Effects of Selection on Estimates of Variance Components Using Gibbs Sampling and Restricted Maximum Likelihood

C. P. VAN TASSELL[1]
Department of Animal Science

G. CASELLA
Biometrics Unit

E. J. POLLAK
Department of Animal Science
Cornell University
Ithaca, NY 14853

ABSTRACT

The effect of selection on estimates of variance components using Gibbs sampling mean and mode, REML, and minimum variance quadratic unbiased estimation was examined through simulation. One hundred replicates were generated for 27 combinations of three levels each of selection schemes, population structures, and heritabilities. All populations consisted of 400 animals. All methods were empirically unbiased, except for the Gibbs sampling estimate of the mode for small variance components, for which the posterior distribution was skewed. Mean squared errors decreased for Gibbs sampling and REML estimates when data were selected, but mean squared errors increased with selection and were largest for minimum variance quadratic unbiased estimation. No pattern existed for differences in mean squared errors for randomly mated and unselected data, suggesting that the differences were due to the direct effect of selection rather than to changes in population structure. Based on these results, the use of minimum variance quadratic unbiased estimation of variance components may be less accurate than other methods for potentially selected field data. Advantages of Gibbs sampling to estimate variance components include simple programming of the Gibbs sampling algorithm and easy calculation of variance of estimates and confidence intervals.

(**Key words**: Gibbs sampling, restricted maximum likelihood, selection, variance components)

Abbreviation key: **GS** = Gibbs sampling, **GSMD** = Gibbs sampling mode, **GSMN** = GS mean, **IG** = inverted gamma, **MIVQUE** = minimum variance quadratic unbiased estimation, **MSE** = mean squared errors, **PS** = population structure, **RP** = randomly mated population, **SP** = selected population, **UP** = unselected population, **VC** = variance components.

INTRODUCTION

Estimation of variance components (**VC**) has long been an important aspect of quantitative genetics. Accurate estimates of VC are important because prediction error variances for predicted genetic values increase as estimated values deviate from the true values (7, 15). The current method of choice for estimation of VC is REML (12). However, computational limitations have restricted the size of the data files that can be considered for estimation of VC. Several strategies have been used to increase that limit, including the use of derivative-free algorithms (10, 11, 20), the use of sparse matrix techniques (13), or both (1, 2). However, these procedures still require the tridiagonalization or Gaussian elimination of the traditional mixed model equations or solutions to the mixed model equations, possibly with fixed effects absorbed. These steps remain as the computational limit for calculation

Received May 27, 1994.
Accepted September 19, 1994.
[1]Current address: A218 Animal Sciences, University of Nebraska, Lincoln 68583-0908.

of REML estimates of VC. Additionally, the variance of these estimates and appropriate confidence intervals can usually be estimated only with the use of approximations because of computational limitations.

An alternative method to estimate VC, Gibbs sampling (**GS**), is investigated here. Gibbs sampling is one method in a larger class of methods, referred to as Marcov chain Monte Carlo methods. Gibbs sampling is based on Bayesian methods for estimation of VC but is evaluated based on its frequentist properties. Use of GS has several advantages: 1) no solution to the mixed model equations is needed; 2) when simple sparse matrix techniques are used, analysis of data files larger than those using REML may be possible; 3) GS yields direct and exact estimates (to any arbitrary precision) of VC and breeding values and confidence intervals for those estimates; and 4) GS is well suited for use on microcomputers and workstations because relatively little information needs to be kept in memory [similar to requirements for iteration data (16)]. Depending on the data structure, the computational demands for GS may be nearly linear for the number of animals in each round, compared with a quadratic to cubic relationship for traditional methods (depending on implementation), although the number of rounds of computations will likely be much larger. Thus, exact (rather than approximate) estimates of VC with confidence intervals may be feasible for large data files on relatively modest computing facilities.

Gibbs sampling is a method of numerical integration that allows inferences to be made about joint or marginal densities, even when those densities cannot be evaluated directly (5). The GS algorithm is based on generation, in sequence, of variables from all of the full conditional densities. The full conditional density is the density of a variable given all other parameters in the model. For example, if GS is used to estimate the distributions of f(a|y), f(b|y), or f(a,b|y), then the full conditional distributions, f(a|b,y) and f(b|a,y), would be required. To use GS to evaluate any of these densities, an arbitrary starting value for one of the variables would be chosen, and then values would be drawn from the full conditional densities in the sequence $a^n \sim f(a|b^{n-1},y)$ and $b^n \sim f(b|a^n,y)$, where ~ indicates that the variable is a random variable from the distribution specified, and the superscript refers to the sequence of the value in the GS chain. If the sequence is repeated enough, the distribution of the a and b samples will be from the distributions f(a|y) and f(b|y), and the a,b sample pairs will be drawn from the f(a,b|y) distribution.

For estimation of VC, the joint density of interest is the distribution of fixed effects, random effects, and VC, all given the data. The marginal densities of interest in this problem are the distribution of fixed effects, random effects, or VC, given the data. In this study, the marginal distribution of the VC, given the data, was of particular interest.

Gibbs sampling has been applied to estimation of VC for a sire model ignoring relationships (4, 27). An animal model has been used by Wang et al. (28) and was used in the present study.

The object of the study was to compare estimates of VC using different methods of estimation. The impact of selection, heritability, and pedigree information were of interest. Three methods of estimation of VC were considered: GS, REML, and minimum variance quadratic unbiased estimation (**MIVQUE**). Data were simulated with and without selection, with several heritabilities, and with different population structures (**PS**).

MATERIALS AND METHODS

Monte Carlo Simulation

Data Simulation. Data for the simulation project were generated using a Monte Carlo procedure similar to that described by Sorensen and Kennedy (22) and applied by van der Werf and de Boer (24) and by Pieramati and Van Vleck (14). This simulation method generated k generations of animals with f females and m males in each generation. The first generation animals were sampled from a conceptually infinite population of unrelated animals with additive genetic variance of σ_a^2 and residual variance of σ_e^2. Genetic and residual effects were distributed normally, and all covariances among base animal genetic and residual effects were null. For generations after the first, a fraction of the males was chosen as sires of the next generation, and each sire was mated to an equal number of females. Each female was mated only once. The mates were

assigned randomly once the sires were chosen; no attempt was made to minimize current or future inbreeding levels. Inbreeding was accounted for only in the Mendelian sampling (Φ_i) variance,

$$\text{Var}(\Phi_i) = \left(\frac{1}{2} - \frac{1}{4}\left(F_{S_i} + F_{D_i}\right)\right)\sigma_a^2$$

(24), where F_{S_i} and F_{D_i} are the inbreeding coefficients (3) of the sire and dam of animal i, respectively. Inbreeding was calculated using the tabular method of calculating relationships (26). No inbreeding depression was simulated. In this study, the number of male and female progeny in each generation were equal, and one-quarter of the males were chosen as sires.

Once the mating assignments were made, one male and one female progeny were generated for each combination of sire and dam. The additive genetic effect of an offspring was simulated as

$$a_i = \frac{1}{2}\left(a_{S_i} + a_{D_i}\right) + \Phi_i,$$

where a_{S_i} and a_{D_i} are the additive genetic effects of the sire and dam of animal i, respectively, and Φ_i, Mendelian sampling, is distributed independently as $N(0,\text{Var}(\Phi_i))$. Finally, the phenotypic value, y_i, of offspring i was calculated as $y_i = a_i + e_i$, where e_i is the independent residual effect distributed $N(0,\sigma_e^2)$.

PS. To evaluate the effect of many versus few generations of selection, three different PS were used: PS1, PS2, and PS3. Population structure 1 had 10 generations of selection with 20 animals of each sex in a generation, PS2 had 5 generations with 40 animals of each sex in a generation, and PS3 had only 2 generations with 100 animals of each sex in a generation. Each PS had a total of 400 animals.

Selection Methods. Three types of selection were applied to all PS to create selected (**SP**), unselected (**UP**), and randomly mated (**RP**) populations. The SP were generated by choosing the sires based on phenotypic value. Each UP was generated corresponding to a SP with an identical relationship structure. The effects of selection on phenotypic value should have been removed from the UP data, but the PS change that was caused by the selection would remain. The procedure is similar to that proposed by Van Vleck and Gregory (25), who suggested simulation of data for a structure obtained from field data. The UP data were generated as follows: 1) calculation of the Cholesky decomposition, **L**, of the numerator relationship matrix, **A**, generated using mass selection (i.e., $\mathbf{LL}' = \mathbf{A}$); 2) calculation of $\mathbf{a} = \mathbf{Lr}$, where **r** is a vector of independent random variables distributed $N(0,\sigma_a^2)$ and 3) $y_i = a_i + e_i$, where a_i is element i of a. Thus, for each population created through selection, a corresponding UP was generated, which had an identical relationship structure but no direct phenotypic selection. In addition, RP were also simulated, for which the same fraction of males was randomly chosen.

Heritabilities. Three heritabilities were examined: .1, .3, and .5. The total variance in the base generation was constant at 20 for all data files. Genetic variances in the base generation were 2, 6, and 10, and residual variance, 18, 14, and 10, for heritabilities of .1, .3, and .5, respectively.

One hundred replicates were generated for each of the 27 combinations of PS, selection method, and heritability.

GS for Estimation of VC

Model. The usual mixed linear model with one random effect was used to analyze the simulated data. The form of the model used was

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e},$$

where **y** is an $n \times 1$ vector of observations, $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed effects, treated in the Bayesian setting as a vector of random effects with a flat prior distribution representing no prior knowledge about the values, **u** is an $r \times 1$ vector of random effects, **e** is an $n \times 1$ vector of random residuals, **X** and **Z** are appropriately dimensioned incidence matrices, and r and n are the number of random effects and observations. Although the only fixed effect considered was an overall mean, the derivation was done for a general animal model. The mean was included because, although the true mean was 0, the phenotypic mean was unlikely to be 0, especially in the selected data files. The phenotypic mean of the selected data files is a

function of the number of generations of selection. Therefore, to be certain that the phenotypic mean did not affect the estimation of VC, the grand mean was included in the model.

Prior Densities and Model Assumptions. Prior distributions are needed to describe the Bayesian model completely. Flat priors for all effects in the model should result in estimates very similar to those obtained using REML (6). However, the use of some improper (i.e., infinite area under the curve) priors, including flat priors, for VC create difficulty in the application of a GS approach because use of these priors results in improper posterior densities of the VC (8, 9, 23).

For the fixed effects, a flat prior was used, so that $f(\boldsymbol{\beta}) \propto$ constant, indicating no prior knowledge about these effects. The random effects were all assumed to be normally distributed, i.e., $\mathbf{u}|\sigma_a^2 \sim N(\mathbf{0},\mathbf{A}\sigma_a^2)$. Independent normally distributed errors were assumed,

$$\mathbf{y}|\boldsymbol{\beta},\ \mathbf{u},\ \sigma_e^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Zu},\ \mathbf{I}\sigma_e^2),$$

where **I** is an identity matrix.

For computational simplicity, a distribution from the conjugate family for the VC in the model was used. The inverted gamma (**IG**) distribution was used as the prior distribution of additive genetic and residual VC. The form of the $IG(\alpha_i,\ \gamma_i)$ distribution is

$$f\left(\sigma_i^2|\alpha_i,\ \gamma_i\right) = \frac{1}{\Gamma(\alpha_i)\gamma_i^{\alpha_i}}\left(\sigma_i^2\right)^{-\alpha_i-1}\exp\left\{\frac{-1}{\gamma_i\sigma_i^2}\right\}$$

$$\sigma_i^2 \geq 0;\ \alpha_i,\ \gamma_i > 0;\ \text{and } i = a,\ e,$$

where α_i is a shape parameter describing the certainty of the knowledge about variance component i, and γ_i is the scale parameter that determines the expected value of variance component i. The expected value of an IG(a,b) variable is $1/[(a - 1)b]$, and the variance is $1/[(a - 1)^2(a - 2)b^2]$.

The inverted chi-square distribution is often used in place of the IG distribution. The inverted chi-square distribution is a special case of the IG, with $\alpha_i = df/2$ and $\gamma_i = 2$, where df is the degrees of freedom associated with the corresponding chi-square distribution. The IG distribution, which is more flexible than the inverted chi-square distribution in the choice of the parameters, was used in the present study.

Initial estimates of VC were calculated using α_i values of 1.000001. The values of γ_i were chosen so that the expected value of the prior distributions were equal to the value of the VC in the base generation of the simulation. Using these values, the mean of the prior distribution was finite, and the variance was infinite. When these priors were used for the data simulated with low heritability (.1), most genetic variance estimates were very near zero, i.e., less than 10^{-4}. To solve this problem, a value for α_i >2 was chosen. A value of 2.000001 was used for all subsequent analyses. The value was chosen so that the variance of the prior was finite and the distribution was as flat as possible, so the estimates should be similar to those obtained with REML. Similar to the previous set of priors evaluated, the value of γ_i was chosen so that the mean of the prior distribution was equal to the variance in the base generation of the simulation. From these assumptions about the distributions, the joint and conditional densities were determined.

Joint Posterior Density. The joint posterior density can be written as the product of the prior and conditional densities previously described. The joint density of the parameters given the data and the prior information is

$$f(\boldsymbol{\beta},\ \mathbf{u},\ \sigma_a^2,\ \sigma_e^2|\mathbf{y},\ \alpha_a,\ \gamma_a,\ \alpha_e,\ \gamma_e)$$

$$\propto f(\boldsymbol{\beta},\ \mathbf{u},\ \sigma_a^2,\ \sigma_e^2,\ \mathbf{y}|\alpha_a,\ \gamma_a,\ \alpha_e,\ \gamma_e)$$

$$\propto f(\mathbf{y}|\boldsymbol{\beta},\ \mathbf{u},\ \sigma_e^2) \times f(\boldsymbol{\beta}) \times f(\mathbf{u}|\sigma_a^2) \times f(\sigma_a^2|\alpha_a,\ \gamma_a) \times f(\sigma_e^2|\alpha_e,\ \gamma_e)$$

$$\propto \left(\sigma_e^2\right)^{-\frac{n}{2}} \times \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})'(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})\right\}$$
$$\times \left(\sigma_a^2\right)^{-\frac{r}{2}} \times \exp\left\{-\frac{1}{2\sigma_a^2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}\right\}$$
$$\times \left(\sigma_a^2\right)^{-\alpha_a - 1} \times \exp\left\{-\frac{1}{\gamma_a\sigma_a^2}\right\} \times \left(\sigma_e^2\right)^{-\alpha_e - 1} \times \exp\left\{-\frac{1}{\gamma_e\sigma_e^2}\right\},$$

which can be rearranged and written as

$$f(\boldsymbol{\beta}, \mathbf{u}, \sigma_a^2, \sigma_e^2|\mathbf{y}, \alpha_a, \gamma_a, \alpha_e, \gamma_e) \propto \left(\sigma_e^2\right)^{-\frac{n}{2} - \alpha_e - 1} \times \left(\sigma_a^2\right)^{-\frac{r}{2} - \alpha_a - 1}$$
$$\times \exp\left\{-\frac{1}{\sigma_e^2}\left[\frac{1}{2}(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})'(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu}) + \frac{1}{\gamma_e}\right]\right\}$$
$$\times \exp\left\{-\frac{1}{\sigma_a^2}\left[\frac{1}{2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + \frac{1}{\gamma_a}\right]\right\}. \quad [1]$$

Full Conditional Densities. The full conditional densities required for GS can be derived from Equation [1] by treating the variables that are known as constants and reorganizing the remaining variables into the form of the kernel of a recognized density.

First, considering only terms that involve $\boldsymbol{\beta}$, the full conditional density of the fixed effects is

$$f(\boldsymbol{\beta}|\mathbf{u}, \sigma_e^2, \mathbf{y})$$
$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})'(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})\right\}$$
$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}[\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{X}'(\mathbf{y} - \mathbf{Zu})]\right\}$$
$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}[(\boldsymbol{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Zu}))'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Zu}))]\right\}$$
$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}[(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})]\right\},$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Zu})$. The full conditional density of the fixed effects is the kernel of a normal density, and, therefore,

$$\boldsymbol{\beta}|\mathbf{u}, \sigma_e^2, \mathbf{y} \sim N(\tilde{\boldsymbol{\beta}}, (\mathbf{X}'\mathbf{X})^{-1}\sigma_e^2). \quad [2]$$

Next, let $\alpha = \sigma_e^2/\sigma_a^2$, and, then, considering only terms that involve $\mathbf{u}$, the full conditional density of the random effects is

$$f(\mathbf{u}|\boldsymbol{\beta}, \sigma_a^2, \sigma_e^2, \mathbf{y})$$
$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})'(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu}) - \frac{1}{2\sigma_a^2}(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u})\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}\left(\mathbf{u}'\left(\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\frac{\sigma_e^2}{\sigma_a^2}\right)\mathbf{u} - 2\mathbf{u}'\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{u} - \tilde{\mathbf{u}})'(\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha)(\mathbf{u} - \tilde{\mathbf{u}})\right\},$$

where $\tilde{\mathbf{u}} = (\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha)^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. This form is also the kernel of a normal density, and, therefore,

$$\mathbf{u}|\boldsymbol{\beta}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N(\tilde{\mathbf{u}}, (\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha)^{-1}\sigma_e^2).$$

However, the conditional distribution of each random effect can be derived using partitioned matrix results (18) and the form of the conditional normal density (17). Let u_i be element i of $\mathbf{u}$, a^{ii} be diagonal element i of $\mathbf{A}^{-1}$, $\mathbf{a}^{-i}$ be row i of $\mathbf{A}^{-1}$ with a^{ii} removed, and $\mathbf{u}_{-i}$ be the vector of random animal effects without element i, u_i. Finally, if animal i has an observation, then $\mathbf{x}_{(i)}$ is the row of $\mathbf{X}$ corresponding to the observation for animal i, and y_i is that observation. Then, if animal i has an observation

$$u_i|\boldsymbol{\beta}, \mathbf{u}_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N\left(\frac{1}{1 + a^{ii}\alpha}(y_i - \mathbf{x}_{(i)}\boldsymbol{\beta} - \mathbf{a}^{-i}\mathbf{u}_{-i}\alpha), \frac{\sigma_e^2}{1 + a^{ii}\alpha}\right), \quad [3]$$

and, if animal i has no observation,

$$u_i|\mathbf{u}_{-i}, \sigma_a^2, \sigma_e^2 \sim N\left(-\frac{1}{a^{ii}\alpha}\mathbf{a}^{-i}\mathbf{u}_{-i}\alpha, \frac{\sigma_e^2}{a^{ii}\alpha}\right).$$

Finally, considering only terms that involve the VC, the conditional distributions of the VC are

$$f(\sigma_a^2|\mathbf{u}) \propto (\sigma_a^2)^{-\frac{r}{2} - \alpha_a - 1} \times \exp\left\{-\frac{1}{\sigma_a^2}\left[\frac{1}{2}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} + \frac{1}{\gamma_a}\right]\right\}$$

and

$$f(\sigma_e^2|\boldsymbol{\beta}, \mathbf{u}, \mathbf{y}) \propto (\sigma_e^2)^{-\frac{n}{2} - \alpha_e - 1} \times \exp\left\{-\frac{1}{\sigma_e^2}\left[\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \frac{1}{\gamma_e}\right]\right\}.$$

These two forms contain kernels of IG densities. Specifically,

$$\sigma_a^2|\mathbf{u} \sim IG\left(\frac{r}{2} + \alpha_a, \frac{1}{\frac{1}{2}(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}) + \frac{1}{\gamma_a}}\right) \quad [4]$$

and

$$\sigma_e^2|\boldsymbol{\beta}, \mathbf{u}, \mathbf{y} \sim IG\left(\frac{n}{2} + \alpha_e, \frac{1}{\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \frac{1}{\gamma_e}}\right). \quad [5]$$

GS Algorithm. The GS algorithm could be simplified because the only fixed effect was an overall mean, μ, and because all animals had observations. The following information was used to reduce computations:

$$(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})'(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu}) = \sum_{i=1}^{n} (y_i - \mu - u_i)^2,$$

$$\mathbf{X'X} = n,$$

and

$$\mathbf{X'}(\mathbf{y} - \mathbf{Zu}) = \sum_{i=1}^{n} (y_i - u_i).$$

Based on the full conditional densities, the GS algorithm used was

1. calculate μ as the arithmetic mean of the observations,
2. calculate $u_i = h^2(y_i - \mu)$, i = 1,...,n,
3. calculate $(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu})'(\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu}) = \sum_{i=1}^{n} (y_i - \mu - u_i)^2$
4. generate σ_e^2 from [5],
5. generate μ from [2],
6. calculate $\mathbf{u'A^{-1}u}$,
7. generate σ_a^2 from [4],
8. calculate $\alpha = \sigma_e^2/\sigma_a^2$,
9. generate u_i from [3], for i = 1, 2, ..., n, and
10. repeat steps 3 through 9.

Because arbitrary values are used for starting a GS chain, a "burn-in" period is required before the samples from the GS can be considered drawn from the joint distribution. In preliminary investigation, the samples drawn over the sequence of the GS were subjectively evaluated for trends and variability. Based on those results, a burn-in of 100 rounds was used for these analyses. To estimate many parameters for a distribution, a sample of independent observations is required. Consecutive samples from a GS chain are correlated, and, in order to obtain independent samples, not all samples can be used. Based on preliminary analysis of samples from the GS chain, samples from every 20th round after burn-in were used. With samples used only every 20th round, the correlation of consecutive samples was <.05. The GS loop was repeated 5000 times for each data file.

Posterior Parameter Estimates. To estimate the GS mean (**GSMN**), the expected value of the VC, given the current value of the sum of squares and priors, was calculated and averaged starting with round 100. The mean of a distribution is unbiased even if correlated samples are used; therefore, all samples after burn-in were used to estimate the mean of a distribution. Based on the expected value of an IG variable,

$$E\left(\sigma_a^2|SSA\right) = \frac{\frac{SSA}{2} + \frac{1}{\gamma_a}}{\frac{r}{2} + \alpha_a - 1},$$

and

$$E\left(\sigma_e^2|SSE\right) = \frac{\frac{SSE}{2} + \frac{1}{\gamma_e}}{\frac{n}{2} + \alpha_e - 1}.$$

Sums of squares for every 20th round after 100 were used to estimate the posterior density of the VC. Based on the estimate of the posterior density for each VC, an estimate of the GS mode (**GSMD**) was calculated as the value with the highest density.

The posterior density of the VC was calculated as the average of the conditional distribution of the VC, i.e., for the posterior of the additive genetic variance,

$$f\left(\sigma^2_{a_i}|\mathbf{y}\right) = \frac{\left(\sigma^2_{a_i}\right)^{-\frac{r}{2} - \alpha_a - 1}}{p \times \Gamma\left(\frac{r}{2} + \alpha_a\right)} \times \sum_{j=1}^{p} \left[\left(\frac{SSA_j}{2} + \frac{1}{\gamma_a}\right)^{\left(\frac{r}{2} + \alpha_a\right)} \times \exp\left\{\frac{1}{\sigma^2_{a_i}}\left(\frac{SSA_j}{2} + \frac{1}{\gamma_a}\right)\right\}\right],$$

$$\sigma^2_{a_i} = .01, .02, ..., 40,$$

and for the posterior of the residual variance:

$$f\left(\sigma^2_{e_i}|\mathbf{y}\right) = \frac{\left(\sigma^2_{e_i}\right)^{-\frac{n}{2} - \alpha_e - 1}}{p \times \Gamma\left(\frac{n}{2} + \alpha_e\right)} \times \sum_{j=1}^{p} \left[\left(\frac{SSE_j}{2} + \frac{1}{\gamma_e}\right)^{\left(\frac{n}{2} + \alpha_e\right)} \times \exp\left\{\frac{1}{\sigma^2_{e_i}}\left(\frac{SSE_j}{2} + \frac{1}{\gamma_e}\right)\right\}\right],$$

$$\sigma^2_{e_i} = .01, .02, ..., 40,$$

where p is the number of samples (i.e., sums of squares) used to estimate the posterior distribution.

REML. The derivative-free REML (20) programs of Meyer (10, 11) were used to estimate VC. The only fixed effect included in the model was the overall mean. The starting value for heritability supplied to the programs was the true simulation values, i.e., .1, .3, and .5.

MIVQUE. The VC were also estimated using a MIVQUE procedure as described by Sorensen and Kennedy (21). Let

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_0\\ \mathbf{C}_1\end{bmatrix} = \begin{bmatrix}\mathbf{C}_{00} & \mathbf{C}_{01}\\ \mathbf{C}_{10} & \mathbf{C}_{11}\end{bmatrix} = \begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z}\\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\gamma\end{bmatrix}^{-1} = \begin{bmatrix}n & \mathbf{1}'\\ \mathbf{1} & \mathbf{I} + \mathbf{A}^{-1}\gamma\end{bmatrix}^{-1},$$

$$\mathbf{W} = [\mathbf{X}\ \mathbf{Z}],$$

$$\hat{\mathbf{S}} = \begin{bmatrix}\hat{\mu}\\ \hat{\mathbf{a}}\end{bmatrix} = \mathbf{CW'y},$$

$$Q_1 = \hat{\mathbf{a}}'\mathbf{A}^{-1}\hat{\mathbf{a}},$$

and

$$Q_2 = \mathbf{y'y} - \mathbf{y'WCW'y},$$

where γ is the prior estimate of α. Then,

$$\begin{bmatrix}\hat{\sigma}^2_a\\ \hat{\sigma}^2_e\end{bmatrix} = \begin{bmatrix}\mathrm{tr}(\mathbf{C}_1'\mathbf{A}^{-1}\mathbf{C}_1\mathbf{WZAZ'W}) & \mathrm{tr}(\mathbf{C}_1'\mathbf{A}^{-1}\mathbf{C}_1\mathbf{W'W})\\ \mathrm{tr}(\mathbf{ZAZ'}) - \mathrm{tr}(\mathbf{CW'ZAZ'W}) & n - \mathrm{tr}(\mathbf{CW'W})\end{bmatrix}\begin{bmatrix}Q_1\\ Q_2\end{bmatrix}$$

$$= \begin{bmatrix}a - 2\gamma\mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}_{11}) + \gamma^2\mathrm{tr}((\mathbf{A}^{-1}\mathbf{C}_{11})^2) & \mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}_{11}) + \gamma\mathrm{tr}((\mathbf{A}^{-1}\mathbf{C}_{11})^2)\\ a\gamma - \gamma^2\mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}_{11}) & n - r(\mathbf{X}) - a + \gamma\mathrm{tr}(\mathbf{A}^{-1}\mathbf{C}_{11})\end{bmatrix}\begin{bmatrix}Q_1\\ Q_2\end{bmatrix}.$$

The true variance ratios were used for γ: 9, 2.33, and 1 for heritabilities of .1, .3, and .5, respectively.

RESULTS AND DISCUSSION

Mean Estimates of VC

Mean estimates of VC for all four methods (GSMN, GSMD, REML, and MIVQUE) with empirical standard errors are presented in Tables 1 and 2. The GSMD estimates appeared to underestimate VC. The underestimation by the GMSD decreased as the VC to be estimated increased and was relatively small for the estimates of the residual variances. The bias was due to the increased skewedness of the distribution as the variance component approached 0. The remaining estimators appeared to be relatively unbiased; that is, none of the remaining estimators seemed to be superior based on their being biased.

Mean Squared Error of Estimates of VC

The mean squared errors (MSE) of estimates of VC for all four methods (GSMN, GSMD, REML, and MIVQUE) are presented in Tables 3 and 4. The MSE was defined as

$$\sum_{i=1}^{n} \frac{1}{n}(\text{estimate}_i - \text{quad}_i)^2,$$

where the quadratic form values (quad) were $\mathbf{u'A^{-1}u}/n$ and $\mathbf{e'e}/n$ for the genetic and residual

TABLE 1. Mean and empirical standard deviations for estimates of genetic variance using Gibbs sampling mean (GSMN), Gibbs sampling mode (GSMD), REML, and minimum variance quadratic variance estimation (MIVQUE) with quadratic form (QUAD), for combinations of heritability (HER), population structure (PS), and selection method (SM).

HER	PS[1]	SM[2]	QUAD[3]		GSMN		GSMD		REML		MIVQUE	
			$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
.1	1	RP	2.0	.01	1.9	.07	1.2	.06	2.2	.12	2.2	.12
.1	1	UP	2.0	.01	1.8	.07	1.2	.06	1.9	.13	1.8	.13
.1	1	SP	2.0	.01	1.7	.06	1.4	.06	1.8	.09	1.6	.19
.1	2	RP	2.0	.01	1.8	.07	1.1	.06	1.8	.13	1.8	.13
.1	2	UP	2.0	.01	1.9	.08	1.2	.07	2.0	.13	2.0	.14
.1	2	SP	2.0	.01	1.8	.06	1.2	.06	2.0	.09	1.9	.15
.1	3	RP	2.0	.01	2.0	.09	1.1	.07	2.4	.17	2.1	.20
.1	3	UP	2.0	.02	1.9	.10	1.1	.09	2.1	.18	2.0	.18
.1	3	SP	2.0	.02	2.0	.10	1.1	.10	2.3	.16	2.2	.18
.3	1	RP	6.1	.04	6.1	.19	5.3	.20	6.3	.22	6.3	.21
.3	1	UP	6.0	.04	5.8	.16	5.0	.16	6.0	.19	6.0	.18
.3	1	SP	6.0	.05	5.9	.14	5.5	.14	6.0	.15	6.1	.28
.3	2	RP	5.9	.04	5.7	.17	4.9	.16	5.9	.19	5.8	.19
.3	2	UP	6.0	.04	5.5	.16	4.8	.17	5.6	.19	5.7	.18
.3	2	SP	6.0	.04	5.9	.15	5.5	.16	6.1	.17	6.1	.26
.3	3	RP	6.0	.04	5.8	.17	5.2	.19	6.2	.20	6.2	.20
.3	3	UP	5.9	.03	5.4	.18	4.7	.19	5.5	.22	5.6	.22
.3	3	SP	6.0	.04	5.9	.17	5.5	.19	6.3	.19	6.3	.22
.5	1	RP	10.1	.07	10.2	.27	9.7	.27	10.3	.29	10.3	.29
.5	1	UP	10.0	.07	9.9	.29	9.3	.29	10.0	.30	9.9	.29
.5	1	SP	9.9	.07	10.0	.18	9.6	.17	9.9	.18	10.0	.31
.5	2	RP	10.2	.07	10.0	.23	9.5	.24	10.1	.24	10.1	.23
.5	2	UP	10.1	.08	10.0	.23	9.4	.23	10.0	.23	10.0	.23
.5	2	SP	10.0	.07	10.1	.18	9.7	.18	10.1	.19	10.2	.29
.5	3	RP	10.0	.07	9.8	.22	9.4	.23	9.9	.23	9.9	.22
.5	3	UP	9.9	.06	9.6	.22	9.1	.23	9.6	.23	9.6	.23
.5	3	SP	10.0	.07	10.1	.22	9.8	.22	10.2	.22	10.3	.26

[1]PS: 1 = 10 generations with 20 animals of each sex in each generation, 2 = 5 generations with 40 animals of each sex in each generation, and 3 = 2 generations with 100 animals of each sex in each generation.

[2]SM: RD = randomly mated population, UP = unselected population, and SP = selected population.

[3]Quad: for genetic variance, $u'A^{-1}u/n$.

variances, respectively, and **u** and **e** are vectors of values generated during the simulation. The values of the quadratic forms were different for each simulated data file. Although not presented, MSE based on the simulation values (2, 6, and 10 for genetic variance and 10, 14, and 18 for residual variance) were very similar to those based on the quadratic forms.

Effects of Selection. The same data were used for all methods of estimation; that is, each data file had VC estimated with each method. Estimation of VC using the same data tended to cause similar patterns for each of the estimation methods when those methods were compared.

There were no systematic differences in the MSE of RP and UP for any of the estimators. This finding suggests that differences observed may have been due to random variability in the simulation of the data files. Additional replicates would be needed to evaluate more precisely the impact of selection on MSE for these two types of data.

However, differences were clear in the MSE of estimates of VC when the SP data were compared with the RP or UP for population structures having multiple generations of selection (PS1 and PS2). For this group of data, the MSE for the estimates obtained by GS (GSMN and GSMD) and REML were smaller for the SP than for the UP or RP. The MSE for MIVQUE, in contrast, were larger for SP than for the UP or RP.

These results agree with those observed in previous studies (14, 21) in which MSE differed for estimates based on SP and RP.

TABLE 2. Mean and empirical standard deviations for estimates of residual variance using Gibbs sampling mean (GSMN), Gibbs sampling mode (GSMD), REML, and minimum variance quadratic variance estimation (MIVQUE) with quadratic form (QUAD), for combinations of heritability (HER), population structure (PS), and selection method (SM).

HER	PS[1]	SM[2]	QUAD[3]		GSMN		GSMD		REML		MIVQUE	
			$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
.1	1	RP	18.0	.12	18.1	.13	17.9	.13	17.9	.15	17.9	.15
.1	1	UP	18.1	.14	18.2	.15	18.1	.15	18.1	.16	18.2	.17
.1	1	SP	18.0	.14	18.2	.14	18.0	.14	18.1	.14	18.2	.18
.1	2	RP	18.1	.11	18.2	.13	18.1	.12	18.1	.15	18.2	.15
.1	2	UP	18.1	.14	18.2	.15	18.1	.15	18.1	.17	18.1	.17
.1	2	SP	18.3	.14	18.6	.16	18.4	.16	18.4	.17	18.4	.20
.1	3	RP	18.1	.14	18.0	.17	18.0	.16	17.7	.21	17.9	.24
.1	3	UP	17.9	.13	17.9	.17	17.9	.17	17.8	.20	17.9	.21
.1	3	SP	18.2	.14	18.2	.18	18.2	.18	17.9	.21	17.9	.22
.3	1	RP	14.1	.09	14.2	.16	14.1	.16	14.0	.17	14.0	.17
.3	1	UP	14.1	.10	14.2	.14	14.1	.14	14.0	.15	14.0	.14
.3	1	SP	14.0	.10	14.0	.13	13.9	.13	14.0	.13	13.9	.20
.3	2	RP	14.0	.09	14.2	.15	14.1	.15	14.1	.16	14.1	.16
.3	2	UP	14.0	.10	14.3	.15	14.2	.15	14.2	.16	14.1	.16
.3	2	SP	14.0	.10	14.1	.14	13.9	.14	14.0	.15	13.9	.20
.3	3	RP	13.8	.09	14.1	.17	13.9	.17	13.7	.19	13.7	.18
.3	3	UP	14.2	.09	14.7	.17	14.6	.18	14.5	.20	14.4	.19
.3	3	SP	14.3	.10	14.5	.16	14.3	.16	14.1	.17	14.2	.18
.5	1	RP	10.0	.07	10.0	.14	9.8	.14	9.9	.14	9.9	.14
.5	1	UP	10.1	.07	10.1	.14	10.0	.14	10.1	.14	10.1	.14
.5	1	SP	9.9	.07	9.9	.11	9.7	.10	9.9	.11	9.8	.17
.5	2	RP	10.1	.07	10.2	.14	10.1	.14	10.1	.15	10.2	.15
.5	2	UP	10.0	.06	10.2	.13	10.1	.13	10.1	.13	10.2	.13
.5	2	SP	9.9	.07	9.9	.14	9.8	.14	9.9	.14	9.8	.20
.5	3	RP	10.0	.07	10.2	.18	10.0	.19	10.1	.19	10.1	.18
.5	3	UP	10.1	.07	10.3	.17	10.1	.17	10.2	.17	10.2	.17
.5	3	SP	10.0	.07	9.9	.18	9.7	.18	9.8	.18	9.8	.21

[1]PS: 1 = 10 generations with 20 animals of each sex in each generation, 2 = 5 generations with 40 animals of each sex in each generation, and 3 = 2 generations with 100 animals of each sex in each generation.

[2]SM: RD = randomly mated population, UP = unselected population, and SP = selected population.

[3]Quad: for residual variance, e'e/n.

Sorensen and Kennedy (21) found an increase in MSE for estimates of VC from SP compared with RP using MIVQUE. Pieramati and Van Vleck (14) found that MSE for REML estimates decreased under similar conditions. Because REML and MIVQUE were never applied to the same data, whether this difference was due to artifacts in the simulated data or to genuine differences caused by selection was uncertain.

The differences in MSE caused by selection did not exist for PS3 for which only two generations of data were present. No consistent pattern existed in the differences for MSE of the methods when only two generations of data were available. The lack of difference suggests that multiple generations of selection are needed to change the MSE.

The effect of selection on phenotypic values, rather than the change in the PS, appears to be the main cause of differences in accuracy for methods of estimation of VC observed in the present study. Differences in the MSE of estimates of VC were large for data based on SP and UP with multiple generations of data (i.e., PS1 and PS2). Because these populations had identical data structures and because the only difference between the two types of data was the presence or lack of selection, selection must have caused the differences. In addition, because the MSE for the UP and RP were similar, the change in PS caused by selection in these populations did not dramatically affect the accuracy of estimation of VC.

Comparison of Estimation Methods. The MSE tended to be smaller for the estimators

TABLE 3. Mean squared errors for estimates of genetic variance using Gibbs sampling mean (GSMN), Gibbs sampling mode (GSMD), REML, and minimum variance quadratic variance estimation (MIVQUE) for the combinations of heritability (HER), population structure (PS), and selection method (SM).

HER	PS[1]	SM[2]	GSMN	GSMD	REML	MIVQUE
.1	1	RP	.46	1.11	1.31	1.34
.1	1	UP	.55	1.11	1.63	1.66
.1	1	SP	.44	.82	.82	3.71
.1	2	RP	.50	1.22	1.52	1.62
.1	2	UP	.67	1.19	1.69	1.98
.1	2	SP	.44	1.11	.91	2.32
.1	3	RP	.85	1.39	3.00	4.19
.1	3	UP	.96	1.58	3.08	3.26
.1	3	SP	.97	1.69	2.74	3.24
.3	1	RP	3.56	4.63	4.79	4.29
.3	1	UP	2.62	3.69	3.59	3.27
.3	1	SP	1.56	1.90	1.83	6.94
.3	2	RP	2.62	3.64	3.27	3.23
.3	2	UP	2.87	4.09	3.57	3.31
.3	2	SP	2.06	2.40	2.41	6.26
.3	3	RP	2.76	4.15	3.83	3.78
.3	3	UP	3.55	5.52	5.04	4.91
.3	3	SP	2.70	3.59	3.74	4.65
.5	1	RP	6.57	6.59	7.43	7.31
.5	1	UP	7.13	7.57	7.72	7.16
.5	1	SP	2.99	3.12	3.15	8.87
.5	2	RP	4.89	5.63	5.25	5.14
.5	2	UP	4.02	4.79	4.15	4.26
.5	2	SP	3.11	3.17	3.24	7.74
.5	3	RP	4.70	5.44	5.07	4.76
.5	3	UP	5.02	5.71	5.55	5.39
.5	3	SP	4.63	5.05	5.01	6.65

[1]PS: 1 = 10 generations with 20 animals of each sex in each generation, 2 = 5 generations with 40 animals of each sex in each generation, and 3 = 2 generations with 100 animals of each sex in each generation.

[2]SM: RD = randomly mated population, UP = unselected populations, and SP = selected population.

based on GS, especially for the data simulated with low heritability. This difference is due, at least in part, to the impact of the prior information used in the GS estimates. To investigate the effect of the expected value of the prior on the estimates of VC, GS estimates were calculated for low heritability data with the priors used for the other heritability levels, .3 and .5. The means and MSE for these estimates are presented in Table 5. These results show the bias introduced by use of a prior with an incorrect expected value, but this bias was relatively small, especially for the estimates based on the priors for a heritability of .3. If MSE are compared, the GSMN estimates compare quite favorably using the moderate heritability, .3. The MSE for the multiple generation data files (PS1 and PS2) are smaller than those for the REML and MIVQUE estimates.

Most traits can likely be categorized a priori into ranges corresponding to the three heritabilities used in this study. These results show that use of this knowledge improves precision of estimation of VC (based on MSE) over current methods. Even if the prior distribution is relatively poor, GS, although possibly biased, may have MSE similar to those for REML and MIVQUE. The effect of the prior distribution decreases as the heritability or the amount of information increases. Therefore, the impact of the prior distribution may be negligible for most parameter estimates based on field data. The effect of the prior was relatively large in this study because small data files (400 animals in each population) were used.

The estimators based on posterior means and modes had quite similar properties when

TABLE 4. Mean squared errors for estimates of residual variance using Gibbs sampling mean (GSMN), Gibbs sampling mode (GSMD), REML, and minimum variance quadratic variance estimation (MIVQUE) for the combinations of heritability (HER), population structure (PS), and selection method (SM).

HER	PS[1]	SM[2]	GSMN	GSMD	REML	MIVQUE
.1	1	RP	.45	.43	.83	.79
.1	1	UP	.52	.49	.99	1.05
.1	1	SP	.37	.32	.46	1.82
.1	2	RP	.51	.46	1.06	1.12
.1	2	UP	.68	.64	.30	1.51
.1	2	SP	.55	.48	.75	1.58
.1	3	RP	1.00	.90	2.77	3.83
.1	3	UP	1.02	.99	2.68	2.79
.1	3	SP	1.15	1.03	2.37	2.78
.3	1	RP	1.63	1.68	1.93	1.75
.3	1	UP	1.21	1.24	1.55	1.37
.3	1	SP	.74	.77	.83	2.62
.3	2	RP	1.51	1.49	1.76	1.66
.3	2	UP	1.41	1.43	1.65	1.48
.3	2	SP	.99	.94	1.05	2.59
.3	3	RP	1.91	2.15	2.48	2.42
.3	3	UP	2.83	3.00	3.76	3.60
.3	3	SP	2.11	2.18	2.51	3.09
.5	1	RP	1.54	1.62	1.67	1.69
.5	1	UP	1.90	2.05	2.00	1.90
.5	1	SP	.74	.78	.76	2.20
.5	2	RP	1.74	1.71	1.81	1.82
.5	2	UP	1.32	1.37	1.32	1.35
.5	2	SP	1.22	1.22	1.27	2.95
.5	3	RP	3.19	3.41	3.35	3.24
.5	3	UP	2.27	2.44	2.53	2.40
.5	3	SP	2.47	2.61	2.67	3.64

[1]PS: 1 = 10 generations with 20 animals of each sex in each generation, 2 = 5 generations with 40 animals of each sex in each generation, and 3 = 2 generations with 100 animals of each sex in each generation.

[2]SM: RD = randomly mated population, UP = unselected populations, and SP = selected population.

TABLE 5. Means and mean squared errors for variance components estimated using Gibbs sampling mean for low heritability using prior values corresponding to .3 and .5 heritabilities (PRIOR) and combinations of population structure (PS), and selection method (SM).

PRIOR	PS[1]	SM[2]	Genetic		Residual	
			$\bar{X}$	MSE	$\bar{X}$	MSE
.3	1	RP	2.94	1.26	17.41	.77
.3	1	UP	2.86	1.33	17.53	.84
.3	1	SP	2.43	.53	17.77	.33
.3	2	RP	2.84	1.19	17.42	.91
.3	2	UP	2.91	1.41	17.48	.94
.3	2	SP	2.62	.66	17.97	.53
.3	3	RP	3.21	2.36	17.01	2.25
.3	3	UP	3.06	2.03	16.97	1.80
.3	3	SP	3.15	2.27	17.19	1.99
.5	1	RP	3.64	3.06	16.99	1.46
.5	1	UP	3.54	2.93	17.13	1.47
.5	1	SP	2.88	1.11	17.54	.48
.5	2	RP	3.51	2.79	16.97	1.73
.5	2	UP	3.58	3.11	17.04	1.68
.5	2	SP	3.14	1.56	17.66	.82
.5	3	RP	3.95	4.76	16.43	3.92
.5	3	UP	3.84	4.39	16.36	3.35
.5	3	SP	3.81	4.16	16.67	3.19

[1]PS: 1 = 10 generations with 20 animals of each sex in each generation, 2 = 5 generations with 40 animals of each sex in each generation, and 3 = 2 generations with 100 animals of each sex in each generation.

[2]SM: RD = randomly mated population, UP = unselected populations, and SP = selected population.

VC were relatively large and the posterior distributions of the VC were fairly symmetric. Differences were large for genetic variance for low heritability data, in which GSMD tended to underestimate genetic variance and, therefore, had larger MSE. Because no computational advantage exists for calculation of the mode of the posterior distribution (which is more difficult to calculate than the mean), the GSMN would be more appropriate than the GSMD for estimation of VC.

Overall, the GSMN and REML estimates were quite similar, especially for the data for traits with high heritability (see Figure 1 for a comparison of the MSE for data simulated with medium heritability and PS2). The GSMN had consistently smaller MSE than REML because of the influence of the prior distribution of the variance component on the posterior distribution. These differences would decrease as the size of the data files increases.

Bayesian estimation using GS produces more than a simple point estimate. Figure 2 shows a posterior distribution for the genetic variance for one of the simulated data files and that use of GS provides an estimate of the distribution that can be used to estimate a mean, confidence interval, or any other function of the distribution.

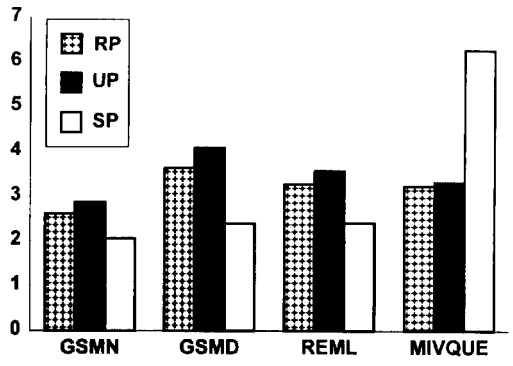

Figure 1. Mean squared errors of genetic variance for randomly mated populations (RP), unselected populations (UP), and selected populations (SP) using Gibbs sampling mean (GSMN), Gibbs sampling mode (GSMD), REML, and minimum variance quadratic unbiased estimation (MIVQUE) estimates of variance components with heritability of .3 and 5 generations of matings with 40 animals of each sex in each generation.

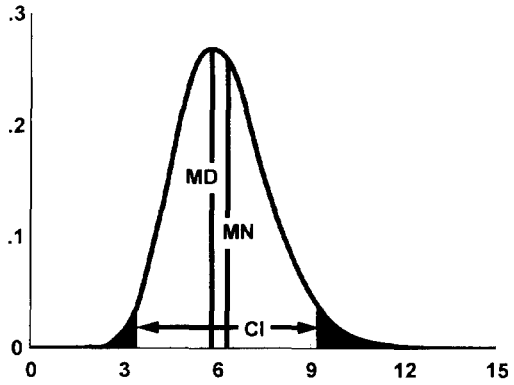

Figure 2. Example of a posterior distribution for genetic variance with mean (MN), mode (MD), and 95% confidence interval (CI) estimated using Gibbs sampling from a replicate with .3 heritability, population structure 2 (5 generations of 40 animals of each sex in each generation), and males chosen with mass selection.

The MSE for MIVQUE estimates had properties for selected data that were very different from those of the GS and REML estimates, although mean estimates were quite similar. If data results from selection, which is likely with field data, MIVQUE, although unbiased if all relationship information is included, may not have smaller MSE than REML, as theory suggests (19).

CONCLUSIONS

All methods for estimation of VC except GSMD were empirically unbiased for all data because all relationships were included in the analyses. The GSMD estimates were biased when VC were small and the posterior distribution was skewed. Selection and data structure affected MSE of the estimation methods differently. There was little impact of selection on the MSE when only one generation of selection was used. When multiple generations were selected, MSE for GSMN, GSMD, and REML decreased compared with that for RP and UP, but MSE for MIVQUE increased and were largest. These results suggest that MIVQUE may not be appropriate for use on field data if selection has been practiced. There were no patterns for the differences in MSE between the RP and UP. Therefore, the change in MSE resulted from the direct effect of selection, not from the change in the PS caused by selection. The MSE for the GSMN were smaller than those from REML because of the impact of the prior distribution of the VC. This difference will decrease as the amount of data used in an analysis increases and the prior distributions have less impact. The use of GS allows calculation of the point estimates and confidence intervals for the posterior distribution of the VC, without approximations or use of normality assumptions. Therefore, for large data files, GS may have advantages over currently used methods.

ACKNOWLEDGMENTS

The authors thank Tony Lednor and Toni Oltenacu for making the IBM RS6000 available for our use. Part of this research was conducted using the resources of the Cornell Theory Center, which receives major funding from the National Science Foundation and New York State.

REFERENCES

1 Boldman, K. G., L. A. Kriese, L. D. Van Vleck, and S. D. Kachman. 1993. A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances (draft). USDA, ARS, USMARC, Clay Center, NE.

2 Boldman, K. G., and L. D. Van Vleck. 1991. Derivative-free restricted maximum likelihood estimation in animal models with a sparse matrix solver. J. Dairy Sci. 74:4337.

3 Falconer, D. S. 1989. Introduction to Quantitative Genetics. 3rd ed. Longman, London, England.

4 Gelfand, A. E., S. E. Hills, A. Racine-Poon, and A.F.M. Smith. 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. J. Am. Stat. Assoc. 85:972.

5 Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intelligence 6:721.

6 Harville, D. A. 1974. Bayesian inference for variance components using only error contrasts. Biometrika 61: 383.

7 Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. J. Anim. Sci. 41:760.

8 Hill, B. M. 1965. Inference about variance components in the one-way model. J. Am. Stat. Assoc. 60: 806.

9 Hobert, J. P., and G. Casella. 1994. Gibbs sampling with improper priors. Tech. Rep. BU-1221-M, Biometrics Unit, Cornell Univ., Ithaca, NY.

10 Meyer, K. 1988. DFREML, a set of programs to estimate variance components under an individual animal model. J. Dairy Sci. 69:1156.

11 Meyer, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative free algorithm. Genet. Sel. Evol. 21:317.
12 Meyer, K. 1990. Present status of knowledge about statistical procedures and algorithms to estimate variance and covariance components. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland XIII:407.
13 Misztal, I., and M. Perez-Enciso. 1993. Sparse matrix inversion for restricted maximum likelihood estimation of variance components by expectation-maximization. J. Dairy Sci. 76:1479.
14 Pieramati, C., and L. D. Van Vleck. 1993. Effect of genetic groups on estimates of additive genetic variance. J. Anim. Sci. 71:66.
15 Schaeffer, L. R. 1984. Sire and cow evaluation under multiple trait models. J. Dairy Sci. 67:1567.
16 Schaeffer, L. R., and B. W. Kennedy. 1986. Computing strategies for solving mixed model equations. J. Dairy Sci. 69:575.
17 Searle, S. R. 1971. Linear Models. John Wiley & Sons, New York, NY.
18 Searle, S. R. 1982. Matrix Algebra Useful for Statistics. John Wiley & Sons, New York, NY.
19 Searle, S. R., Casella, G., and C. E. McCulloch. 1992. Variance Components. John Wiley & Sons, New York, NY.
20 Smith, S. P., and H.-U. Graser. 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. J. Dairy Sci. 69:1156.
21 Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Dairy Sci. 59:1213.
22 Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of response to selection using least-squares and mixed model methodology. J. Dairy Sci. 58:1097.
23 Tiao, G. C., and W. Y. Tan. 1965. Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components. Biometrika 52:37.
24 van der Werf, J.H.J., and I.J.M. de Boer. 1990. Estimation of additive genetic variance when base populations are selected. J. Anim. Sci. 68:3124.
25 Van Vleck, L. D., and K. E. Gregory. 1992. Multiple-trait restricted maximum likelihood for simulated measures of ovulation rate with underlying multivariate normal distributions. J. Anim. Sci. 70:57.
26 Van Vleck, L. D., E. J. Pollak, and E.A.B. Oltenacu. 1987. Genetics for the Animal Sciences. W. H. Freeman and Co., New York, NY.
27 Wang, C. S., J. J. Rutledge, and D. Gianola. 1993. Marginal inferences about variance components in a mixed linear model using Gibbs sampling. Genet. Sel. Evol. 25:41.
28 Wang, C. S., J. J. Rutledge, and D. Gianola. 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. Genet. Sel. Evol. 26:91.