Comparison of restricted maximum likelihood and Method ℜ for estimating heritability and predicting breeding value under selection[1]

R. J. C. Cantet[2], A. N. Birchmeier, M. G. Santos-Cristal, and V. S. de Avila

Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires,
(1417) Buenos Aires, Argentina

ABSTRACT: Method ℜ and Restricted Maximum Likelihood (REML) were compared for estimating heritability (h^2) and subsequent prediction of breeding values (a) with data subject to selection. A single-trait animal model was used to generate the data and to predict breeding values. The data originated from 10 sires and 100 dams and simulation progressed for 10 overlapping generations. In simulating the data, genetic evaluation used the underlying parameter values and sires and dams were chosen by truncation selection for greatest predicted breeding values. Four alternative pedigree structures were evaluated: complete pedigree information, 50% of phenotypes with sire identities missing, 50% of phenotypes with dam identities missing, and 50% of phenotypes with sire and dams identities missing. Under selection and with complete pedigree data, Method ℜ was a slightly less consistent estimator of h^2 than REML. Estimates of h^2 by both methods were biased downward when there was selection and loss of pedigree information and were unbiased when no selection was practiced. The empirical mean square error (EMSE) of Method ℜ was several times larger than the EMSE of REML. In a subsequent analysis, different combinations of generations selected and generations sampled were simulated in an effort to disentangle the effects of both factors on Method ℜ estimates of h^2. It was observed that Method ℜ overestimated h^2 when both the sampling that is intrinsic in the method and the selection occurred in generations 6 to 10. In a final experiment, BLUP(a) were predicted with h^2 estimated by either Method ℜ or REML. Subsequently, five more generations of selection were practiced, and the mean square error of prediction (MSEP) of BLUP(a) was calculated with estimated h^2 by either method, or the true value of the parameter. The MSEP of empirical BLUP(a) using Method ℜ was greater than the MSEP of empirical BLUP(a) using REML. The latter statistic was closer to prediction error variance of BLUP(a) than the MSEP of empirical BLUP(a) using Method ℜ, indicating that empirical BLUP(a) calculated using REML produced accurate predictions of breeding values under selection. In conclusion, the variability of h^2 estimates calculated with Method ℜ was greater than the variability of h^2 estimates calculated with REML, with or without selection. Also, the MSEP of EBLUP(a) calculated using estimates of h^2 by Method ℜ was larger than MSEP of EBLUP(a) calculated with REML estimates of h^2.

Key Words: Predicted Difference, Selection, Selection Methods

Introduction

Dispersion parameters of animal populations are rarely known. Thus, BLUP of breeding values (a) have been calculated using estimates of these parameters in the mixed model equations (**MME**; Henderson, 1984). This two-stage method of calculating BLUP has been referred to by statisticians as "empirical BLUP" (**EBLUP**; Robinson, 1991), to differentiate it from "true" BLUP where the variance components are known. Harville (1985) indicated that, although EBLUP may be unbiased, its mean square error of prediction (**MSEP[EBLUP]**) is larger than the MSEP of BLUP.

Selection affects statistical properties of estimates of dispersion parameters (Gianola and Fernando, 1986; Im et al., 1989), notably bias and mean square error (Rothschild et al., 1979; Ouweltjes et al., 1988; Cantet, 1990). If the distribution of dispersion parameters is affected by selection, MSEP[EBLUP(a)] would also be affected. Use of less accurate estimates of the dispersion

parameters may also affect accuracy of predicted breeding values in selected populations more than in random mating populations. Recently, Method $\Re$ (Reverter et al., 1994a,b) has been proposed as an alternative to REML for estimating the dispersion parameters, because Method $\Re$ is more feasible to calculate than REML. Derivation of Method $\Re$ does not invoke any general principle for obtaining an estimator, such as minimizing mean square error or maximizing a likelihood (Lehman, 1983). Properties of Method $\Re$ under selection are unknown, so objectives of this research were 1) to compare performance of Method $\Re$ with respect to REML under no selection (random choice of replacements) and selection on EBLUP(a), in relation to empirical expectation and mean square error, 2) to disentangle the effects of moment of selection and moment of sampling on the empirical expectation of Method $\Re$ estimates of h^2, and 3) to compare the magnitude of MSEP[EBLUP(a)] calculated with either REML or Method $\Re$.

Materials and Methods

Stochastic simulation was used to predict breeding values with data from selected populations and unknown heritability, because an analytical study of the precision of EBLUP(a) is mathematically complex (Robinson, 1991).

Simulation Procedures

A species with overlapping generations, a yearly mating season, a trait measured in both sexes (e.g., body weight), and females with single progeny in any given year was simulated. The base population consisted of 10 sires and 100 dams. The number of breeding animals was kept constant during either 10 or 15 subsequent cycles of selection and mating in the simulation experiments conducted in this research. The total number of records in each replicate was either 1,110 or 1,610. The heritability was estimated by either REML or Method $\Re$ using the records of the animals in the base population and of the animals in generations 1 to 10. The single-trait animal model used for data generation and breeding value prediction was as follows:

$$y_{ijk} = s_i + g_j + bx_{ijk} + a_{ijk} + e_{ijk}$$

where y_{ijk} is the trait under selection; s_i and g_j are classification variables representing the effects of the ith sex and j[th] contemporary group, respectively; b represents the regression coefficient of the trait on the x_{ijk} age; and a_{ijk} is the random breeding value of animal k and e_{ijk} the random error. For each record, sex was assigned at random with equal probabilities, with males weighing 20 kg more than females. There were 10 contemporary groups (each with 10 records) in every generation, and a total of 100 contemporary groups in every replicate. Contemporary groups were assigned

such that the first animal went to the group 1, the second to group 2, the eleventh to group 1, and so on. Constants for contemporary group effects ranged from 0 to 100 kg. The regression (b) coefficient of the trait on the age of the animal when y_{ijk} was measured was 0.5 kg/d. Age (in days) was assigned at random by drawing from an N(400, 10,000).

The variance of y_{ijk} in the animal model was $\sigma^2 = \sigma_A^2 + \sigma_e^2$, where σ_A^2 is the additive genetic variance or variance of the breeding values (a_{ijk}) and σ_e^2 is the error variance. The phenotypic variance (σ^2) was 500. Heritability (h^2) of the trait is $h^2 = \sigma_A^2/\sigma^2$. For the base population, breeding values were sampled from an N(0, σ_A^2) distribution, with σ_A^2 set equal to $h^2\sigma^2$. In any generation after the base, the breeding value of animal k was calculated as follows:

$$a_k = \frac{1}{2} a_s + \frac{1}{2} a_D + \phi_k$$

(Bulmer, 1985), where a_s and a_d are the breeding values of the sire and of the dam of k, respectively. The random variable ϕ_k is the Mendelian effect of animal k. The variance of breeding values and of ϕ_k were, respectively, equal to:

$$\text{Var}(a_k) = (1 + F_k)h^2\sigma^2$$

$$\text{Var}(\phi_k) = \frac{1}{2}\left[1 - \frac{(F_S + F_D)}{2}\right] h^2\sigma^2$$

where F_k, F_s, and F_D are the inbreeding coefficients of the individual, its sire and its dam, respectively. The ϕ_k were sampled from an N(0,Var(ϕ_k)) distribution. Inbreeding coefficients were calculated using the algorithm of Quaas (1976). Males and females were randomly mated but sire-daughter and dam-son matings were avoided. The three oldest males (replacement rate = 0.33) and the 10 oldest females (replacement rate = 0.10) were culled each generation. The replacements were from progeny born during the same year and selected on either BLUP(a) or EBLUP(a) with h^2 estimated by REML or Method $\Re$. In all cases truncation selection using the greatest BLUP(a) or EBLUP(a) was practiced. All records were used to build the MME.

Estimation of Heritability

Heritability was estimated by either REML or Method $\Re$ at the end of generation 10. Let the animal model in matrix notation be:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{e}$$

with $\boldsymbol{y}$, $\boldsymbol{\beta}$, $\boldsymbol{a}$, and $\boldsymbol{e}$ the vector of records, fixed effects, breeding values, and errors, respectively. The incidence matrix $\boldsymbol{X}$ relates records to parameters in β, and the matrix $\boldsymbol{Z}$ relates records to breeding values. The distri-

bution of the breeding values is N($\boldsymbol{0}$, $\boldsymbol{A}$ h^2 σ^2), where $\boldsymbol{A}$ is the additive relationship matrix.

REML. The REML estimate of h^2 was calculated as follows:

$$\hat{\text{h}}^2_{REML} = \frac{\hat{\sigma}^2_A}{\hat{\sigma}^2_A + \hat{\sigma}^2_e}$$

with the REML estimates of the variance components σ^2_A and σ^2_e calculated by an EM-type algorithm (Dempster et al., 1977), in order to speed up convergence:

$$\hat{\sigma}^{2[k]}_e = \frac{\hat{\boldsymbol{e}}^{[k]'}\hat{\boldsymbol{e}}^{[k]}}{n - p - q + \text{tr}(\boldsymbol{C}^{aa}\,\boldsymbol{A}^{-1})\,\alpha^{[k-1]}}$$

$$\hat{\sigma}^{2[k]}_A = \frac{\hat{\boldsymbol{a}}^{[k]'}\,\boldsymbol{A}^{-1}\hat{\boldsymbol{a}}^{[k]}}{q - tr(\boldsymbol{C}^{aa}\,\boldsymbol{A}^{-1})\,\alpha^{[k-1]}}$$

with n, q, and p, the number of records, animals, and fixed effects in β, respectively. The scalar $\alpha^{[k-1]}$ is the ratio of $\hat{\sigma}^{2[k-1]}_e$ to $\hat{\sigma}^{2}_A{}^{[k-1]}$ in iteration $[k-1]$. Also, $\hat{\boldsymbol{a}}^{[k]}$ and $\hat{\boldsymbol{e}}^{[k]}$ are EBLUP of $\boldsymbol{a}$ and $\boldsymbol{e}$, respectively, calculated at iteration k. The matrix $\boldsymbol{C}^{aa}$ is the portion of the inverse of the coefficient matrix of the MME associated with the breeding values. True values of the parameters were used in the MME as priors.

Method $\Re$. A sampling scheme of 50% was employed as suggested by Reverter et al. (1994a), keeping the same set of solutions in both the sampled ($\mathbf{H}$) and the whole ($\mathbf{W}$) data sets, so that the complete $\boldsymbol{A}$ matrix was used. The sampling scheme consisted of eliminating the "even" records for 50 animals and the "odd" records for the next 50 individuals, and so on. Other similar 50% sampling schemes that were tried (sampling all even records or all odd records) gave similar results and are not reported. The MME using W and H data sets were calculated starting with the true value of h^2 as prior. In order to estimate h^2 by Method $\Re$, we iterated on the following ratio proposed by Reverter et al. (1994a):

$$R_c = \frac{\hat{\boldsymbol{a}}'_w\,\boldsymbol{A}^{-1}\,\hat{\boldsymbol{a}}_H}{\hat{\boldsymbol{a}}'_H\,\boldsymbol{A}^{-1}\,\hat{\boldsymbol{a}}_H}$$

If $R_c < 1$, h^2 is overestimated so that its estimated value should be decreased. On the contrary, if $R_c > 1$, then h^2 is underestimated and its estimated value should be increased. When R_c is equal to 1 the estimated h^2 is equal to the "true" one.

Convergence Criterion. The same convergence criterion was used for both methods of estimation. At the end of each round the current value of h^2 was calculated, and convergence was considered to be attained when the value of the difference between two successive iterates of h^2 was equal to 0.0001. In all cases, Method $\Re$ converged to the stopping criterion in 40 or fewer iteration, whereas the number of REML estimates needed to reach convergence ranged from 2 to 60. Both methods met the convergence criterion in all replicates.

Simulation Schemes

Empirical Expected Values and Mean Square Errors of Heritability Estimates. Based on the loss of pedigree information during selection for predicted breeding values, four situations were used to compare the consistency and the empirical mean square error (**EMSE**) of Method $\Re$ with REML: a) complete information: all records and all pedigree relationships were retained to estimate h^2; b) 50% sires not identified: the paternities of half the records were lost at random; c) 50% dams not identified: the maternities of half the records were lost at random; d) S + D not identified: the paternities of 12.5%, the maternities of 12.5%, and the paternities and maternities of 12.5% of the records were lost at random. Situation a) is the ideal case to estimate h^2 (Sorensen and Kennedy, 1984) and is termed "ignorable selection" by Im et al. (1989). The other situations represent "non-ignorable selection" (Im et al.,1989), and van der Werf and de Boer (1990) observed REML to be biased in this case. All four situations were compared with the case in which the choice of replacement animals was at random: e) "no selection."

To calculate EMSE of h^2 estimates, let θ_{ik} be the estimated value of h^2 due to method i (i = REML or Method $\Re$) for replicate k ($k = 1,..., 100$), and let θ be the true value of the parameter. Then, EMSE$_i$ for method of estimation i was calculated as follows:

$$\text{EMSE}_i = \frac{\sum\limits_{k=1}^{100}(\theta_{ik} - \theta)^2}{100}$$

Time of Selection and Sampling. Estimates of h^2 will be "unbiased" if all data used for selection decisions are taken into account by the estimation method (Im et al., 1989). Such is not the case with $\Re$, because of the sampling procedure involved. In order to disentangle the effects of selection and of sampling, different sets of situations were simulated: f) selection and sampling during generations 1 to 5; g) selection during generations 1 to 5 and sampling during generations 6 to 10; h) selection during generations 6 to 10 and sampling during generations 1 to 5; and i) selection and sampling during generations 6 to 10.

Mean Square Error of Prediction of Breeding Values. Some notation is needed for the statistics used to compare empirical MSEP of breeding values calculated with either REML or Method $\Re$ estimates of h^2. Prediction error variance (**PEV**) of breeding values is denoted as PEV[BLUP(a)], whereas mean square errors of prediction of EBLUP(a) are denoted as MSEP[EBLUP(a) + REML] or MSEP[EBLUP(a) + R]. Let $\boldsymbol{a}_{ikl}$ be the breeding value and $\hat{\boldsymbol{a}}_{ikl}$ be either the BLUP(a) or EBLUP(a) of animal l ($l = 1,..., 1,610$; 110 in the base population plus 15 generations of 100 animals each), within replicate k ($k = 1,..., 100$). The subscript i corresponds to either the use of true h^2, or $\hat{\text{h}}^2$ estimated by REML or by Method $\Re$, in the mixed-model equations.

Table 1. Average estimates of heritability (h^2) derived by REML and Method $\Re$ from simulated data with true values for h^2 of 0.1, 0.4, and 0.7

	$h^2 = 0.1$		$h^2 = 0.4$		$h^2 = 0.7$	
Item	REML	$\Re$	REML	$\Re$	REML	$\Re$
Complete information[a]	0.100	0.100	0.400	0.412	0.697	0.761
50% Sires (S) not identified[b]	0.083	0.082	0.321	0.374	0.554	0.657
50% Dams (D) not identified[c]	0.092	0.112	0.365	0.438	0.637	0.757
S+D not identified[d]	0.062	0.060	0.295	0.351	0.526	0.644
No selection[e]	0.097	0.109	0.398	0.408	0.693	0.717

[a]Complete genealogical information: no loss.
[b]After simulating the data, a random 50% of sire identifications were lost.
[c]After simulating the data, a random 50% of dam identifications were lost.
[d]S+D not identified: the paternities of 12.5%, the maternities of 12.5%, and the paternities and maternities of 12.5% of the records were lost at random.
[e]No selection: random choice of replacements, with complete genealogical information.

The statistics PEV[BLUP(a)], MSEP[EBLUP(a) + REML], and MSEP [EBLUP(a) + R], were calculated in two steps. At the end of every replicate, the 1,610 differences ($\hat{\boldsymbol{a}}_{ikl} - \boldsymbol{a}_{ikl}$) were averaged using the following formula:

$$\text{MSEP}_{ik} = \frac{\sum\limits_{l=1}^{1,610} (\hat{\boldsymbol{a}}_{ikl} - \boldsymbol{a}_{ikl})^2}{1,610}$$

Finally, the 100 MSEP_{ik} were averaged to obtain MSEP_i (i = true h^2, or $\hat{h}^2$ estimated by REML or by Method $\Re$) as follows:

$$\text{MSEP}_i = \frac{\sum\limits_{k=1}^{100} \text{MSEP}_{ik}}{100}$$

Results

Empirical Expected Values of Heritability Estimates

The REML and $\Re$ estimates of h^2, as averages of 100 replicates, are shown in Table 1. When all relationships among animals were used to estimate h^2 under selection (complete information), REML estimates were closer to the base population h^2 than method $\Re$ estimates. When a fraction of the parental identifications, or maternal identifications, or both, were misssing, estimates of h^2 by REML and by Method $\Re$ were biased due to a "nonignorable" selection process (Im et al., 1989). When no selection was practiced and replacement animals were chosen at random, both methods of estimation were seemingly unbiased.

Mean Square Errors of Heritability Estimates

The EMSE of h^2 estimates are shown in Table 2. The EMSE of Method $\Re$ were approximately 2.7 to 8.7 times larger than the EMSE of REML.

Timing of Selection and Sampling

The theoretical results of Im et al (1989) may suggest that the sampling involved to estimate h^2 by Method $\Re$ would produce estimates closer to the parameters for situations g) and h) than for f) and i). This result was observed when selection and Method $\Re$ sampling occurred in generations 6 to 10 (situations h and i). Because there was no sampling involved to obtain REML estimates, they were not affected by these sampling procedures. Also, generations of selection did not affect REML estimates of h^2 (columns 1, 3, and 5, Table 3) because all data used for selection purposes were involved in the estimation process. The EMSE values of h^2 estimates corresponding to the schemes discussed in this section are not shown, because they are similar to the EMSE in Table 2.

Mean Square Error of Prediction of Breeding Values When Heritability is Unknown

Effects of method of estimation of h^2 on the mean square error of prediction of EBLUP(a) in populations under random mating or selected by BLUP(a) or EBLUP(a) are presented in Table 4. The first row shows empirical PEV, using BLUP(a) calculated with the true value of h^2. The analyses of data arising from random mating resulted in higher PEV[BLUP(a)] than analyses of selected data. The opposite was true when h^2 was estimated by either REML (row 2) or Method $\Re$ (row 3). Under either random mating or selection, PEV[BLUP(a)] was always smaller than MSEP[EBLUP(a) + REML], which was, in turn, smaller than MSEP[EBLUP(a) + R].

Discussion

An objective of this research was to check on the consistency and EMSE of Method $\Re$ under different types of selection. The first selection scheme was identified as "complete information" in Table 1 and has the following characteristics: 1) complete pedigree informa-

Table 2. Average mean square error for estimates of heritability (h^2) derived by REML and Method $\Re$ from simulated data with true values for h^2 of 0.1, 0.4, and 0.7

Item	$h^2 = 0.1$		$h^2 = 0.4$		$h^2 = 0.7$	
	REML	$\Re$	REML	$\Re$	REML	$\Re$
Complete information[a]	0.00104	0.00348	0.00322	0.00944	0.00261	0.02262
50% Sires (S) not identified[b]	0.00086	0.00317	0.00320	0.01006	0.00253	0.01109
50% Dams (D) not identified[c]	0.00134	0.00704	0.00329	0.01659	0.00271	0.01927
S+D not identified[d]	0.00081	0.00291	0.00441	0.01371	0.00432	0.01832
No selection[e]	0.00115	0.00432	0.00334	0.00910	0.00198	0.01644

[a]Complete genealogical information: no loss.
[b]After simulating the data, a random 50% of sire identifications were lost.
[c]After simulating the data, a random 50% of dam identifications were lost.
[d]S+D not identified: the paternities of 12.5%, the maternities of 12.5%, and the paternities and maternities of 12.5% of the records were lost at random.
[e]No selection: random choice of replacements, with complete genealogical information.

tion to a base population of unselected, unrelated, and noninbred animals (Sorensen and Kennedy, 1984) and 2) all data employed to estimate h^2 were retained and their distribution was completely specified (Im et al., 1989). Under this scheme, the mean of REML estimates was closer to true parameter values than the mean of Method $\Re$ estimates. Using the missing data theory of Rubin (1976), Im et al. (1989) concluded that if estimates of the dispersion parameters are calculated using all data employed in making selection decisions, the selection process may be "ignored," and estimation may proceed as if selection never occurred. Computer simulations have shown that methods based on likelihood functions, such as Maximum Likelihood (**ML**; Rothschild et al., 1979; Cantet, 1990) or REML (Ouweltjes et al., 1988; Cantet, 1990), were not biased by "ignorable" selection. The three remaining selection schemes were simulated by losing a fraction of parental identities. These schemes produced estimates of h^2 by both methods that were biased downward (rows 2 to 4 in Table 1). When replacement animals were chosen at random (no selection), both REML and Method $\Re$ estimates were close to the true values of h^2 (last row in Table 1). The magnitude of EMSE for Method $\Re$ estimates of h^2 were systematically several times larger than EMSE of REML (Table 2), indicating that Method $\Re$ was far less accurate in this research. There is ample evidence

that the estimation of dispersion parameters in selected populations using likelihood-based methods produces smaller EMSE than quadratic methods (ML vs Method I of Henderson, Rothschild et al., 1979; REML vs Henderson's simple method, Ouweltjes et al., 1988; ML and REML vs MIVQUE and Henderson type III, Cantet, 1990). Data reported in Table 3 included all combinations of generation of selection and generation of sampling and showed that the generation when sampling occurred had little effect on Method $\Re$ estimates of h^2 when selection and sampling occurred in the initial (1 to 5) generations. This effect was larger when selection and sampling occurred in the final (6 to 10) generations. The use of the complete relationship matrix by Method $\Re$ may have partially accounted for the selection practiced in our simulation, even though some of the records on which selection decisions have been based were not included in the sampling to calculate Method $\Re$. Sorensen and Kennedy (1984), using MIVQUE, and van der Werf and de Boer (1990), using REML, observed that use of the complete relationship matrix produced seemingly unbiased estimates of h^2 under selection, when not all the records on which selections decisions were made were used in the estimation.

In practice, animal breeders calculate BLUP(a) by replacing the true value of h^2 by an estimate. Kackar and Harville (1981) found that this two-stage proce-

Table 3. Average estimates of heritability (h^2) derived by REML and Method $\Re$ from simulated data with true values for h^2 of 0.1, 0.4, and 0.7

Item	$h^2 = 0.1$		$h^2 = 0.4$		$h^2 = 0.7$	
	REML	$\Re$	REML	$\Re$	REML	$\Re$
Selection I, Sampling I[a]	0.101	0.108	0.392	0.395	0.695	0.716
Selection I, Sampling F[b]	0.101	0.111	0.392	0.393	0.695	0.710
Selection F, Sampling I[c]	0.099	0.104	0.391	0.391	0.695	0.698
Selection F, Sampling F[d]	0.099	0.107	0.391	0.427	0.695	0.725

[a]Selection and Method $\Re$ sampling in generations 1 to 5.
[b]Selection in generations 1 to 5 and Method $\Re$ sampling in generations 6 to 10.
[c]Selection in generations 6 to 10 and Method $\Re$ sampling in generations 1 to 5.
[d]Selection and Method $\Re$ sampling in generations 6 to 10.

Table 4. Mean square error of prediction for breeding values for randomly mated and selected populations when heritability was estimated by either REML or Method ℜ

Item	$h^2 = 0.1$ Random	Selection	$h^2 = 0.4$ Random	Selection	$h^2 = 0.7$ Random	Selection
True	39.12	37.92	98.30	95.48	104.31	101.08
REML	40.08	47.51	99.50	103.27	105.31	108.80
ℜ	41.62	63.47	101.55	113.79	112.64	198.38

dure, more appropriately called EBLUP(a) (Robinson, 1991; Harville and Carriquiry, 1992), yields unbiased estimators of linear functions of the fixed effects and predictors of functions of random effects, provided that 1) the data vector is symmetrically distributed about its expected value; 2) the estimators of the variance components are translation-invariant (i.e., the expected value of the estimating statistics does not depend on the fixed effects); and 3) the variance components are an even function of the data vector. A statistic S is said to be an even function of a data vector y if $S(y) = S(-y)$. Restricted Maximum Likelihood (Patterson and Thompson, 1971) uses even and translation invariant statistics (Kackar and Harville, 1981). Harville and Carriquiry (1992) pointed out that MSEP of EBLUP(a) (say, "PEV[EBLUP(a)]") calculated as if variance components were known seems to underestimate MSEP[-EBLUP(a)] by approximately two times Var[EBLUP(a) – BLUP(a)]. In other words; "PEV" of breeding value calculated under the pretense that dispersion parameters are known underestimate MSEP. The use of concepts such as bias, PEV, or MSEP in the mixed linear model requires that the matrices X, Z, and A be fixed under repeated sampling. This approach was followed by Henderson (1975) to derive his conditional selection model. However, as pointed out by Gianola (1999), the idea is unrealistic from an animal breeding point of view. Fixing the incidence and the relationship matrices would imply, for example, that in all conceptual replicates the same contemporary group effect and age are assigned to the same animals, and the same relationships among animals will be established. Therefore, fixing X, Z, and A restricts the number of cases examined, making the results applicable only to those situations. At the same time, calculating MSEP in a conditional scheme suggests some caution to using the analytical results of Harville (1985) on the decomposition of prediction error. Harville (1985) decomposed prediction error into four terms due to 1) inherent, "in the sense that it would be unavoidable even if there were complete knowledge of the joint distribution" of a and y, 2) unknown distributional form, 3) unknown means, and 4) unknown dispersion parameters. Harville (1985) gave conditions under which all four components are uncorrelated, one of them being multivariate normality of a and e. The MSEP[EBLUP(a)] calculated with h^2 estimated by either REML or Method ℜ were higher than PEV[BLUP(a)] because of the presence of the fourth term. Another possible explanation may lie in the generation of positive covariances among some of the terms. Thus, the term due to unknown dispersion parameters may be highly correlated with the term related to the unknown means, indicating that a high error in estimating h^2 induces a higher prediction error due to having to estimate the means. In the present research, the increased variability in estimates of h^2 affected EBLUP(a). Therefore, as EMSE of Method ℜ was higher than EMSE of REML, MSEP of EBLUP(a) calculated with Method ℜ was greater than that calculated with REML estimates. The latter statistic was closer to PEV[BLUP(a)], indicating that EBLUP(a) calculated using REML estimates produced more accurate predictions of breeding value than Method ℜ, especially under selection. In conclusion, the variability of h^2 estimates calculated with Method ℜ was greater than the variability of h^2 estimates calculated with REML, with or without selection. Also, the MSEP of EBLUP(a) calculated using estimates of h^2 by Method ℜ was larger than MSEP of EBLUP(a) calculated with REML estimates of h^2.

Implications

For genetic evaluations with actual data, dispersion parameters are unknown. However, prediction of breeding values proceeds as if dispersion parameters are known. This approach underestimates the possible changes in predicted breeding value, although the effect is not large when variance components are estimated by REML. Estimates of h^2 by Method ℜ are more variable than estimates of h^2 by REML, either for unselected data or for data undergoing selection, with a complete or an incomplete relationship matrix.

Literature Cited

Bulmer, M. G. 1985. The mathematical theory of quantitative genetics. 2nd ed. Clarendon Press, Oxford, U.K.

Cantet, R. J. C. 1990. Estimation and prediction problems in linear mixed models for maternal genetic effects. Ph.D. dissertation. University of Illinois, Urbana.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Stat. Soc. 39:1–38.

Gianola, D. 1999. Statistical methods in animal breeding: Wherefrom and whereto? In: Proc. 50th Annu. Mtg. European Assoc. Anim. Prod., Zurich. p 25.

Gianola, D., and R. L. Fernando. 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63:217–244.

Harville, D. A. 1985. Decomposition of prediction error. J. Am. Stat. Soc. 80:132–138.

Harville, D. A., and A. L. Carriquiry. 1992. Classical and Bayesian prediction as applied to an unbalanced mixed linear model. Biometrics 48:987–1003.

Henderson, C. R. 1975: Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447.

Henderson, C. R. 1984. Applications of linear model in animal breeding. 1st ed. University of Guelph, Ontario, Canada.

Im, S., R. L. Fernando, and D. Gianola. 1989. Likelihood inferences in animal breeding under selection: A missing-data theory view point. Genet. Sel. Evol. 21:399–414.

Kackar, R. N., and D. A. Harville. 1981. Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. Comm. Stat. 10:1249–1261.

Lehman, E. L. 1983.Theory of Point Estimation. 1st ed. John Wiley & Sons, New York.

Ouweltjes, W., L. R. Schaeffer, and B.W. Kennedy. 1988. Sensitivity of methods of variance component estimation to culling type of selection. J. Dairy Sci. 71:773–779.

Patterson, H. D., and R. Thompson. 1971. Recovery of inter-block information when blocks sizes are unequal. Biometrika 58:545–554.

Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:940–953.

Reverter, A., B. L. Golden, R. M. Bourdon, and J. S. Brinks. 1994a. Method $\Re$ Variance components procedure: application on the simple breeding value model. J. Anim. Sci. 72:2247–2253.

Reverter, A., B. L. Golden, R. M. Bourdon, and J. S. Brinks. 1994b. Multiple-trait method $\Re$ estimation of (co)variance components. In: Proc. 5th World Congr. Genetics Applied to Livest. Prod. 18:406–409.

Robinson, G. K. 1991. That BLUP is a good thing: The estimation of random effects. Stat. Sci. 6:15–51.

Rothschild, M. F., C. R. Henderson, and R. L. Quaas. 1979. Effects of selection on variances and covariances of simulated first and second records. J. Dairy Sci. 62:996–1002.

Rubin, D. B. 1976. Inference and missing data. Biometrika 63:581–592.

Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci. 59:1213–1223.

van der Werf, J. H. J., and I. J. M. de Boer. 1990. Estimation of additive genetic variances when based populations are selected. J. Anim. Sci. 68:3124–3132.