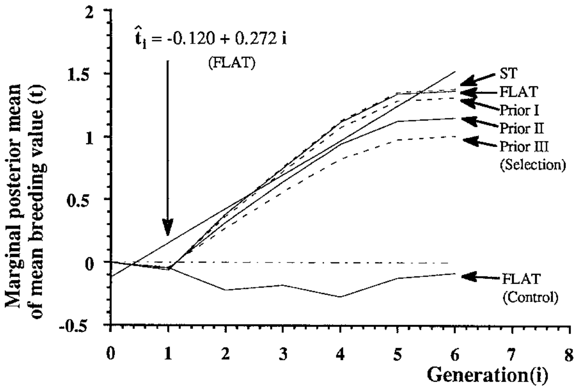


Fig. 5 Plot of mean breeding values against generation. Each point is an estimated posterior mean of an appropriate posterior distribution under a particular set of priors of varience components (see Table 1 for prior specifications). For example, at generation 2, a point (Flat) is the posterior mean of the marginal distribution of t_2, $p(t_2 | \mathbf{y})$. The straight line is based on flat priors. ST means are based on a classical analysis

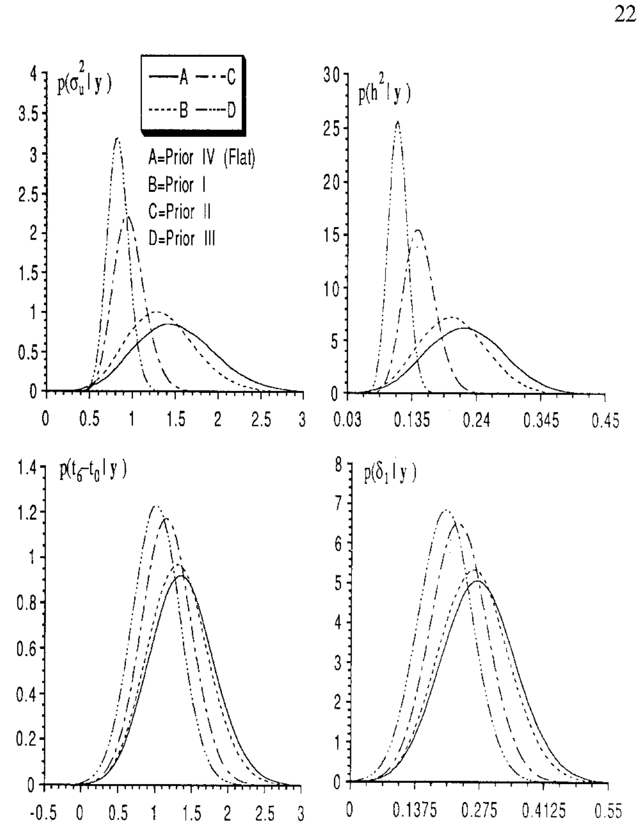


Fig. 6 Estimated marginal posterior distributions of four selected parameters under priors I to IV: genetic variance (σ_u^2), heritability (h^2), total response ($t_6 - t_0$), and slope (δ_1) of linear regression of generation mean additive genetic value on generation. See Table 1 for prior specifications, and Table 2 for estimated summaries

where $\hat{\delta}_0$ and $\hat{\delta}_1$ are the estimated posterior means of the distributions of $\delta_0 | \mathbf{y}$ and $\delta_1 | \mathbf{y}$, respectively, under flat priors. The classical point estimates (ST) of generation means were similar to those obtained in the Bayesian analysis with flat priors. When evidence from this population that was external to the experiment was taken into consideration (Priors I–III), the results indicate that the rate of selection response may have been lower than that suggested by the experimental data. This is because these priors gave increasing weight to the evidence from Estany and Sorensen (in preparation), which indicated lower genetic variance and heritability of litter size in the population than that suggested by the experiment.

This point was studied further. The posterior distributions of four selected parameters (σ_u^2, h^2, TR and δ_1) under Priors I to IV are overlaid in Fig. 6, for comparison. As the prior degree of belief parameters (v_i, $i = u, c, e$) increased, variances of posterior distributions became smaller. The results under Prior I were not very different from those under Prior IV (flat priors). Prior influence under Priors II and III was strong, as expected. As noted earlier, the posterior mean of h^2 (0.112) under Prior III was about 50% of that obtained under the flat prior. The estimated total response ($TR = t_6 - t_0$) in terms of the posterior mean ± SD under Prior III was 1.011 ± 0.315. The slope of the linear regression of breeding values on generation was 0.201 ± 0.057. Although the hypothesis of ineffective selection was still rejected, the information provided by the external evidence indicates that the true rate of reponse is much lower, i.e., about 1.6% of the mean of the base population on an annual basis. It is to be noted that the posterior means of TR and δ_1 under the flat priors fell in regions of relatively high density of the posterior distributions under Prior III. Thus, the experiment cannot be considered to be an "outlier", as is not inconsistent with what is expected from prior information. In the light of all the available evidence, the rate of response to family selection for litter size (as practiced here) in Danish Landrace is expected to be around 1.6% per year. Equivalently, about 6 years of genetic selection would be needed to attain an increase in litter size of one pig.

Analysis excluding control line 1, with flat priors for variance components

It was observed that the behaviour of the control line in replicate 1 was peculiar (Appendix 1, Table A2). In generation one, mean litter size was 10.14, comparable to the base population mean of 10.26 pigs. However, there was a sharp drop of 1.73 pigs per litter (mean 8.41) in generation 2 in this line, for an unknown reason, and mean litter size stayed around 8.0 onwards. A re-analysis excluding data from this control line was conducted using flat priors for variance components. The model was similar to the one applied to the whole data. The reduced data consisted of 805 records from 446 dams.

The estimated posterior means ± SD of additive genetic variance (σ_u^2) and heritability were 0.67 ± 0.38 and 0.11 ± 0.06, respectively, while the estimated posterior means ± SD of permanent environmental variance and repeatability were 1.11 ± 0.41 and 0.28 ± 0.05. The point estimates of genetic parameters were closer to those obtained using the whole data with Prior III than with flat priors. Similarly, the estimated posterior mean of total response (*TR*) was 0.62 ± 0.40, much smaller than the estimate of 1.37 obtained using the whole data with flat priors. The posterior probability that total response was greater than zero was 0.96, still indicating that selection for increased litter size in Danish Landrace pigs was probably effective, the posterior odds in favor of effective selection being 16:1. The estimated posterior means of the regression through the origin (δ), and of the slope (δ_1) of the regression of breeding values on generation were 0.10 ± 0.07 ($6 \times 0.10 = 0.60$ pigs in total) and 0.12 ($-0.08 + 6 \times 0.12 = 0.64$ pigs in total) pigs per litter per generation, respectively. These two posterior distributions again indicate that selection was probably effective, because the posterior probabilities that $\delta > 0$ and $\delta_1 > 0$ were 0.95 and 0.97, respectively.

Discussion

We have presented a Bayesian analysis of a selection experiment for litter size in Danish Landrace pigs where BLUP of breeding value was the selection criterion. The analysis was based on constructing marginal posterior distributions of masures of genetic response, such that all sources of uncertainty about parameters other than the one of interest were taken into account. Our analysis is in marked contrast with a classical one, in which genetic response was assessed from estimated breeding values which, in turn, depended on the values of variance component estimates. In the Bayesian analysis, all information about genetic response is contained in a marginal posterior distribution, so a full Bayesian inference became possible. On the other hand, a classical analysis gave only a point estimate of genetic response; the variance of the estimate is unknown, although some approximations that ignore selection are available (e.g., Sorensen and Kennedy 1983). More generally, the sampling distribution of the classical, three-stage estimation procedure, is completely unknown. The Bayesian approach took uncertainty about nuisance parameters and the selection process into account, and gave a marginal posterior distribution that was estimated by Monte Carlo procedures.

The total genetic response to selection was an increase of 1.37 pigs after five cycles of selection; the linear regression (through the origin) of breeding values on generation was 0.272 (0.245) pigs per litter per generation, the heritability and repetability estimates were 0.22 and 0.32, respectively, under flat priors of variance components. All mean point estimates were different from 0 based on Bayesian confidence intervals, which had coverage probabilities of no less than 0.95. Selection was effective in increasing litter size of Danish Landrace pigs based on these results.

The Bayesian analysis with flat priors of varience components excluding one of the control lines gave smaller total response of 0.62 pigs. The linear regression (through the origin) of breeding values on generation was 0.12 (0.10) pigs per litter per generation, and heritability and repeatability estimates were 0.11 and 0.28, respectively. All three measures of response to selection were significantly different from 0.

The point estimates of genetic parameters for litter size obtained using the whole data in this study were within the range of literature estimates, though on the high side, whereas those obtained excluding records from one of the control lines were closer to literature values. Haley et al. (1988) gave an average heritability estimate of litter size around 0.10, based on results from different sources.

The experiment was not replicated extensively, and the effective population size was relatively small for a selection experiment. It could be argued that the significant response observed could be due to genetic drift, rather than to the force of selection. However, the posterior distribution of measures of response to selection takes into account both the selection process and drift variance; the latter is automatically adjusted for via the matrix of additive relationship among all animals.

Inbreeding is known to affect litter size in pigs. However, inbreeding effects were controlled via a planned mating scheme resulting in a similar increase of inbreeding in selection and control lines, each generation. A restriction in inbreeding level could, in the short term, have somewhat lowered the response to selection.

The size of the fraternity in which a gilt is born has been suggested to influence the size of the litters she would produce subsequently (e.g., Nelson and Robinson 1976; Rutledge 1980). This possible complication was removed by standardizing the first litter of a gilt into two males and two females after birth.

Genes with large effects affecting litter size in this population are unknown. The reason for the large difference in raw means of litter size between the selection and control lines in replicate 1 (Appendix I, Table A2) is unclear, and this may have inflated the estimates of response to selection. A possible explanation for a low mean litter size in one of the control lines could be a chromosomal abnormality in the population, which is known to lower prolificacy in pigs (Popescu 1989). In view of the above, the rate of genetic response to selection for litter size assessed in the experiment from the whole data should be viewed with caution.

Bayesian analysis can be regarded as a compromise between prior information and that contained in the likelihood, from data at hand. It can be thought of either in terms of how data modify prior information, or in terms of how priors influence the analysis. When data are strongly informative about the parameters, the likelihood dominates, and vice versa. Under flat priors for fixed effects and variance components, the marginal density of a parameter is equivalent to a normalized marginalized likelihood. From a robust Bayesian perspective (e.g., Berger 1985), different weak and plausible priors should be tried on the same data; if the conclusions from the alternative priors are vastly different, the results of the analysis should be viewed cautiously, because the likelihood is noninformative, and can be influenced by any prior. This is a strength of Bayesian analysis, because it prevents making false conclusions based on weak likelihoods. In this study, we used four prior sets to analyze the data. The flat prior (Prior IV) represented a situation where we had no strong prior information about fixed effects and variance components. It has some drawbacks, e.g., it is not invariant under re-parameterization. A mildly informative prior (Prior I) for variance components gave similar results to those obtained with the flat priors, indicating that the likelihood had some strength. This was also ascertained from the estimated marginal densities, many of which were nearly symmetrical. Prior III was derived using a method of moments fit from a published data set for this pig population (Estany and Sorensen, in preparation). The analysis with Prior III can be thought of as a combined analysis, via Bayes theorem, of the published data with that of the selection experiment. Beacause the priors had much more information about parameters than that contained in the experiment, it had a profound influence on inferences. The estimates of response to selection ($\widehat{TR} = 1.01$ and $\hat{\delta}_1 = 0.20$) should be considered more realistic than those derived under flat priors. In any case, the Bayesian analysis provided a formal way of combining different sources of information about response to selection for litter size.

The point estimates from a three-stage classical analysis were comparable to those obtained with the Bayesian analysis under flat priors. This was because the marginal posterior distribution of heritability was nearly normal; hence the posterior distributions of measures of response to selection were also nearly normal. In this situation, it is not surprising that the classical analysis, which is based on some asymptotic considerations, worked well. However, a great appeal of the Bayesian analysis is the availability of the whole marginal posterior distribution, from which point estimates can be computed, exact interval estimates can be constructed, and probability statement can be made about a parameter of interest. In the classical analysis, only a point estimate and, at best, an approximate standard error of the estimate, can be obtained.

The Bayesian analysis via Gibbs sampling was computationally expensive. In each of the four analyses made, the CPU time elapsed was approximately 22.3 h using a HP9000/828 running HPUX 8.02. The post Gibbs analysis of density estimation was expensive too, due to the large number of samples saved. However, this cost is minimal relative to the cost of generating the data.

In conclusion, a Bayesian analysis of a selection experiment via Gibbs sampling provided a rich inferential framework that can be strengthened by a consideration of relevant external evidence.

Appendix 1

Characteristics of the data from the selection experiment

Table A1 Number of records by generation and line (*REP* replicate, *S* selection, *C* control, *First* no. of first litters, *Second* no. of second litters, *N* first + second)

Rep	Line	Records	Generation						Total
			1	2	3	4	5	6	
1	S	First	31	32	32	32	28	28	183
		Second	27	23	24	26	23	22	145
		N	58	55	56	58	51	50	328
	C	First	11	12	14	14	13	13	77
		Second	11	10	11	12	11	10	65
		N	22	22	25	26	24	23	142
2	S	First	28	30	27	21	27	27	160
		Second	23	21	21	21	23	23	132
		N	51	51	48	42	50	50	292
	C	First	15	15	20	22	15	16	103
		Second	13	11	13	17	14	14	82
		N	28	26	33	39	29	30	185
Total		First	85	89	93	89	83	84	523
		Second	74	65	69	76	71	69	424
		N	159	154	162	165	154	153	947

Appendix 2

Specification of prior values for v_i and s_i^2 (**i** = **u**, **c**, **e**) using a method of moments fit

An independent field data set on litter size of Danish Landrace pigs consisting of 19 666 records for 7 851 dams, most of them with multiple parties (Estany and Sorensen, in preparation), was used to derive the prior values. An animal model with repeated records similar to (1) was employed. REML estimates (Meyer 1988) of the variance components were $\hat{\sigma}_u^2 = 0.810, \hat{\sigma}_c^2 = 0.531$, and $\hat{\sigma}_e^2 = 6.129$. These REML estimates were used as hyperparameters in (8) and (9), i.e., $s_i^2 = \hat{\sigma}_i^2$ $(i = u, c, e)$. Note that in this data set:

$$\hat{\sigma}_p^2 = \hat{\sigma}_u^2 + \hat{\sigma}_c^2 + \hat{\sigma}_e^2 = 7.470, \; \hat{h}^2 = \hat{\sigma}_u^2/\hat{\sigma}_p^2 = 0.108$$

and

$$\hat{c}^2 = \hat{\sigma}_c^2/\hat{\sigma}_p^2 = 0.071.$$

The asymptotic standard errors were: $SE(\hat{h}^2) = 0.014$ and $SE(\hat{c}^2) = 0.012$, so that h^2 and c^2 were well estimated.

Assuming that each of the variance components has a scaled inverted chi-square distribution, the coefficient of variation (CV) can be shown to be:

$$CV(\sigma_i^2) = \sqrt{2/(v_i - 4)}$$

so the degree of belief parameter is:

$$v_i = 4 + 2/[CV(\sigma_i^2)]^2. \tag{A1}$$

Next, we estimate $CV(\sigma_i^2)$ from the REML estimates and their estimated standard errors. Recalling that $CV(x) = SE\,(x)/x$ and assuming that $\hat{\sigma}_p^2$ is known without error,

$$\begin{aligned}\widehat{CV}(\hat{\sigma}_u^2) &= SE(\hat{\sigma}_u^2)/\hat{\sigma}_u^2 = SE(\hat{\sigma}_p^2\hat{h}^2)/\hat{\sigma}_u^2 \\ &= (7.470 \times 0.014)/0.810 = 0.129.\end{aligned}$$

$$\begin{aligned}\widehat{CV}(\hat{\sigma}_c^2) &= SE(\hat{\sigma}_c^2)/\hat{\sigma}_c^2 = SE(\hat{\sigma}_p^2\hat{c}^2)/\hat{\sigma}_c^2 \\ &= (7.470 \times 0.012)/0.531 = 0.1688.\end{aligned}$$

Table A2 Raw means and standard deviations (SD) of litter size by generation and line (*REP* replicate, *S* selection, *C* control, *N* no. of records)

Rep	Line	Statistics	Generation						Total
			1	2	3	4	5	6	
1	S	N	58	55	56	58	51	50	328
		Mean	9.97	10.35	10.52	10.53	10.24	10.12	10.29
		SD	2.57	2.36	2.49	2.27	2.21	1.95	2.32
	C	N	22	22	25	26	24	23	142
		Mean	10.14	8.41	8.80	7.85	8.29	8.00	8.56
		SD	2.59	2.75	3.08	3.07	2.37	3.33	2.94
2	S	N	51	51	48	42	50	50	292
		Mean	10.61	11.59	10.58	11.52	10.98	10.82	11.01
		SD	2.37	2.69	2.41	2.48	2.41	2.93	2.57
	C	N	28	26	33	39	29	30	185
		Mean	10.64	10.81	10.61	10.46	9.17	10.20	10.32
		SD	2.36	2.58	2.61	2.40	3.12	2.54	2.62
Total		N	159	154	162	165	154	153	947
		Mean	10.31	10.56	10.29	10.35	9.97	10.05	10.26
		SD	2.47	2.74	2.64	2.73	2.64	2.77	2.67

For σ_e^2, we used the result that if h^2 and c^2 are known, $SE(\hat{\sigma}_e^2) = \hat{\sigma}_e^2\sqrt{2/(n-p)}$, where n is the total number of records in the data, and p is the rank of $\mathbf{X}$ (803 in this case). Then:

$$\widehat{CV}(\hat{\sigma}_e^2) = \mathrm{SE}(\hat{\sigma}_e^2)/\hat{\sigma}_e^2 = (\hat{\sigma}_e^2\sqrt{2/(n-p)})/\hat{\sigma}_e^2$$

$$= \sqrt{2/(n-p)} = \sqrt{2/(19\,666-803)}.$$

Finally, using $\widehat{CV}(\hat{\sigma}_i^2)$ in place of $CV(\sigma_i^2)$ in (A1):

$$v_u = 4 + 2/(0.129)^2 \approx 124,$$

$$v_c = 4 + 2/(0.1688)^2 \approx 74$$

and

$$v_e = 4 + (19\,666 - 803) = 18\,867.$$

Acknowledgements We thank the College of Agricultural and Life Sciences (CALS), University of Wisconsin-Madison, and the National Pork Procedure Council, USA, for supporting this work. Part of this research was financed by the Danish Agricultural and Veterinary Research Council. We thank Irma and Erik Bach for the collection of the data and for meticulously looking after the animals. Computations were made on the HP9000/827 of CALS, with additional computing resources generously provided by the San Diego Supercomputer Center. We also thank two anonymous reviewers for comments on the paper.

References

Avalos E, Smith C (1987) Genetic improvement of litter size in the pig. Anim Prod 44:153–164

Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer-Verlag, New York

Bichard M, Bovey M, Seidel L, David P, Tomkins C (1983) New developments in scientific pig breeding. Report No. 3, Pig Improvement Company

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409

Haley CS, Avalos E, Smith C (1988) Selection for litter size in the pig. Anim Breed Abstr 56:317–329

Hill WG (1976) Order statistics of correlated variables and implications in genetic selection programs. Biometrics 32:889–902

IMSL, Inc. (1989) IMSL®, STAT/LIBRARY: FORTRAN subroutines for statistical analysis. Houston, Texas

Kennedy BW (1990) Use of mixed model methodology in analysis of designed experiments. In: Gianola D, Hammond K (eds) Advances in statistical methods for genetic improvement of livestok. Springer-Verlag, New York Heidelberg Berlin, pp 77–97

Lamberson WR, Johnson RK, Zimmerman DR, Long TE (1991) Direct responses to selection for increased litter size, decreased age at puberty, or random selection following selection for ovulation rate in swine. J Anim Sci 69:3129–3143

Legault C (1985) Selection of breeds, strains and individual pigs for prolificacy. J Reprod Fertil (suppl) 33:151–166

Le Roy P, Legault C, Gruand J, Ollivier L (1987) Realized heritability for litter size in selection of hyperprolific sows. Genet Sel Evol 19:351–363

Meyer K (1988) DFREML – A set of programs to estimate variance components under an individual animal model. J Dairy Sci 71 (suppl 2):33–34

Nelson EE, Robison OW (1976) Effects of post-natal maternal environment on reproduction of gilts. J Anim Sci 43:71–77

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. Biometrika 58:545–554

Popescu PC (1989) Cytogenetique des Mammiferes d'elevage. INRA, Paris

Robertson A (1961) Inbreeding in artificial selection programs. Genet Res 2:189–194

Rutledge JJ (1980) Fraternity size in swine reproduction. 1. Effect on fecundity of gilts. J Anim Sci 51:868–870

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London, New York

Smith C, Dickerson GE, Tess MW, Bennett GL (1983) Expected relative responses to selection for alternative measures of life cycle economic efficiency of pork production. J Anim Sci 56:1306–1314

Sorensen DA, Kennedy B (1983) The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. Theor Appl Genet 66:217–220

Sorensen DA, Vernersen A (1991) Large scale selection for number of born piglets using an animal model. 42nd meeting EAAP, Berlin, 1 (Abstract):105

Sorensen DA, Wang CS, Jensen J, Gianola D (1994) Bayesian analysis of genetic change due to selection using Gibbs sampling. Genet Sel Evol (in press)

Tess MW, Bennett GL, Dickerson GE (1983) Simulation of genetic changes in life cycle efficiency of pork production. II. Effects of components on efficiency. J Anim Sci 56:354–368

Toro MA, Silio L, Rodrigalez J, Dobao MT (1988) Inbreeding and family index selection for prolificacy in pigs. Anim Prod 46:79–85

Wang CS, Rutledge JJ, Gianola D (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. Genet Sel Evol 25:41–62

Wang CS, Rutledge JJ, Gianola D (1994) Bayesian analysis of mixed linear models via Gibbs sampling with an aplication to litter size in Iberian pigs. Genet Sel Evol (in press)