Accounting for outliers and heteroskedasticity in multibreed genetic evaluations of postweaning gain of Nelore-Hereford cattle

F. F. Cardoso,*2 G. J. M. Rosa,† and R. J. Tempelman‡

*EMBRAPA Pecuária Sul (Brazilian Agricultural Research Corporation South Cattle and Sheep Center), Bagé, RS 96401-970; †Department of Dairy Science, University of Wisconsin, Madison 53706; and ‡Department of Animal Science, Michigan State University, East Lansing 48824

ABSTRACT: The objectives of this study were to demonstrate the utility of hierarchical Bayesian models combining residual heteroskedasticity with robustness for outlier detection and muting and to evaluate the effects of such joint modeling in multibreed genetic evaluations. A 3 × 2 factorial specification of 6 residual variance models based on several distributional (Gaussian, Student’s t, or Slash) and variability (homo- skedastic or heteroskedastic) assumptions was used to analyze 22,717 postweaning gain records from a Nelore-Hereford population (40,082 animals in the pedigree). To illustrate the utility of the 2 robust distributional specifications (Student’s t and Slash) for outlier detection and muting, 3 records from the same contemporary group (an extreme residual outlier, a mild residual outlier, and a near-zero residual) were chosen for further study. The posterior densities of the corresponding weighting variables of these records were used to assess their degree of Gaussian outlyingness and the ability of the robust models to mute the effects of deviant records. The Student’s t heteroskedastic provided the best-fit model among the 6 specifications and was preferred for genetic merit inference. Kendall rank correlations of the posterior means of the additive genetic effects of the animals, used to compare the selection order of the Student’s t and Gaussian models, were reasonably high across all animals within the most frequent genotypes, ranging from 0.83 to 0.91 and from 0.89 to 0.95 for the homoskedastic and the heteroskedastic versions, respectively. However, when considering only animals ranked in the top 10% by the customary Gaussian homoskedastic model, these rank correlations were reduced considerably, ranging from 0.29 to 0.57 and from 0.72 to 0.85 between the 2 residual densities within the homoskedastic and heteroskedastic versions, respectively. Rank correlations between the homoskedastic and heteroskedastic versions within each of the Gaussian and Student’s t error models tended to be smaller, with a range from 0.68 to 0.90 across all animals and from 0.28 to 0.67 for animals ranked in the top 10%. These results support the implementation of robust models accounting for sources of heteroskedasticity to increase the precision and stability of multibreed genetic evaluations with proper statistical treatment of deviant records.

Key words: Bayesian inference, beef cattle, genetic evaluation, heteroskedasticity, multibreed, robust model

INTRODUCTION

The accuracy of EBV depends on the quality of the phenotypic and pedigree data available. Performance records influenced by factors that are not explicitly specified in the statistical model but that have potentially large effects (i.e., outliers) can severely bias the genetic parameter estimates and genetic evaluations (Strandén and Gianola, 1998). Preferential treatment, disease, inappropriate contemporary group (CG) formation, and animal misidentification are potential examples of incomplete recording in beef cattle populations.

Data editing for genetic evaluation systems generally involves deleting records that are considered extremely deviant (usually 3 or more SD) from the phenotypic mean of their subclass or that fall outside the range of 60 to 140% of the ratio record or mean of its subclass (Bertrand and Wiggans, 1998). Although this strategy seems suitable for very obvious data recording errors
(e.g., digit transposition), it could be unduly inefficient if the true residual distribution strays from normality and the variances are heterogeneous across various data subclasses. Robust heteroskedastic error models, recently proposed by Cardoso et al. (2005) for genetic evaluations, combine features of structural variance models (Foulley et al., 1992; Kizilkaya and Tempelman, 2005) with those of heavy-tailed densities from the normal-independent family (Lange and Sinsheimer, 1993; Rosa et al., 2003), such as the Student’s t and the Slash distributions, representing an alternative to the deletion of records that could be plausible yet considered extreme relative to the normality assumption.

The objectives of this study were as follows: 1) to demonstrate the utility of hierarchical Bayesian models combining residual heteroskedasticity with robustness for outlier detection and muting and 2) to further evaluate the effects of such joint modeling in multibreed genetic evaluations.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the data were obtained from an existing performance records database.

Nelore-Hereford Data

The data consisted of 22,717 postweaning gain (PWG) records (40,082 animals in the pedigree file) of a beef cattle population composed of Herefords, Nelores, and their crosses, belonging to 15 herds within a large-scale breeding program in Brazil, the Conexão Delta G. The animals were grown on extensive pasture conditions in 3 regions: region 1, located between 14 and 16°S latitude (5,410 records); region 2, located between 21 and 23°S (3,110 records); and region 3, located between 30 and 32°S (14,197 records). The distribution of the breed composition groups was highly unbalanced between 30 and 32°.

The average PWG ± SD was 98.2 ± 41.2 kg. More details on the structure of this data set were presented by Cardoso and Tempelman (2004).

Model Specification and Posterior Inference

Hierarchical Bayesian models combining heteroskedasticity and robustness as proposed by Cardoso et al. (2005) were employed. A $3 \times 2$ factorial representation was used in the specification of 6 residual variance models, with 1 factor being the distributional specification (Gaussian, Student’s t, or Slash) and the second factor specifying variability (homoskedastic vs. heteroskedastic). The resulting 6 residual variance models were as follows: 1) Gaussian homoskedastic (G-HO), 2) Student’s t homoskedastic (T-HO), 3) Slash homoskedastic (S-HO); 4) Gaussian heteroskedastic (G-HE), 5) Student’s t heteroskedastic (T-HE), and 6) Slash heteroskedastic (S-HE).

For all 6 residual variance models, the same linear mixed model was specified for data as an additive function of the location parameters, as in Cardoso et al. (2005):

$$y_{ijklmn} = \mu + \text{reg}_i + \text{cj}_j + \text{sex}_k + \beta_{\text{PWP}}(\text{pwp}_i) + \beta_{\text{AOD}}(\text{aod}_m) + \beta_{\text{AOD}^2}(\text{aod}_m^2) + \beta_{\text{sex}A}(\text{sex}_k \times \text{f}_n) + \beta_{\text{PWP}A}(\text{pwp}_i \times \text{f}_n) + \beta_{\text{AOD}A}(\text{aod}_m \times \text{f}_d) + \beta_{\text{PWP}A^2}(\text{aod}_m^2 \times \text{f}_d) + \beta_{\text{f}(\text{f}_n)} + \beta_{\text{A}(\text{f}_n)} + \beta_D(\delta_n) + \beta_{\text{AA}}(2(1-f_n)f_n) + \text{u}_n + \epsilon_{ijklmn}.$$

Here, $y_{ijklmn}$ is the PWG record on the $n$th animal within the $i$th region, $j$th CG, and $k$th sex; $\mu$ is the overall mean; $\text{reg}_i$ is the effect of the $i$th region ($i \in [1, 2, 3]$); $\text{cj}_j$ is the effect of the $j$th CG ($j \in [1, 2, ..., 940]$), with $\text{cj}_j \sim N(0, \sigma_{\text{cj}}^2)$ for all $j$; $\text{sex}_k$ is the effect of the $k$th sex ($k \in [0, 1]$ or $k = 0$ for male or female, respectively); the $\beta$s are regression coefficients associated with postweaning test period in days (PWP, 106 ≤ PWP ≤ 483), age of dam in years (AOD, 2 ≤ AOD ≤ 12), “fixed” additive effect (A) for the Nelore breed, Nelore-Hereford dominance effect (D), Nelore-Hereford additive × additive epistatic effect (AA), and interactions between A and sex, PWP, and AOD.

The coefficient $f_n$ ($f_d$) represents the expected proportion of Nelore alleles for animal $n$ (dam $d$), whereas $\delta_n$ is the heterozygosity coefficient, obtained as the probability that 1 allele is derived from the Nelore breed and the other allele is derived from the Hereford breed for a randomly chosen locus from individual $n$. Furthermore, $2(1-f_n)f_n$ is the additive × additive epistatic coefficient for animal $n$ based on Kinghorn’s (1980) definition of recombination loss. Now, $u_n$ is the additive genetic effect of the $n$th animal, assuming that $u = u_{n,10082} = 1 - N(0, G)$, where $G$ is a multibreed genetic (co)variance matrix that is a function of the breed-specific genetic variances for the Nelore and Hereford breeds, the segregation variance between the 2 breeds, and genetic relationships among animals (Lo et al., 1993).

Finally, $\epsilon_{ijklmn}$ is a residual term with distribution according to the 6 alternative specifications as outlined above by specifying residual variance conditionally on particular record-specific weighting variables. More specifically, $\epsilon_{ijklmn} \sim N(0, \sigma^2_{\text{rkn}})$, where $\sigma^2_{\text{rkn}} = \sigma^2_{\text{rjkn}}$ is the residual scale parameter specific to the record on animal $n$, as determined by a multiplicative function of an overall scale parameter ($\sigma^2$) and CG ($\gamma$), sex ($\gamma$), breed ($\gamma$), and heterozygosity ($\gamma$) factors for the 3 heteroskedastic error models (G-HE, T-HE, and S-HE). For the 3 homoskedastic error models (G-HO, T-HO, and S-HO), residual variances were more simply specified across all records as $\sigma^2_{\text{rkn}} = \sigma^2_w^{-1}$. Here, $w_n$ is a weighting variable specific for the record on
animal $n$ and whose distribution conditionally defines the marginal residual distribution for the data density. For example, the Gaussian error specifications (G-HO and G-HE) were obtained by setting $w_n = 1$ for all $n$. Otherwise, there are alternative distributional specifications on $w_n$, each yielding a different heavier-tailed marginal density that models the deviation of the distribution of the residuals from a Gaussian distribution (Lange and Sinsheimer, 1993).

We specifically considered 2 heavy-tailed densities: the Student’s $t$ and the Slash distributions, and other alternatives are discussed in Lange and Sinsheimer (1993) and Rosa et al. (2003). The Student’s $t$ residual distribution (T-HE and T-HO) was obtained by assuming that $w_n \sim \text{Gamma}(\nu_e, \nu_e/2)$ for all $n$ with $w_n > 0$, where $\nu_e > 0$ represents the degree of freedom or robustness parameter. The Slash residual distribution (S-HE and S-HO) was alternatively specified by $w_n \sim p(w_n|\nu_e) = \nu_e w_n^{-1}$, for all $n$, with $0 < w_n \leq 1$ and $\nu_e > 0$. See Cardoso et al. (2005) for additional details.

Bayesian inference was based on Markov chain Monte Carlo (MCMC) chains of 200,000 cycles after 15,000 cycles of burn-in for all 6 models; again, full conditional densities and implementation details are provided in Cardoso et al. (2005). Posterior means, modes, key percentiles, and SD of the parameters were obtained from MCMC samples. Kendall rank correlations between posterior means of the additive genetic effects obtained by the various models were used to compare the ordering of the genetic evaluations of the animals for PWG selection within genotype.

RESULTS AND DISCUSSION

Detection and Muting of Outliers

The scatter plot of standardized residuals by CG, derived from the analysis of PWG using G-HO (Figure 1), showed 0.2% of the records with residuals lying outside the range of ±4.0 SD. This percentage is over 30 times that expected given the G-HO error assumption. Beyond attenuating the effect of extreme records on parameter estimates, heavy-tailed error models can be used to aid the detection of outliers (Rosa et al., 2003). The posterior density of $w_n$ provides useful information to assess the corresponding record’s degree of outlyingness relative to a Gaussian error assumption. Low values of $w_n$ (i.e., $w_n$ closer to zero) suggest that the corresponding record $y_{ijklmn}$ is deviant, whereas values of $w_n$ close to 1 indicate that $y_{ijklmn}$ matches well with the model prediction. To illustrate this point in the context of heteroskedastic multibreed genetic evaluations, we deliberately chose for further study the records of 3 F1 males from the same CG containing 160 animals and within region 2. Based on the G-HO specification, the first record ($y_1$) represents a mild outlier, being 3.08 SD above zero; the second record ($y_2$) represents a near-zero residual or a perfect model fit (0.02 SD); and the third record ($y_3$) consists of an extreme outlier, being −5.57 SD from zero (Figure 1).

Posterior densities of the weighting variables $w_1$, $w_2$, and $w_3$, corresponding to the 3 records $y_1$, $y_2$, and $y_3$, respectively, are presented for the T-HE and S-HE error models in Figure 2. We emphasize that only the heteroskedastic versions of these 2 robust models are shown in Figure 2, because they provided a much better fit to this PWG data set based on a Bayesian model choice criterion (Cardoso et al., 2005). One essential difference between the posterior distributions of $w_n$ in the Student’s $t$ and Slash error models is that $w_n$ is defined anywhere on the positive real line for the Student’s $t$, whereas $w_n$ is defined only between 0 and 1 in the Slash error model (Lange and Sinsheimer, 1993). This distinction results in substantially different posterior densities for each of $w_1$, $w_2$, and $w_3$ between the T-HE and S-HE error models. Nevertheless, both models successfully identify $y_3$ as fairly deviant. For instance, the posterior mean and 95% posterior probability interval (PPI) for $w_3$ were, respectively, 0.18 and [0.05, 0.42] for the T-HE model, where the 95% PPI is defined by the [25th, 97.5th] percentiles of the posterior density. Corresponding values for the S-HE model were 0.10 and [0.02, 0.28]. Both densities are then concentrated around low values for $w_3$, thereby implying $y_3$ to be an outlier relative to a G-HE error model specification and such that its effect on fixed and random effects inferences are attenuated. Conversely, $w_2$ has a somewhat flatter distribution relative to the corresponding parameter spaces as defined for the T-HE and S-HE error models. The posterior mean and 95% PPI for $w_2$ corresponding to these 2 models were 1.11 and [0.31, 2.39] and 0.72 and [0.25, 0.99], respectively. That is, these results suggest that $y_2$ is not outlying relative to a G-HE specification.

The case of $y_1$ illustrates the interdependence between robustness and heteroskedasticity specifications. The standardized residual of this record from G-HE model suggested that the influence of $y_1$ on statistical inference is still somewhat downweighted under this model, albeit to a lesser extent relative to $y_3$. This downweighting for $y_1$ was also observed under the S-HE error model, because the posterior mean and 95% PPI for $w_1$ were then, respectively, 0.36 and [0.06, 0.89].

For the Student’s $t$ distribution, the prior expectation of $w_n$ is the neutral value of 1.00; therefore, the containment of 1.00 in the confidence set for $w_n$ could be used as an objective criterion to judge outlyingness with re-
spect to a Gaussian error specification for a desired level of precision. For the example above, the 95% PPI under T-HE model contained 1.00 for \( w_1 \) and \( w_2 \) but not for \( w_3 \), indicating that only \( y_3 \) would be considered an outlier by this criterion.

Conceivably, posterior inference on \( w_n \) could also be used for data quality monitoring if, for example, small posterior means (\(<1.00\)) appeared to be clustered within certain herds or there seemed to be substantial differences for the average posterior means of these weighting variables across herds. In such cases, herds having a high frequency of low posterior means for the weighting variables could be identified as having quality problems with their data collection process. Nevertheless, this issue deserves further investigation before its application can be advised for monitoring data quality in beef cattle performance programs.

The data edits used for determining which records are outliers in animal breeding are often ad hoc in nature and rely upon empirical justifications. For example, the ratio record-mean of its subclass approach advocated by Bertrand and Wiggans (1998), if applied to our PWG data, would have resulted in the deletion of 1,517 records (6.7% of the data), which fall outside the range of 60 to 140%. We believe that only a fraction of these observations likely correspond to recording errors such that the approach of Bertrand and Wiggans (1998) might be too rigid for our particular data set. Evidently, record editing and deletion should be pursued for those cases when a recording error is obvious, and methods for detecting and correcting abnormal records, such as the one used for US dairy test-day yields (Wiggans et al., 2003), provide useful safeguards for current genetic evaluation systems. Nevertheless, it is important to recognize that the proposed heavy-tailed error models differentially weight each record for inferences such that records deemed to be Gaussian-deviant (\( w_n \ll 1 \)) provide proportionately smaller contributions to parameter estimates compared with other records. Therefore, these models provide a much more elegant statistical treatment of deviant records than the rather abrupt alternative of outlier detection and deletion (or adjustment).

**Genetic Evaluations**

Inferences on random additive genetic effects are used to rank animals for selection of breedstock and to predict the expected progeny differences in purebred populations. However, for multibreed populations, fixed additive and nonadditive genetic effects as well as non-additive random effects are also of interest, because they determine multibreed expected progeny differences (MB-EPD; Elzo and Wakeman, 1998; Pollak and Quaas, 1998).

**Comparisons Within Breed Compositions.** We did not consider random nonadditive genetic effects such that inference on random additive genetic effects would suffice for selection of breedstock within breed composition. The Kendall rank correlations between error models G-HO and T-HO, G-HO and G-HE, G-HE and T-HE, and between T-HO and T-HE for posterior means of these additive genetic effects within the most frequent genotypes are presented in Table 1. The rank correlation between the Gaussian and Student’s \( t \) error models within either heteroskedasticity classification (G-HO...
Figure 2. Posterior distribution of weighting variables corresponding to record 1 ($w_1$, a mild outlier), record 2 ($w_2$, a nearly perfect model fit), and record 3 ($w_3$, an extreme outlier) under 2 robust models: a Student’s $t$ heteroskedastic model (panel A) and a Slash heteroskedastic model (panel B).

The rank correlation among the additive genetic values of these top animals ranked in the top 10% using the G-HO error model, the rank correlation among the additive genetic values of these top animals by G-HO and T-HO error models decreased considerably (Table 1). These latter results have greater implications for genetic evaluations and selection decisions than the correlation results involving all animals. The correlations for top-ranked animals might be expected to be lower, because Gaussian-deviant records should lead to more extreme genetic merit predictions on the contributing animal under the G-HO error model compared with the muting effect of the T-HO error model (Stranden and Gianola, 1998), particularly when the outlying record is the primary phenotypic source of information for these animals. Therefore, as pointed out by an anonymous reviewer, the implications of using heavy-tailed error models would be expected to be more evident for replacement seedstock and cows with few offspring than for greatly proven sires in extensively managed beef cattle popula-
tions. This is further evident from Figure 3 (top graph), in which posterior means of additive genetic effects for Nelores, Herefords, and F1’s obtained by the G-HO model are plotted against the corresponding posterior means under the T-HO error model. In this plot, we observed that several animals with extreme posterior means for genetic effects under the G-HO model had these estimates shrunk closer to 0 under the T-HO error model. For example, the F1 animal associated with $y_3$ (the extreme outlier described previously) had a posterior mean of $-13$ kg under the G-HO model but only $-4$ kg under the T-HO error model (Figure 3, top graph). This result is in agreement with simulated results reported by Strandén and Gianola (1998) for homoskedastic models, in which the Student’s $t$ error model was shown to attenuate the effects of preferential treatment and led to more reliable inference on genetic effects of animals associated with outlying records. We also observed a lower rank correlation between the G-HE and T-HE error models when considering only the top 10% of animals than that for all animals (Table 1); however, the magnitude of the decrease in correlation and the degree of change in genetic effects predictions (Figure 4, top graph) were not as substantial, because they were between the G-HO and T-HO error models. Our results suggest then that modeling residual heteroskedasticity is potentially more important than modeling robustness for genetic evaluations. This was further seen by lower rank correlations between the G-HO and G-HE error models compared with between the G-HO and T-HO error models (Table 1). One possible reason for the relatively low correlation between genetic predictions obtained by the G-HO and G-HE error models for the top 10% of animals is that the G-HE error model allows for more balanced selection of animals across environments in accounting for heteroskedasticity, whereas a larger proportion of animals from the most variable environments tend to be ranked near the top when heterogeneous variances are ignored in statistical modeling (Hill, 1984; Gianola, 1986). Moreover, the large difference in genetic variance estimates between the G-HO and G-HE error models (i.e., Herefords had a larger genetic variance than the Nelores in the G-HO error model, whereas the opposite was true for the G-HE error model; see also Cardoso et al., 2005) affects the dispersion of posterior mean additive genetic effects within different breed compositions (Figure 3, bottom graph), and, consequently, the manner in which these predictions overlap within the whole population. Furthermore, several Hereford animals had their genetic effects predictions shrunk closer toward zero with the G-HE compared with the G-HO error models. However, this was not true for the posterior mean genetic effect of the animal with record $y_3$, which was large under both models.

Table 1 and Figures 3 and 4 provide evidence that heteroskedasticity and robustness specifications interact, because rank correlations for posterior mean genetic effects were greater between the heteroskedastic Gaussian and Student’s $t$ error models (G-HE and T-HE) compared with their homoskedastic counterparts (G-HO and T-HO). Similarly, these correlations were larger between homoskedastic and heteroskedastic $t$ error models (T-HO and T-HE) than between homoskedastic and heteroskedastic Gaussian error models (G-HO and G-HE). These results indicate that the effect of adding either residual error specification factor (heteroskedasticity or robustness) to the model on stable genetic evaluations is diminished when the other factor has been modeled.

**Comparisons Across Breed Compositions.** If selection decisions are to be made between breedstock across various breed compositions, fixed additive ($A$) and non-additive ($D$ and $AA$) genetic effects and the interaction of $A$ with other effects, as assumed in equation 1, must be considered in addition to random additive genetic

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Including all animals within genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelore</td>
<td>7,445</td>
<td>0.91</td>
<td>0.86</td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>Hereford</td>
<td>19,976</td>
<td>0.84</td>
<td>0.76</td>
<td>0.89</td>
<td>0.84</td>
</tr>
<tr>
<td>F1</td>
<td>8,718</td>
<td>0.83</td>
<td>0.68</td>
<td>0.93</td>
<td>0.77</td>
</tr>
<tr>
<td>Braford</td>
<td>1,452</td>
<td>0.85</td>
<td>0.82</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td>Considering only animals ranked in the top 10% by the G-HO model within genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelore</td>
<td>745</td>
<td>0.29</td>
<td>0.39</td>
<td>0.72</td>
<td>0.38</td>
</tr>
<tr>
<td>Hereford</td>
<td>1,998</td>
<td>0.36</td>
<td>0.31</td>
<td>0.77</td>
<td>0.67</td>
</tr>
<tr>
<td>F1</td>
<td>872</td>
<td>0.57</td>
<td>0.44</td>
<td>0.85</td>
<td>0.64</td>
</tr>
<tr>
<td>Braford</td>
<td>145</td>
<td>0.37</td>
<td>0.28</td>
<td>0.63</td>
<td>0.55</td>
</tr>
</tbody>
</table>

1No. = sample size of the breed composition group.

Figure 3. Scatter plot of posterior means of additive genetic effects for postweaning gain obtained by the Gaussian homoskedastic (G-HO) and Student’s t homoskedastic (T-HO) models (top) and by the G-HO and Gaussian heteroskedastic (G-HE) models (bottom) for the Nelore, Hereford, and F1 breed composition groups. The highlighted F1 animals are from the same contemporary group and are associated with records representing the G-HO model; 1 represents a mild outlier, being +3.08 SD from zero; 2 represents a near-zero residual (perfect fit); and 3 represents an extreme outlier, at −5.57 SD from zero.

Effects to provide MB-EPD. In this case, expected progeny performance and their differences will be a function of genetic coefficients (breed proportion, heterozygosity, and recombination loss) that will vary according to any breedstock and the genotype of its potential mate. In other words, the selection of candidate parents across breed compositions and prediction of their MB-EPD must occur within a specified mating genotype.

To illustrate this point, we simulated matings of the 2 F1 animals (F1-1 with $u = 7.8$ kg and F1-2 with $u = −0.4$ kg) previously considered in this paper in relation to records $y_1$ and $y_2$, along with 2 Herefords and 2 Nelores of identical additive genetic effects, and compared their future progeny genotypes, genetic coefficients, genotypic means, MB-EPD, their MB-EPD percentiles across the whole population, and the relative rank.
Figure 4. Scatter plot of posterior means of the additive genetic effects for postweaning gain obtained by the Gaussian heteroskedastic (G-HE) and Student’s $t$ heteroskedastic (T-HE) models (top) and by the Student’s $t$ homoskedastic (T-HO) and T-HE models (bottom) for the Nelore, Hereford, and F$_1$ breed composition groups. The highlighted F$_1$ animals are from the same contemporary group and are associated with records representing the G-HO model; 1 represents a mild outlier, being $+3.08$ SD from zero; 2 represents a near-zero residual (perfect fit); and 3 represents an extreme outlier, at $-5.57$ SD from zero.

among these 6 selection candidates, according to alternate dam genotypes (Table 2). When animals Hereford-1, F$_1$-1, and Nelore-1 are mated to Nelore dams, they present substantially different MB-EPD (Table 2) rankings across the whole multibreed population, being within the first, 54th, and 76th percentiles, respectively, despite having the same random additive genetic effects. In this case, differences in MB-EPD are due to different progeny genotypic means (Table 2). However, if these same 3 animals were to be mated to Hereford dams, their MB-EPD would be oppositely ranked, being within the 22nd, 14th, and first percentiles, respectively, across the multibreed population. Moreover, among the 6 animals considered in this example in Table 2, Nelore-1 ranks first, third, or fifth depending on whether the genotype of its mate is Hereford, F$_1$, or
Nelore, respectively. These results further demonstrate the strong dependence between mating genotype and MB-EPD ranking across breed composition. Finally, it should be observed from Table 2 that the difference between the MB-EPD of any of the 2 selection candidates of the same breed composition does not change across the different mating genotypes, being always equal to the difference in their additive genetic effects; these results support the assertion that the latter information suffices for selection within breed composition, at least when nonadditive random genetic effects are not considered.

Practical Implementation and Computational Issues. We have established the importance of jointly considering robustness and heteroskedascity in multibreed genetic evaluations using MCMC inference and a relatively small data set. For much larger data sets involving several million records, such as, for example, that to be analyzed by the National Beef Cattle Consortium in the United States (Pollak, 2006), computational issues will likely arise using MCMC (Jamrozik, 2004). These issues, however, should not discourage the consideration of the proposed robust structural error variance models in beef cattle populations similar to the Nelore-Hereford cross studied in this paper. We have demonstrated these models to substantially outperform the conventional homoskedastic Gaussian error model for data fit. If the presence of outliers is due to the data collection process, the number of extreme records should be expected to be proportional to the data set size such that the importance of model robustness is not likely to diminish even for very large datasets. We further anticipate that computationally attractive analytical procedures will be developed; for example, parameters specifying the Student's t error model could be conceptually estimated using maximum likelihood based on the expectation-maximization algorithm (Lange et al., 1989). Genetic evaluation systems would be more computationally tractable if the df parameter was treated as known much like what is commonly assumed for heritabilities. Presumably, values for these parameters would be based on previous estimates or estimates based on a subset of the data; alternatively, plausible values could be chosen from a discrete set by using a model choice criterion (Stranden and Gianola, 1998). Similarly, structural mixed models for modeling heterogeneous residual variances could also be implemented via the expectation-maximization algorithm (San Cristobal et al., 1993; Weigel et al., 1993) or by simply pursuing heterogeneous residual variance adjustments, similar to those that have been routinely used for dairy cattle genetic evaluations (Wiggans and VanRaden, 1991).
In conclusion, the implementation of robust multibreed genetic evaluations that jointly model sources of heterogeneous genetic and residual variances is recommended to reliably infer upon genetic merit of crossbred animals. Robustness allows for differential weighting of records for inference, such that extremely deviant records provide smaller contributions to genetic evaluations compared with other records. The heterogeneous variances specification properly accounts for distinct genetic and environmental variability inherent to multibreed populations. In our analysis of a Nelore-Hereford cross, a robust Student's $t$ heterogeneous model was shown to better fit the data and to lead to substantially different ranking of top animals for selection as compared with the normal homogeneous error model. Consequently, the use of genetic evaluation models based on normally distributed residuals with homogeneous genetic and residual variances may hinder genetic progress in multibreed beef cattle populations.

**LITERATURE CITED**


