EFFECTS OF MISSPECIFICATION ERRORS IN THE DISPERSION MATRIX OF RANDOM EFFECTS ON ACCURACY OF BLUP OF BREEDING VALUE

R.J.C.Cantet[1]

Departamento de Producción Animal, Facultad de Agronomía,
Universidad de Buenos Aires, Argentina.

Summary

The effects of misspecifying the dispersion matrix of random effects on prediction error variance (PEV) and accuracy of breeding value (BV) were studied analytically. The first type of misspecification arises when excluding a set of random effects from the model, that may or may not be correlated to BV. The analysis was based on two operational models: the long model, which includes both random effects, and the short model which only includes BV's. Expressions for PEV(BV) were obtained for: 1) the long model, 2) the short model, and 3) the short model assuming the long model is the correct one. It is shown analytically that accuracy of BV prediction in 2) is always greater or equal than accuracy calculated under 1). However, when the long model is correct PEV(BV) in 1) is greater or equal than accuracy of BV calculated as in 3). The other misspecification error studied was the use of a dispersion matrix of BV which accounts for less variation. In this case too accuracy of BV calculated under the model that accounts for less additive variance is greater or equal than accuracy of BV calculated with the correct covariance matrix, when each model is considered to be true. On the contrary, accounting for more variance produce smaller or equal PEV(BV) than accounting for less variance of BV, when the model that accounts for more variance is the correct one. Results are used to discuss accuracy of BV prediction in relation to selection response, with special emphasis on comparing alternative selection strategies by means of stochastic simulations.

Introduction

Prediction of breeding values (BV) is an essential step in the process of selection. Generally, animal breeders predict BV using Best Linear Unbiased Prediction (BLUP, Henderson, 1984), by means of mixed linear models. This is acomplished by writing the structure of the variance-covariance matrices of the data and of the random effects, the latter

---

[1]. Facultad de Agronomía, UBA. Av. San Martín 4453.
(1417) Buenos Aires, Argentina.

usually based on quantitative genetic theory. There are situations in which the model of analysis includes other random effects beside BV. Examples are maternal effects (Willham, 1963), dominance effects (Falconer, 1981), and marked quantitative trait loci effects (MQTL, Fernando and Grossman, 1989). However, the use of these complex models is not always feasible because of increased computing demands, or lack of estimates of the dispersion parameters. When these reasons prevail, and an approximate genetic model is fit to the data, the dispersion matrix is misspecified.

Henderson (1975) is one of the first references in the animal breeding literature to dealt with the effects of misspecification of the dispersion matrix of random effects on prediction error variance of BV (PEV(BV)). More recently, Van Vleck (1993) studied the effect of ignoring inbreeding on PEV(BV). Also, Cantet and Fernando (1995) considered the effects of using homogeneous additive variance on PEV(BV) when the true specification requires heterogeneous additive variances. The effects of misspecification on bias and PEV of predicted BV can be objectively assesed by means of stochastic simulation. de Boer and van Arendonk (1992) and Johansson *et al.* (1993) studied the effects of misspecifying the variance of dominance effects when inbreeding develops, on bias and PEV. In unpublished research conducted by the author, genetic response to simulated selection by BLUP including or excluding MQTL effects were compared. In a different study, Ruane and Colleau (1995) compared selection by BLUP of total genetic merit taking into account or ignoring the variance of MQTL effects. As stochastic simulations usually require a large amount of computing efforts and time to run, it is important to have a thorough understanding of the effects of misspecification errors in the dispersion matrix of random effects. Therefore, the objective of this presentation is to show *analytically* that misspecification of the covariance matrix of random effects may lead to contradicting conclusions with regard to PEV(BV) and accuracy when either: 1) the complete model is assumed to be true, or 2) the complete and the approximate model are considered to be correct.

1. Effects of ignoring a second set of random effects on PEV(BV) and accuracy.

The situation was first considered by Henderson (1975). We first discuss two nested models without any reference of which one is the "true" model. After that, the expressions of PEV(BV) in both models are compared analytically. Finally, we consider the larger model to be true and study the consequences of fitting the smaller model on PEV. Using the terminology of Goldberger (1991), we call the model with both sets of random effects, the "long" model. Whereas the model without the second random effect will be called the "short" model, rather than the "reduced model". This is to avoid any confusion with the "reduced animal model" of Quaas and Pollak (1980).

1.1. The long model.

Consider the mixed linear model:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Za} + \boldsymbol{Wu} + \boldsymbol{e} \qquad [1]$$

where $\boldsymbol{y}$ is the n x 1 vector of records; β is a p x 1 vector of location parameters; $\boldsymbol{a}$ is the qa x 1 random vector of BV; $\boldsymbol{u}$ is a q_u x 1 random vector that may or may not be uncorrelated with $\boldsymbol{a}$; and $\boldsymbol{e}$ is an n x 1 vector of random residuals uncorrelated to $\boldsymbol{a}$ and $\boldsymbol{u}$. The vectors $\boldsymbol{\beta}$, $\boldsymbol{a}$ and $\boldsymbol{u}$ are related to $\boldsymbol{y}$ through known incidence matrices $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\boldsymbol{W}$, respectively. Matrix $\boldsymbol{X}$ is parameterized to full rank p. Expected values of random vectors are:

$$\mathrm{E}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{a}\\ \boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X}\beta\\ \boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix} \qquad [2]$$

and their covariances are:

$$\mathrm{Var}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{a}\\ \boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{V} & \boldsymbol{ZG_a}+\boldsymbol{WG_{ua}} & \boldsymbol{ZG_{au}}+\boldsymbol{WG_u} & \boldsymbol{R}\\ \boldsymbol{G_aZ'}+\boldsymbol{G_{au}W'} & \boldsymbol{G_a} & \boldsymbol{G_{au}} & \boldsymbol{0}\\ \boldsymbol{G_uW'}+\boldsymbol{G_{ua}Z'} & \boldsymbol{G_{ua}} & \boldsymbol{G_u} & \boldsymbol{0}\\ \boldsymbol{R} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}\end{bmatrix} \qquad [3]$$

where

$$\boldsymbol{V} = \boldsymbol{ZG_aZ'} + \boldsymbol{ZG_{au}W'} + \boldsymbol{WG_{ua}Z'} + \boldsymbol{WG_uW'} + \boldsymbol{R} \qquad [4]$$

It will be assumed that the covariance matrices $\boldsymbol{G_a}$, $\boldsymbol{G_u}$ and $\boldsymbol{G_{au}}$ are function of the scalars σ^2_A, σ^2_u and σ_{au}, respectively. If these parameters are unknown, the analysis is conditional on the estimated values: $\sigma^2_A = \hat{\sigma}^2_A$, $\sigma^2_u = \hat{\sigma}^2_u$, $\hat{\sigma}_{au} = \sigma_{au}$ and $\sigma^2_e = \hat{\sigma}^2_e$, the latter being the residual variance. An example of [1] is the model with maternal effects (Willham, 1963). When $\sigma_{au} = 0$, $\boldsymbol{u}$ may be MQTL effects (Fernando and Grossman, 1989) or dominance effects in absence of inbreeding (Harris, 1964; de Boer and van Arendonk, 1992; Johansson *et al.*, 1993). In general, for the additive and dominance model $q_a = q$, the number of animals, and $q_u = q_a = q$. For the additive model with one MQTL $q_a = q$, and $q_u = 2q_a = 2q$.

Mixed model equations (Henderson, 1984) for the long model are:

$$\begin{bmatrix}\boldsymbol{Z'HZ} + \boldsymbol{G_{11}} & \boldsymbol{Z'HW} + \boldsymbol{G_{12}}\\ \boldsymbol{W'HZ} + \boldsymbol{G_{21}} & \boldsymbol{W'HW} + \boldsymbol{G_{22}}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{a}}\\ \hat{\boldsymbol{u}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{Z'Hy}\\ \boldsymbol{W'Hy}\end{bmatrix} \qquad [5]$$

with:

$$\boldsymbol{H} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\,\boldsymbol{R}^{-1} \qquad [6]$$

and

$$\begin{bmatrix} \boldsymbol{G_a} & \boldsymbol{G_{au}} \\ \boldsymbol{G_{ua}} & \boldsymbol{G_u} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{G_{11}} & \boldsymbol{G_{12}} \\ \boldsymbol{G_{21}} & \boldsymbol{G_{22}} \end{bmatrix} \qquad [7]$$

Accuracy of BV for animal *i* is:

$$\mathrm{r_i} = \sqrt{1 - \frac{C_{ii}^{aa}}{\sigma_A^2}} \qquad [8]$$

where $\boldsymbol{C}^{aa}_{ii}$ is the diagonal element in the inverse of the coefficient matrix in [5] corresponding to the BV of animal *i*, and it is equal to PEV(BV) for BLUP(a_i). On letting:

$$\begin{bmatrix} \boldsymbol{Z'HZ} + \boldsymbol{G_{11}} & \boldsymbol{Z'HW} + \boldsymbol{G_{12}} \\ \boldsymbol{W'HZ} + \boldsymbol{G_{21}} & \boldsymbol{W'HW} + \boldsymbol{G_{22}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{bmatrix} \qquad [9]$$

and using standard results on the inverse of partitioned matrices, PEV($\boldsymbol{a}$) = $\boldsymbol{C}^{aa}$ (Henderson, 1984) in the long model can be expressed as:

$$\boldsymbol{C^{aa}} = \left(\boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\, \boldsymbol{\Sigma_{22}^{-1}} \boldsymbol{\Sigma_{21}} \right)^{-1} \qquad [10]$$

1.2. An equivalent model.

An equivalent model (Henderson, 1985) to the long one is:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{\epsilon} \qquad [11]$$

with the same expected value of [1]:

$$\mathrm{E}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{a} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X\beta} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix} \qquad [12]$$

and variance-covariance matrix equal to:

$$\mathrm{Var}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{a} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \boldsymbol{V_r} & \boldsymbol{ZG_a} + \boldsymbol{WG_{ua}} & \boldsymbol{R_r} \\ \boldsymbol{G_aZ'} + \boldsymbol{G_{au}W'} & \boldsymbol{G_a} & \boldsymbol{ZG_{au}W'} \\ \boldsymbol{R_r} & \boldsymbol{WG_{ua}Z'} & \boldsymbol{R_r} \end{bmatrix} \qquad [13]$$

where:

$$\boldsymbol{V_r} = \boldsymbol{ZG_aZ'} + \boldsymbol{R_r} \qquad \boldsymbol{R_r} = \boldsymbol{WG_uW'} + \boldsymbol{R} \tag{14}$$

In this model the residuals are correlated to ***a***, so that MME have to be solved by procedures suggested by Schaeffer and Henderson (1983). Additional difficulties for setting up MME are related to the need of inverting $\boldsymbol{R_r}$, and to perform the associated matrix multiplications.

1.3. The short model.

The short model is not equivalent to the long one, and is the result of ignoring ***u*** in [1] with:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{e}^* \tag{15}$$

The expected value of [15] is:

$$\mathrm{E}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{a}\\ \boldsymbol{e}^*\end{bmatrix} = \begin{bmatrix}\boldsymbol{X\beta}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix} \tag{16}$$

and the variance-covariance matrix is:

$$\mathrm{Var}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{a}\\ \boldsymbol{e}^*\end{bmatrix} = \begin{bmatrix}\boldsymbol{V_I} & \boldsymbol{ZG_a} & \boldsymbol{R}\\ \boldsymbol{G_aZ'} & \boldsymbol{G_a} & \boldsymbol{0}\\ \boldsymbol{R} & \boldsymbol{0} & \boldsymbol{R}\end{bmatrix} \tag{17}$$

The dispersion matrix of the data is now equal to:

$$\boldsymbol{V_I} = \boldsymbol{ZG_aZ'} + \boldsymbol{R} \tag{18}$$

Let $\boldsymbol{a_I}$ be the prediction of ***a*** computed from the short model. Then, MME to obtain BLUP($\boldsymbol{a_I}$) are:

$$(\boldsymbol{Z'HZ} + \boldsymbol{G_a}^{-1})\,\hat{\boldsymbol{a}} = \boldsymbol{Z'Hy} \tag{19}$$

which leads to the following expression for PEV($\boldsymbol{a_I}$):

$$\boldsymbol{C_I^{aa}} = (\boldsymbol{Z'HZ} + \boldsymbol{G_a}^{-1})^{-1} = \boldsymbol{\Sigma_{11}^{-1}} \tag{20}$$

The last equality holds after definition [9].

1.4. Differences in PEV(*a*) between the long and short models.

It will be shown in this section that accuracy of selection for BV in the long model (r), is less or equal to accuracy of BV in the short model (r_I), *i.e.* in a model where ***u*** is ignored. Begin by observing that:

$$\boldsymbol{C}^{aa} - \boldsymbol{C}_I^{aa} = \left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right)^{-1} - \boldsymbol{\Sigma}_{11}^{-1} \qquad [21]$$

$$= \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \left(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \right)^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} - \boldsymbol{\Sigma}_{11}^{-1} \qquad [22]$$

$$= \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \left(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \right)^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \qquad [23]$$

$$= \boldsymbol{C}_I^{aa} \, (\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{W} + \boldsymbol{G}_{21}) \, \boldsymbol{C}^{uu} \, (\boldsymbol{W}'\boldsymbol{H}\boldsymbol{Z} + \boldsymbol{G}_{12}) \, \boldsymbol{C}_I^{aa} \qquad [24]$$

Now, as $\boldsymbol{C}^{uu}$ is a positive definite variance-covariance matrix (Henderson, 1984), expression [24] is non negative definite by theorem A.9 on Toutenburg (1982; p.183). If $\boldsymbol{C}^{aa}{}_I\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{W}+\boldsymbol{G}_{21}$ is square and non-singular, then [24] is positive definite. However, the rank of a product of matrices is less or equal than the minimum of the ranks of the two matrices (Seber, 1977; A.2.1. in p.383), so that:

$$\mathrm{r}\ [\boldsymbol{C}_I^{aa}\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{W}] \le \text{Minimum} \left(\mathrm{r}[\boldsymbol{C}_I^{aa}],\ \mathrm{r}[\ \boldsymbol{Z}'\boldsymbol{H}\boldsymbol{W}\ ] \right) = \text{Minimum}\ (\mathrm{p},\mathrm{q}) \qquad [25]$$

In general in animal models p<<q, i.e. the number of linearly independent fixed effects is smaller than the number of animals. In case p≥q, the matrix in [24] is positive definite.

Theobald (1974) showed that a difference of mean square error matrices being non-negative definite, is equivalent to differences of mean square errors of individual elements being non-negative. Hence, PEV($\boldsymbol{a}$) in the true model is greater or equal to PEV($\boldsymbol{a}_I$), *i.e.* $\boldsymbol{C}^{aa}{}_{ii} - \boldsymbol{C}^{aa}{}_{Iii} \ge 0$. Thus, on squaring r_i in [8] and solving for $\boldsymbol{C}^{aa}{}_{ii}$ we have:

$$\boldsymbol{C}_{ii}^{aa} \ge \boldsymbol{C}_{iiI}^{aa} \;\; => \;\; 1 - r_i^2 \ge 1 - r_{Ii}^2 \;\; => \;\; r_{Ii}^2 \ge r_i^2 \qquad [26]$$

Therefore, if the short model is fitted and PEV(BV) is computed as if this is the correct model, accuracy is always greater or equal to the one computed by fitting the long model, and conditionally on this being the correct one. The result merely says that PEV($\boldsymbol{a}$) in the long model is larger, or at most equal, due to the additonal uncertainty of having to predict more random variables. However, comparison of fitted (or operational) models should be done under the pretense of just one model being true (Seber, 1977; Goldberger, 1991). Now, suppose the long model is correct. Then, PEV computed under the short model (PEV($\boldsymbol{a}_I$)) is no longer [20], but can be computed as:

$$\mathrm{Var}(\hat{\boldsymbol{a}}_I - \boldsymbol{a}) = \mathrm{Var}(\hat{\boldsymbol{a}}_I) + \mathrm{Var}(\boldsymbol{a}) + \mathrm{cov}(\hat{\boldsymbol{a}}_I, \boldsymbol{a}') + \mathrm{cov}(\boldsymbol{a}, \hat{\boldsymbol{a}}_I') \qquad [27]$$

And, as BLUP($\boldsymbol{a}_I$) = $\hat{\boldsymbol{a}}_I$ is:

$$\hat{\boldsymbol{a}}_I = \boldsymbol{C}_I^{aa} \, \boldsymbol{Z}'\boldsymbol{H}\boldsymbol{y} \qquad [28]$$

we can obtain Var($\hat{\boldsymbol{a}}_I$) as:

$$\mathrm{Var}\ (\boldsymbol{\hat{a}_I}) = \boldsymbol{C_I^{aa} Z' H V H Z\ C_I^{aa}}$$

Also, cov($\boldsymbol{\hat{a}_I}$,$\boldsymbol{a}$) is:

$$\mathrm{cov}\ (\boldsymbol{\hat{a}_I},\boldsymbol{a}) = \boldsymbol{C_I^{aa} Z' H}\ \mathrm{cov}[\boldsymbol{y},(\boldsymbol{Za} + \boldsymbol{Wu})'] = \boldsymbol{C_I^{aa} Z' H (G_a Z' + G_{au} W')}$$

Note that Var($\boldsymbol{\hat{a}_I}$) is not equal to cov($\boldsymbol{\hat{a}_I}$,$\boldsymbol{a}$), which is the case when the short model is both the operational and the true model (Henderson, 1984). Finally, on using [29] and [30], PEV($\boldsymbol{a_I}$) is:

$$\mathrm{Var}(\boldsymbol{\hat{a}_I} - \boldsymbol{a}) = \boldsymbol{C_I^{aa} Z' H V H Z C_I^{a}} + \boldsymbol{G} - \boldsymbol{C_I^{aa} Z' H (Z G_a + W G_{au})} - \boldsymbol{(G_a Z' + G_{ua} W') H Z C_I^{aa}} \qquad [31]$$

As this PEV was derived under the assumption of an incorrect specification of the covariance matrix of the data, the property of being "minimum variance in the class of linear unbiased predictors" (Henderson, 1984) of [5] ensures us that the difference between [31] minus [10] (=$\boldsymbol{C^{aa}_I}$ - $\boldsymbol{C^{aa}}$) is non-negative definite. We illustrate the principle using a small example with 4 animals: a sire and a dam in the same contemporary group, and two full-sibs out of the two parents in another contemporary group. Suppose the long model includes the contemporary group efect as fixed, and the BV, the dominant deviations and the residuals as random effects. Further suppose that $\sigma^2_A = 0.25$, $\sigma^2_u = 0.25$ and $\sigma^2_e = 1$. Then, PEV(a) of the sire is equal to PEV(a) of the dam, and PEV(BV) for the two full-sibs are also equal to each other. Now, PEV of the sire in the short model is 0.225, in the long model is 0.229 and in the short model, but computed as if the long model is correct by means of [31], 0.830. Corresponding figures for any of the full-sibs are 0.243, 0.244 and 0.952. This is a very small example but serves right to illustrate the point. There is no pretension whatsoever to extrapolate any magnitude among PEVs for larger data sets and more complex models.

2. Effects of misspecification of the dispersion matrix of breeding values on prediction error variance.

Henderson (1975) obtained the expression for $\boldsymbol{C^{aa}}$ when the covariance matrix of BV is misspecified:

$$\mathrm{Var}(\boldsymbol{\hat{a}} - \boldsymbol{a}) = \boldsymbol{C_I^{aa}} + \boldsymbol{C_I^{aa} G_{aI}^{-1} (G_a - G_{aI}) G_{aI}^{-1} C_I^{aa}} \qquad [32]$$

where $\boldsymbol{G_{aI}}$ is the incorrect covariance matrix of BV, and $\boldsymbol{C^{aa}_I}$ is PEV computed under $\boldsymbol{G_{aI}}$. If the specification of Var($\boldsymbol{a}$) is correct, then $\boldsymbol{G_a}$=$\boldsymbol{G_{aI}}$ and $\boldsymbol{C^{aa}}$=$\boldsymbol{C^{aa}_I}$. By means of a simple numerical example, Henderson (1975) indicated that PEV($\boldsymbol{a}$) computed with $\boldsymbol{G_{aI}}$ (but assuming $\boldsymbol{G_a}$ is correct) is greater than when calculated under the true model. Van Vleck (1993) used this expression to study the effect of ignoring inbreeding in $\boldsymbol{G_a}$ over PEV($\boldsymbol{a}$). Again, by a numerical example he observed that ignoring inbreeding reduces diagonals of $\boldsymbol{G_a}$ and increases PEV($\boldsymbol{a}$).

Cantet and Fernando (1995) discussed the case of using $\boldsymbol{G_{aI}} = \boldsymbol{A}\sigma^2_{\mathrm{A}}$ instead of a matrix ($\boldsymbol{G_a}$) that accounts for heterogeneity of additive variance in predicting $\boldsymbol{a}$ for crossbred individuals originated in two populations under an additive model of inheritance. In this situation, Lo *et al* (1993) showed that $\boldsymbol{G_a}$ can be written as:

$$\boldsymbol{G_a} = (\boldsymbol{I}-\boldsymbol{P})^{-1}\ \boldsymbol{G_\epsilon}\ (\boldsymbol{I}-\boldsymbol{P}')^{-1} \qquad [33]$$

where the matrix $\boldsymbol{P}$ relates progeny to parents. The diagonal matrix $\boldsymbol{G_\varepsilon}$ has diagonal elements as linear functions of the additive variances in populations L (σ^2_{AL}) and I (σ^2_{AI}), say, and of segregation variance (σ^2_{ALI}). This latter parameter arises from differences in gene frequencies between the two populations (Lo *et al*, 1993). Now, let $\boldsymbol{D}$ be diagonal with the i^{th} diagonal element equal to $0.5[1\text{-}0.5(\mathrm{F}_{\mathrm{S}i}+\mathrm{F}_{\mathrm{D}i}]$, if the father (S$i$) and the mother (D$i$) of i are known, and $\mathrm{F}_{\mathrm{S}i}$ is the inbreeding coefficient of Si. Also, $\boldsymbol{D}_{ii} = 0.25(3\text{-}\mathrm{F}_{\mathrm{S}i})$, if only the sire of i is known, and $\boldsymbol{D}_{ii} = 0.25(3\text{-}\mathrm{F}_{\mathrm{D}i})$ if only the dam of i is known. Finally, if both parents of i are unknown $\boldsymbol{D}_{ii} = 1$. With this definition of $\boldsymbol{D}$ and after some algebra, [32] becomes equal to:

$$\mathrm{Var}(\hat{\boldsymbol{a}}-\boldsymbol{a}) = \boldsymbol{C_I^{aa}} + \boldsymbol{C_I^{aa}}(\boldsymbol{I}-\boldsymbol{P}')\ \boldsymbol{D}^{-1}\ (\boldsymbol{G_\epsilon}-\boldsymbol{D})\ \boldsymbol{D}^{-1}\ (\boldsymbol{I}-\boldsymbol{P})\ \boldsymbol{C_I^{aa}} \qquad [34]$$

Therefore, PEV($\boldsymbol{a}$) will be incorrectly estimated by the second term on the right of [34]. As this term depends on the structure of $\boldsymbol{G_\varepsilon}$ and $\boldsymbol{D}$ no general result can be given. However, if $\boldsymbol{C_I^{aa}}(\boldsymbol{I}\text{-}\boldsymbol{P}')\boldsymbol{D}^{-1}(\boldsymbol{G_\varepsilon}\text{-}\boldsymbol{D})\boldsymbol{D}^{-1}(\boldsymbol{I}\text{-}\boldsymbol{P}')\boldsymbol{C_I^{aa}}$ is positive definite, it adds up to $\boldsymbol{C_I^{aa}}$ and true PEV is underestimated. This happens if both ($\boldsymbol{G_\varepsilon}\text{-}\boldsymbol{D}$) and $\boldsymbol{C_I^{aa}}(\boldsymbol{I}\text{-}\boldsymbol{P}')\boldsymbol{D}^{-1}$ are positive definite (Toutenburg, 1982; theorem A.9, page 183). Then, $[\boldsymbol{C_I^{aa}}(\boldsymbol{I}\text{-}\boldsymbol{P}')\boldsymbol{D}^{-1}]^{-1} = \boldsymbol{D}(\boldsymbol{I}\text{-}\boldsymbol{P}')^{-1}(\boldsymbol{C_I^{aa}})^{-1}$ and $\boldsymbol{C_I^{aa}}(\boldsymbol{I}\text{-}\boldsymbol{P}')\boldsymbol{D}^{-1}$ is positive definite. Finally, if ($\boldsymbol{G_\varepsilon}\text{-}\boldsymbol{D}$) is positive definite its diagonal elements are positive (Seber, 1977; page 388), which in turn happens when the diagonal elements of $\boldsymbol{G_\varepsilon}$ are strictly greater than corresponding elements of $\boldsymbol{D}$. For example, this may happen whenever σ^2_{ALI} contributes to genetic variance (such is the case for backcrosses, F2, etc), and this parameter is ignored. Under these conditions PEV($\boldsymbol{a}$) will be underestimated, and the amount of underestimation will depend on the magnitude of σ^2_{ALI}.

Now, suppose on the contrary that PEV($\boldsymbol{a}$) is compared to PEV($\boldsymbol{a_I}$) when both models are considered to be correct. Take for example a model with MQTL effects (Fernando and Grossman, 1989). Suppose that BV and MQTL effects are to be predicted together in $\boldsymbol{t}$, say, such that:

$$\boldsymbol{t} = \boldsymbol{a} + \boldsymbol{Wv} \qquad \mathrm{Var}(\boldsymbol{t}) = \boldsymbol{G_a} + \boldsymbol{WG_v}\ \boldsymbol{W}' = \boldsymbol{G} \qquad [35]$$

where $\boldsymbol{W} = \boldsymbol{I_{qa}} \otimes [1\ 1]$, and $\boldsymbol{G_v}$ is the covariance matrix of the MQTL effects $\boldsymbol{v}$. Van Arendonk *et al* (1994) indicated how to obtain BLUP($\boldsymbol{t}$) by means of an efficient manner of inverting $\boldsymbol{G}$. In the alternative model $\boldsymbol{G}$ is incorrectly specified by taking $\boldsymbol{G}^{*}=\boldsymbol{G_a}$. This is equivalent to predicting $\boldsymbol{t}$ with its variance matrix being $\boldsymbol{G_a}$ instead of $\boldsymbol{G}$, although PEV($\boldsymbol{a_I}$) is computed as if the specification of Var($\boldsymbol{t}$) is correct. With the model assumptions PEV's in both models are:

$$\mathrm{PEV}(\boldsymbol{a}) = (\boldsymbol{Z}'\boldsymbol{HZ} + \boldsymbol{G}^{-1})^{-1} = \boldsymbol{C}^{aa} \qquad \mathrm{PEV}(\boldsymbol{a}_I) = (\boldsymbol{Z}'\boldsymbol{HZ} + \boldsymbol{G}_a^{-1})^{-1} = \boldsymbol{C}_I^{aa}$$

Now we follow the reasoning of Toutenburg (1982; page 52) to compare mean square errors of an alternative estimator of $\boldsymbol{\beta}$ *vs.* the ordinary least squares estimator, in a fixed effects model. In doing so, we obtain the difference of the inverses of PEV's, which is equal to:

$$(\boldsymbol{C}_I^{aa})^{-1} - (\boldsymbol{C}^{aa})^{-1} = \boldsymbol{G}_a^{-1} - \boldsymbol{G}^{-1} = \boldsymbol{G}_a^{-1}\boldsymbol{W}\,(\boldsymbol{W}'\boldsymbol{G}_a^{-1}\boldsymbol{W} + \boldsymbol{G}_v^{-1})^{-1}\boldsymbol{W}'\boldsymbol{G}_a^{-1} \qquad [37]$$

and is non-negative definite by a similar argument to the one that led to [24]. Therefore, PEV($\boldsymbol{a}$)-PEV($\boldsymbol{a}_I$) is non-negative definite, which means that in the model ignoring $\boldsymbol{G}_v$ (i.e. misspecifying the covariance matrix of random effects), accuracy is always greater or equal than in the model in which $\boldsymbol{G}$ is correctly specified, and when both models are considered to be correct.

An identical situation occurs in case $\boldsymbol{G}_{aI} = \boldsymbol{I}\sigma^2_{\mathrm{A}}$ and $\boldsymbol{G}_a = \boldsymbol{A}\sigma^2_{\mathrm{A}}$, *i.e.* when additive relationships are ignored. As in [37] form the difference between the inverses of PEVs (incorrect minus correct) to obtain:

$$(\boldsymbol{C}_I^{aa})^{-1} - (\boldsymbol{C}^{aa})^{-1} = (\boldsymbol{I} - \boldsymbol{A}^{-1})\left(\frac{\sigma_e^2}{\sigma_A^2}\right) = \boldsymbol{A}^{-1}\,(\boldsymbol{A} - \boldsymbol{I})\left(\frac{\sigma_e^2}{\sigma_A^2}\right) \qquad [38]$$

Consider the last equality in [38]. Unless identical genotypes are in $\boldsymbol{a}$, $\boldsymbol{A}$ is positive definite and all its off-diagonal elements are the additive relationships between individuals, which are greater or equal to 0. Thus, the difference ($\boldsymbol{A}$-$\boldsymbol{I}$) is non-negative definite, and so is $(\boldsymbol{C}^{aa}{}_I)^{-1}$-$(\boldsymbol{C}^{aa})^{-1}$. This in turn means that $\boldsymbol{C}^{aa}$-$\boldsymbol{C}^{aa}{}_I$ is non-negative definite, *i.e.* PEV($\boldsymbol{a}$) is always greater or equal than PEV($\boldsymbol{a}_I$). As a consequence, accuracy calculated ignoring relationships is greater or equal than accuracy computed including $\boldsymbol{A}$. By means of numerical calculations, Wilmink and Dommerholt (1985) observed that ignoring additive relationships overestimated accuracy in a sire model.

Discussion

A motivation for this study was unpublished work by the author. A stochastic simulation was performed to compare selection response from a long model with BV plus MQTL effects, with a short model in which MQTL effects were ignored. Results showed that genetic response to selection for predicted BV plus predicted MQTL effects slightly outperformed selection on predicted BV alone. However, the genetic mean of BV to selection by BLUP(BV) ignoring MQTL was greater than the response calculated when selection was on BLUP(BV) + BLUP(MQTL). This result that at first may look counterituitive as more information is used to predict genetic merit, can be partially explained by differences in accuracy of prediction. As shown here, if MQTL effect are not included in the model, PEV(BV) ignoring MQTL and computed as if the short model is true, is always greater or equal than PEV(BV) in models including MQTL effects. It follows that accuracy of selection is decreased by the increase in PEV(BV). If intensity of selection remains constant, which may not be true since selection candidates have a correlated structure of their BLUP(BV) (Hill, 1976; Meuwissen, 1991), then selection response for BV alone decreases. This fact should be taken into account while analyzing the results of a simulation. For example, a comparison of selection on BV + MQTL *vs.* selection on BV, when both predictions are calculated using the long model, is a comparison at equal PEV(BV) or accuracy, and at probably similar selection intensity. Henderson (1988) showed that PEV(BV) calculated from equivalent models are equal. Therefore, comparisons of selection based on BLUP(BV) computed under equivalent models (such as [1] and [11]) are at the same PEV(BV). However, fitting model [11] may be difficult in practice, and so may be the interpretation of the results of such a comparison.

Although we concentrated here on the effects of excluding a random effect in the model on PEV(BV), a similar problem occurs while estimating mean parameters in a model where one or more random effects are excluded. Take for example the case of estimating crossbreeding mean parameters. Komender and Hoeschele (1989) reported the results of a simulation to compare ordinary (OLS) *vs.* generalized least squares (GLS) estimators of a vector $\boldsymbol{\beta}$, which included crossbreeding parameters, when the animals were related. This was equivalent to using $\boldsymbol{V} = \boldsymbol{I}\sigma 2\text{e}$ instead of $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{G}_a\boldsymbol{Z}' + \boldsymbol{I}\sigma^2_e$. Although the authors warned on the use of OLS to estimate $\boldsymbol{\beta}$, they reported higher standard errors for the GLS estimates as compared with the OLS estimates. This may be due to using $\text{Var}(\boldsymbol{\beta}^{\text{OLS}}) = (\boldsymbol{X'X})^{-1}\,\sigma^2_e$ rather than the proper variance of OLS with this misspecification (Seber, 1977, page 144):

$$\text{Var}\ (\boldsymbol{\beta}^{\boldsymbol{OLS}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\ \boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Misspecification of $\boldsymbol{G}_a$ has a similar effect on PEV(BV) when both the correct and the incorrect models are taken to be true. In this case it is hard to envision any other fair comparison than the one including the model with the correct $\boldsymbol{G}_a$ *vs.* an equivalent model which uses $\boldsymbol{G}_{aI}$. The latter model should include the extra variation not accounted for $\boldsymbol{G}_{aI}$ into the residual matrix $\boldsymbol{R}$. Fitting such a model may not be difficult in case that $\boldsymbol{R}$ is a diagonal matrix. Note that this may not be the situation in the examples discussed by Van Vleck (1993) and Cantet and Fernando (1995), as off-diagonal elements (additive relationships) to the right of the diagonal of $\boldsymbol{G}_a$ calculated with either homogeneous or heterogeneous additive genetic variance, depend on the diagonal elements which are incorrectly specified in $\boldsymbol{G}_{aI}$.

Misspecification of $\boldsymbol{G}_a$ is also related to selection bias. It has been shown that if all data employed to make selection decisions are available, then BLUP of BV can be computed ignoring selection (Goffinet, 1983; Fernando and Gianola, 1990). This result only holds when the correct covariance matrix of ***a*** is used to compute BLUP. An example is the result of Johansson *et al.* (1993) who obtained an increased trend for prediction error bias *i.e.* E[BLUP(***a***)-***a***], as well as dominance deviations after 10 generations of selection, when the dispersion matrix of dominance effects was initially misspecified. With selection inbreeding developes, which in turn generates an additional covariance between additive and dominance effects plus two other parameters besides σ^2_A and σ^2_u (Harris, 1964). The rise of these parameters constitutes a dynamic specification. For example, the initial value of σ_{au} is 0, but later takes on a non-zero value. It is worth adding that we intentionally began by specifying E(***a***) = ***0*** (see [2]), and not E(***a***) = ***Qg***. In the latter ***g*** is a vector of means of BV called genetic groups (Westell *et al.*, 1988; Quaas, 1988), and ***Q*** is the matrix that relates the BV to genetic groups. These parameters are used to account for selection bias due to genetic trend and loss of relationships among animals. Therefore, models with genetic groups implicitly assume that $\boldsymbol{G_a}$ is incorrectly specified.

Literature

Cantet,R.J.C.; Fernando,R.L.: 1995. Prediction of breeding values with additive animal models for crosses from two populations. Genet.Sel.Evol. 32:323-334.

de Boer,I.J.M.; van Arendonk,J.A.M.: 1992. Prediction of additive and dominance effects in selected or unselected populations with inbreeding. Theor.Appl.Genet. 84:451-459.

Falconer,D.S.: 1981. *Introduction to quantitative genetics*. 2nd. Ed. Longman, London.

Fernando,R.L.; Grossman,M.: 1989. Marker assisted selection using best linear unbiased prediction. Genet.Sel.Evol. 21:467-477.

Fernando,R.L.; Gianola,D.: 1990. Statistical inferences in populations undergoing selection or non-random mating. in: *Advances in statistical methods for genetic improvement of livestock.* Gianola,D., K.Hammond, ed. Springer-Verlag, New York.

Goffinet,B.: 1983. Selection on selected records. Genet. Sel.Evol. 15:91-98.

Goldberger,A.S. 1991.: *A course in econometrics.* Harvard University Press, Cambridge.

Harris,D.L.1964.:Genotypic covariances between inbred relatives.Genetics 50:1319-1348.

Henderson,C.R.: 1975. Comparison of alternative sire evaluation methods. J.Anim.Sci. 41:760-770.

Henderson,C.R.: 1984. *Application of Linear Models in Animal Breeding.* University of Guelph Press. 2nd. printing. Guelph.

Henderson,C.R.: 1985. Equivalent linear models to reduce computations. J.Dairy Sci. 68:2267-2277.

Henderson,C.R.: 1988. Exact prediction error variances for full model computed from reduced model. J.Anim.Sci. 41:760-770.

Hill,W.G.: 1976. Order statistics of correlated variables and implications in genetic selection programs. Biometrics 32,889-902.

Johansson,K.; Kennedy,B.W.; Quinton,M.: 1993. Prediction of breeding values and dominance effects from mixed models with approximations of the dominance relationship matrix. Livest.Prod. Sci. 34:213-223.

Komender,P.; Hoeschele,I.: 1989. use of mixed-model methodology to improve estimation of crossbreeding parameters. Livest.Prod.Sci. 21:101-113.

Lo L.L.; Fernando,R.L.; Grossman,M.G.: 1993. Genotypic covariance between relatives in multibreed populations: Additive model. Theor.Appl.Genet. 87:423-430.

Meuwissen,T.H.E.: 1991. Reduction of selection differentials in finite populations with a nested full-half sib family structure. Biometrics 47:195-203.

Quaas,R.L.: 1988. Additive models with groups and relationships. J.D.Sci. 71:1338-1345.

Quaas,R.L.; Pollak,J.: 1980. Mixed model methodology for farm and ranch beef cattle testing programs. J.Anim.Sci. 51:1277-1287.

Ruane,J.; Colleau,J.J.: 1995. Marker-assisted selection in an outbred population. Second European Workshop on Advanced Biometrical Methods in Animal Breeding. Salzburg, Austria.

Seber,G.A.F.: 1977. *Linear regression analysis.* J.Wiley & Sons, New York.

Schaeffer,L.R.; Henderson,C.R.: 1983. Best linear unbiased prediction when error vector is correlated with other random vectors in the model. Genet.Sel.Evol. 15:395-400.

Theobald,C.M.: 1974. Generalizations of mean square error applied to ridge regression. J. Roy. Stat.Soc. B. 36:103-106.

Toutenburg,H.: 1982. *Prior information in linear models.* J. Wiley & Sons, New York.

Van Arendonk,J.A.M.; Tier,B.; Kinghorn,B.P. 1994. Use of multiple genetic markers in prediction of breeding values. Genetics 137:319-329.

Van Vleck,L.D.: 1993. Variance of prediction error with mixed model equations when relationships are ignored. Theor.Appl.Genet. 85:545-549.

Westell,R.A.; Quaas,R.L.; VanVleck,L.D.: 1988. Genetic groups in an animal model. J.Dairy Sci. 71:1310-1318.

Willham,R.L.: 1963. The covariance between relatives for characters composed of components contributed by related individuals. Biometrics 19:18-27.

Wilmink, J.B.M.; Dommerholt,J.: 1985. Approximate reliability of Best Linear Unbiased Prediction in models with and without relationships. J.Dairy Sci. 68:946-952.