Theor Appl Genet (1993) 85: 545–549

THEORETICAL
AND
APPLIED GENETICS

Variance of prediction error with mixed model equations when relationships are ignored*

L. D. Van Vleck

Roman L. Hruska Meat Animal Research Center, USDA-ARS, A218 Animal Sciences, University of Nebraska, Lincoln, USA

Received August 11, 1992; Accepted September 3, 1992

Summary. Formulas are presented to illustrate the calculation of correct variances of prediction error (PEV) and the correlation between true and predicted values (r_{TI}) when the incorrect variance–covariance matrix for the random effects is used in mixed-model equations (MME). The example with progeny records of highly related and inbred sires showed that PEV were underestimated from the diagonals of the inverse of the coefficient matrix of the MME when sires were assumed unrelated and not inbred and were overestimated when relationships among sires were calculated with Henderson's simple rules for the inverse of the numerator relationship matrix, $\mathbf{A}^{-1}$, which do not consider inbreeding. When Quaas' rules for $\mathbf{A}^{-1}$, which do consider inbreeding, are used, the correct PEV are obtained. In the example, calculations of r_{TI} from the diagonals of the inverse of the coefficient matrix were too large when relationships and inbreeding were ignored and were obviously wrong when the approximation to the numerator relationship matrix, $\mathbf{A}$, was based on the simple rules for calculating $\mathbf{A}^{-1}$. If the correct $\mathbf{A}$ is used in the MME, the calculation of r_{TI} may be incorrect if inbreeding of the evaluated individual is not considered. If inbreeding is known, adjustment for inbreeding is easy for calculation of r_{TI}.

Key words: Mixed models – Prediction error – Numerator relationships – Inbreeding – Palm trees

Introduction

The properties of mixed-model equations (MME) for joint prediction of realized values of random effects such as breeding values and estimation of fixed effects affecting records have been developed and described by Henderson (1975 b, 1984). The mixed-model equations are usually a preferred computing alternative to using generalized least squares to estimate fixed effects and then predicting realized values of random effects from the adjusted records. Incorporation in the model of numerator relationships in the variance–covariance structure of the genetic values of individuals or parents to be genetically evaluated increases the accuracy of evaluation and reduces the prediction error variance (PEV). Accuracy of prediction, defined as the correlation between true and predicted value (r_{TI}), can be calculated from the inverse of the coefficient matrix of the mixed-model equations. The purpose of this note is to describe a method of calculating the true accuracy of prediction when relationships are ignored or are approximated based on the principles outlined by Henderson (1975 a). A simple method of calculating accuracy from mixed-model equations usually ignores inbreeding of evaluated individuals (Van Vleck 1988) and can lead to the incorrect calculation of r_{TI} unless inbreeding is considered. A secondary purpose is to describe the correct calculation in cases when inbreeding coefficients are not zero. The procedures will be illustrated with data from oil palm breeding that stimulated this discussion.

Methods

The usual general mixed model (Henderson 1984) is:

$$y = \mathbf{X}\beta + \mathbf{Z}u + e,$$

where y is the vector of observations, β is the vector of fixed effects associated with y by the matrix $\mathbf{X}$, u is the vector of random effects associated with y by the matrix $\mathbf{Z}$ (for this discussion u will be limited to a vector of additive genetic (breeding) values or to a vector of transmitting abilities which are equiva-

---

* Published as paper no. 9947, Journal Ser, Nebraska Agric Res Div, University of Nebraska, Lincoln, Neb.

lent to one-half of additive genetic values), and $\boldsymbol{e}$ is a vector of random residual effects. The first and second moments are:

$$\mathrm{E}\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\beta} \\ \mathrm{u} \\ \mathrm{e} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\beta \\ \beta \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathrm{V}\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{u} \\ \boldsymbol{e} \end{pmatrix} = \begin{pmatrix} \mathbf{ZGZ'} & \mathbf{ZG} & \mathbf{R} \\ \mathbf{GZ'} & \mathbf{G} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{pmatrix}.$$

The corresponding mixed model equations (MME) are:

$$\begin{pmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z+G^{-1}} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'R^{-1}}\boldsymbol{y} \\ \mathbf{Z'R^{-1}}\boldsymbol{y} \end{pmatrix}.$$

For this discussion the usual assumption for a single trait analysis will be made that $\mathbf{R}=\mathbf{I}\sigma^2$ where σ^2 is the residual variance. The assumption that $\mathbf{X'X}$ is full rank will be made for notational ease, although generalized inverse results can be shown to give the same answers for variances of prediction errors for the random effects. If $\mathbf{R}^{-1}=\mathbf{I}(1/\sigma^2)$ and the right- and left-hand sides of the equations are multiplied by σ^2 the MME are:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\sigma^2\mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'}\boldsymbol{y} \\ \mathbf{Z'}\boldsymbol{y} \end{pmatrix}.$$

The solutions are:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{XX} & \mathbf{C}^{XZ} \\ \mathbf{C}^{ZX} & \mathbf{C}^{ZZ+} \end{pmatrix} \begin{pmatrix} \mathbf{X'}\boldsymbol{y} \\ \mathbf{Z'}\boldsymbol{y} \end{pmatrix} \quad \text{where}$$

$$\begin{pmatrix} \mathbf{C}^{XX} & \mathbf{C}^{XZ} \\ \mathbf{C}^{ZX} & \mathbf{C}^{ZZ+} \end{pmatrix} = \begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\sigma^2\mathbf{G}^{-1} \end{pmatrix}^{-1}.$$

The inverse of a non-full rank coefficient matrix can be any generalized inverse without affecting the PEV.

The properties of the solutions are (Henderson 1975b, 1984):

$\mathrm{E}[\boldsymbol{k}'\hat{\boldsymbol{\beta}}]=\mathbf{k}'\boldsymbol{\beta}$ if $\mathbf{k}'\boldsymbol{\beta}$ is an estimable function,
$\mathrm{E}[\hat{\boldsymbol{u}}]=\mathbf{0}$, and with $\mathrm{V}(\boldsymbol{u})=\mathbf{G}$,
$\mathrm{V}(\hat{\boldsymbol{u}})=\mathbf{G}-\sigma^2\mathbf{C}^{ZZ+}$,
$\mathrm{Cov}(\hat{\boldsymbol{u}},\boldsymbol{u}')=\mathrm{Cov}(\boldsymbol{u},\hat{\boldsymbol{u}}')=\mathbf{G}-\sigma^2\mathbf{C}^{ZZ+}$, and
$\mathrm{V}(\hat{\boldsymbol{u}}-\boldsymbol{u})=\sigma^2\mathbf{C}^{ZZ+}$.

The correspondence with selection index can be used to obtain the $\mathrm{r_{TI}}$. From the selection index, $\mathrm{V}(\hat{\mathrm{u}}_i-\mathrm{u}_i)=(1-\mathrm{r}^2_{\mathrm{TI}})\,\sigma^2_{\mathrm{u}}$ for an individual u_i. Thus, equate $\mathrm{diag}\{\mathbf{C}^{ZZ+}\}_i\,\sigma^2=(1-\mathrm{r}^2_{\mathrm{TI}})\,\sigma^2_{\mathrm{u}}$ and solve to find $\mathrm{r}^2_{\mathrm{TI}}=1-\mathrm{diag}\{\mathbf{C}^{ZZ+}\}_i\,\sigma^2/\sigma^2_{\mathrm{u}})$ (Van Vleck 1988). A possible error occurs when u_i is additive genetic value or transmitting ability when the individual evaluated is inbred. The appropriate σ^2_{u} is $\mathrm{E}[\mathrm{u}^2_i]$. If individual **i** has inbreeding coefficient F_i, then $\mathrm{E}[\mathrm{u}^2_i]=(1+\mathrm{F}_i)\,\sigma^2_{\mathrm{u}}$ where σ^2_{u} is the additive genetic variance for an individual genetic model or σ^2_{u} is one-fourth the additive genetic variance for a sire (transmitting ability) genetic model. Thus:

$$\begin{aligned} \mathrm{r}^2_{\mathrm{TI}} &= 1-\mathrm{diag}\{\mathbf{C}^{ZZ+}\}_i\,(\sigma^2/\mathrm{E}[\mathrm{u}^2_i]) \\ &= 1-\mathrm{diag}\{\mathbf{C}^{ZZ+}\}_i\,[\sigma^2/(1+\mathrm{F}_i)\,\sigma^2_{\mathrm{u}}]. \end{aligned}$$

If inbreeding is not zero and is ignored in the calculation of $\mathrm{r}^2_{\mathrm{TI}}$, the error in calculation of $\mathrm{r}^2_{\mathrm{TI}}$ can be substantial, i.e., when a simple but wrongly applied method is used.

Another way to calculate $\mathrm{r}^2_{\mathrm{TI}}$ is from the definitional form of a correlation coefficient. Then,

$$\begin{aligned} \mathrm{r}^2_{\mathrm{TI}} &= \mathrm{diag}\{\mathrm{E}[\hat{\boldsymbol{u}}\boldsymbol{u}']\}^2_i/\mathrm{diag}\{\mathrm{E}[\hat{\boldsymbol{u}}\hat{\boldsymbol{u}}']\}_i\,\mathrm{diag}\{\boldsymbol{u}\,\boldsymbol{u}'\}_i \\ &= \mathrm{diag}\{\mathbf{G}-\sigma^2\mathbf{C}^{ZZ+}\}^2_i/\mathrm{diag}\{\mathbf{G}-\sigma^2\mathbf{C}^{ZZ+}\}_i(1+\mathrm{F}_i)\,\sigma^2_{\mathrm{u}} \\ &= [(1+\mathrm{F}_i)\,\sigma^2_{\mathrm{u}}-\mathrm{diag}\{\mathbf{C}^{ZZ+}\}_i\,\sigma^2]/(1+\mathrm{F}_i)\,\sigma^2_{\mathrm{u}} \\ &= 1-\mathrm{diag}\{\mathbf{C}^{ZZ+}\}_i\,[\sigma^2/(1+\mathrm{F}_i)\,\sigma^2_{\mathrm{u}}], \end{aligned}$$

where

$$\mathrm{diag}\{\mathbf{G}\}_i=\mathrm{diag}\{\mathrm{E}[\boldsymbol{u}\,\boldsymbol{u}']\}_i=(1+\mathrm{F}_i)\,\sigma^2_{\mathrm{u}}.$$

With a transmitting ability model to accommodate records of half-sib progeny, a common approximation is to ignore numerator relationships among the sires and inbreeding of the sires. Another approximation sometimes used in animal breeding is to calculate the inverse of the numerator relationship matrix with the simple rules of Henderson (1976), pretending thereby in applying the rules that the individuals are not inbred, rather than with the rules of Quaas (1976), which consider inbreeding exactly. The MME are set up augmented for individuals that contribute to relationships but do not themselves have records or progeny with records (Henderson 1977). The solutions and PEV from the augmented equations are the same as when the inverse of the true relationship matrix associated with individuals with records or with progeny with records is used in the MME. In this discussion equations to describe in general the calculation of PEV and $\mathrm{r}^2_{\mathrm{TI}}$ will be developed assuming the true $\mathbf{G}$ is known. An example will be developed to illustrate the effect of the two approximations described earlier in this paragraph.

Incorrect G *matrix*

Let $\mathbf{D}$ be substituted for $\sigma^2\mathbf{G}^{-1}$ in the mixed-model equations. If $\mathbf{D}=\sigma^2\mathbf{G}^{-1}$, the model and equations will be correct. In the example to follow, calculations for incorrect $\mathbf{D}=\mathbf{I}\lambda$ (male parent model with male parents assumed to be unrelated and not inbred) and for incorrect $\mathbf{D}=\sigma^2\mathrm{A}^{-1}_*$ (A_* is an incorrect numerator relationship matrix) will be illustrated.

The "MME" with $\mathbf{D}$ substituted for $\sigma^2\mathbf{G}^{-1}$ are:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z+D} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^0 \\ \boldsymbol{u}^0 \end{pmatrix} = \begin{pmatrix} \mathbf{X'}\boldsymbol{y} \\ \mathbf{Z'}\boldsymbol{y} \end{pmatrix}.$$

The symbolic solutions are:

$$\begin{pmatrix} \boldsymbol{\beta}^0 \\ \boldsymbol{u}^0 \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{XX}_0 & \mathbf{C}^{XZ}_0 \\ \mathbf{C}^{ZX}_0 & \mathbf{C}^{ZZ+}_0 \end{pmatrix} \begin{pmatrix} \mathbf{X'}\boldsymbol{y} \\ \mathbf{Z'}\boldsymbol{y} \end{pmatrix}.$$

The solutions for $\boldsymbol{u}$, $\boldsymbol{u}^0$, are unbiased as with correct $\mathbf{G}^{-1}$:

$$\begin{aligned} \mathrm{E}[\boldsymbol{u}^0] &= \mathrm{E}[(\mathbf{C}^{ZX}_0\mathbf{X'X}+\mathbf{C}^{ZZ+}_0\mathbf{Z'X})\,\boldsymbol{\beta}+(\mathbf{C}^{ZX}_0\mathbf{X'Z}+(\mathbf{C}^{ZZ+}_0\mathbf{Z'Z})\,\boldsymbol{u}] \\ &= (\mathbf{0})\,\boldsymbol{\beta}+(\mathbf{C}^{ZX}_0\mathbf{X'Z}+\mathbf{C}^{ZZ+}_0\mathbf{Z'Z})\,\mathrm{E}[\boldsymbol{u}]=\mathbf{0}. \end{aligned}$$

Also, the model for $\boldsymbol{u}^0$ is $\mathbf{C}^{ZX}_0\mathbf{X'}\boldsymbol{y}+\mathbf{C}^{ZZ+}_0\mathbf{Z'}\boldsymbol{y}$.

With $\mathrm{E}[\boldsymbol{u}^0]=\mathbf{0}$ then,

$$\mathrm{V}(\boldsymbol{u}^0)=\mathrm{E}[\boldsymbol{u}^0\,\boldsymbol{u}^{0\prime}]=\mathrm{E}[(\mathbf{C}^{ZX}_0\mathbf{X'}+\mathbf{C}^{ZZ+}_0\mathbf{Z'})\,\boldsymbol{y}\boldsymbol{y}'(\mathbf{Z}\mathbf{C}^{ZZ+}_0+\mathbf{X}\mathbf{C}^{XZ}_0)].$$

In the expectation, any products such as $\mathrm{E}[\boldsymbol{\beta}\boldsymbol{u}']$, $\mathrm{E}[\boldsymbol{\beta}\boldsymbol{e}']$, and $\mathrm{E}[\boldsymbol{u}\,\boldsymbol{e}']$ will be zero. Thus, the expectations can be taken separately for the $\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'$, $\mathbf{Z}\boldsymbol{u}\boldsymbol{u}'\mathbf{Z}'$, and $\boldsymbol{e}\,\boldsymbol{e}'$ parts of $\mathrm{E}[\boldsymbol{y}\boldsymbol{y}']$. The part of the expectation with $\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'$ is zero because $\mathbf{C}^{ZX}_0\mathbf{X'X}+\mathbf{C}^{ZZ+}_0\mathbf{Z'X}-\mathbf{0}$:

$$\begin{aligned} &\mathrm{E}[(\mathbf{C}^{ZX}_0\mathbf{X'}+\mathbf{C}^{ZZ+}_0\mathbf{Z'})\,\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'(\mathbf{Z}\mathbf{C}^{ZZ+}_0+\mathbf{X}\mathbf{C}^{XZ}_0)] \\ &\quad =(\mathbf{0})\,\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{0})=\mathbf{0}. \end{aligned}$$

The $\mathbf{Z}\boldsymbol{u}\boldsymbol{u}'\mathbf{Z}'$ part results in:

$$\begin{aligned} &\mathrm{E}[\mathbf{C}^{ZX}_0\mathbf{X'}+\mathbf{C}^{ZZ+}_0\mathbf{Z'})\,\mathbf{Z}\boldsymbol{u}\boldsymbol{u}'\mathbf{Z}'(\mathbf{Z}\mathbf{C}^{ZZ+}_0+\mathbf{X}\mathbf{C}^{XZ}_0)] \\ &\quad =(\mathbf{C}^{ZX}_0\mathbf{X'Z}+\mathbf{C}^{ZZ+}_0\mathbf{Z'Z})\,\mathbf{G}\,(\mathbf{Z'Z}\mathbf{C}^{ZZ+}_0\mathbf{Z'X}\mathbf{C}^{XZ}_0) \\ &\quad =(\mathbf{I}-\mathbf{C}^{ZZ+}_0\mathbf{D})\,\mathbf{G}\,(\mathbf{I}-\mathbf{D}\mathbf{C}^{ZZ+}_0). \end{aligned}$$

The $\boldsymbol{e}\,\boldsymbol{e}'$ part is:

$$\begin{aligned} &\mathrm{E}[(\mathbf{C}^{ZX}_0\mathbf{X'}+\mathbf{C}^{ZZ+}_0\mathbf{Z'})\,\boldsymbol{e}\,\boldsymbol{e}'(\mathbf{Z}\mathbf{C}^{ZZ+}_0+\mathbf{X}\mathbf{C}^{XZ}_0) \\ &\quad =\mathbf{I}\sigma^2[(\mathbf{C}^{ZX}_0\mathbf{X'Z}+\mathbf{C}^{ZZ+}_0\mathbf{Z'Z})\,\mathbf{C}^{ZZ+}_0 \\ &\qquad\quad +(\mathbf{C}^{ZX}_0\mathbf{X'X}+\mathbf{C}^{ZZ+}_0\mathbf{Z'X})\,\mathbf{C}^{XZ}_0] \\ &\quad =\sigma^2[\mathbf{I}-\mathbf{C}^{ZZ+}_0\mathbf{D}]\,\mathbf{C}^{ZZ+}_0+(\mathbf{0})\,\mathbf{C}^{XZ}_0. \end{aligned}$$

Thus,

$$\mathrm{E}[\boldsymbol{u}^0\boldsymbol{u}^{0\prime}] = (\mathbf{I}-\mathbf{C}_0^{ZZ+}\mathbf{D})\,\mathbf{G}\,(\mathbf{I}-\mathbf{D}\,\mathbf{C}_0^{ZZ+}) + \sigma^2(\mathbf{I}-\mathbf{C}_0^{ZZ+}\mathbf{D})\,\mathbf{C}_0^{ZZ+}$$
$$= \mathbf{G}-\mathbf{G}\,\mathbf{D}\,\mathbf{C}_0^{ZZ+}-\mathbf{C}_0^{ZZ+}\mathbf{D}\,\mathbf{G}+\mathbf{C}_0^{ZZ+}\mathbf{D}\mathbf{G}\mathbf{D}\mathbf{C}_0^{ZZ+}$$
$$+\sigma^2\,\mathbf{C}_0^{ZZ+}-\sigma^2\,\mathbf{C}_0^{ZZ+}\mathbf{D}\,\mathbf{C}_0^{ZZ+}.$$

Also needed will be $\mathrm{Cov}(\boldsymbol{u}^0, \boldsymbol{u}')$.

With the model substituted for $\boldsymbol{u}^0$:

$$\mathrm{E}[\boldsymbol{u}^0\boldsymbol{u}'] = \mathrm{E}[(\mathbf{C}_0^{ZX}\mathbf{X}'+\mathbf{C}_0^{ZZ+}\mathbf{Z}')\,\boldsymbol{y}\,\boldsymbol{u}'].$$

The $\mathbf{X}\boldsymbol{\beta}$ part of E [y u] will drop our for two reasons, including $\mathrm{E}[\boldsymbol{\beta}\,\boldsymbol{u}'] = \mathbf{0}$.
The $\boldsymbol{e}\,\boldsymbol{u}'$ part of $\mathrm{E}[\boldsymbol{y}\,\boldsymbol{u}']$ will drop out because $\mathrm{E}[\boldsymbol{e}\,\boldsymbol{u}'] = \mathbf{0}$.
The $\mathbf{Z}\,\boldsymbol{u}\,\boldsymbol{u}'$ part of $\mathrm{E}[\boldsymbol{y}\,\boldsymbol{u}']$ leads to:

$$\mathrm{E}[\boldsymbol{u}^0\boldsymbol{u}'] = (\mathbf{C}_0^{ZX}\mathbf{X}'\mathbf{Z}+\mathbf{C}_0^{ZZ+}\mathbf{Z}'\mathbf{Z})\,\mathbf{G} = (\mathbf{I}-\mathbf{C}_0^{ZZ+}\mathbf{D})\,\mathbf{G}.$$

Thus, prediction error variance, $\mathrm{V}(\boldsymbol{u}^0-\boldsymbol{u}) = \mathrm{V}(\boldsymbol{u})+\mathrm{V}(\boldsymbol{u}^0)-\mathrm{Cov}(\boldsymbol{u}^0,\boldsymbol{u}')-\mathrm{Cov}(\boldsymbol{u},\boldsymbol{u}^{0\prime})$, becomes:

$$\mathrm{V}(\boldsymbol{u}^0-\boldsymbol{u}) = \mathbf{C}_0^{ZZ+}\mathbf{D}\mathbf{G}\mathbf{D}\mathbf{C}_0^{ZZ+}+\sigma^2\,\mathbf{C}_0^{ZZ+}-\sigma^2\,\mathbf{C}_0^{ZZ+}\mathbf{D}\,\mathbf{C}_0^{ZZ+}$$
$$= \sigma^2\,\mathbf{C}_0^{ZZ+}[\mathbf{I}+(1/\sigma^2)\,\mathbf{D}\mathbf{G}\mathbf{D}\mathbf{C}_0^{ZZ+}-\mathbf{D}\,\mathbf{C}_0^{ZZ+}].$$

If $\mathbf{D} = \sigma^2\mathbf{G}^{-1}$ (the correct model) then $\boldsymbol{u}^0 = \hat{\boldsymbol{u}}$ and $\mathrm{V}(\boldsymbol{u}^0-\boldsymbol{u}) = \sigma^2\,\mathbf{C}_0^{ZZ+}$ with $\mathbf{C}_0^{ZZ+} = \mathbf{C}^{ZZ+}$.

For the next two cases, assume that **A** is the true numerator relationship matrix. In practice, A is difficult to obtain for large populations. With $\boldsymbol{u}$ either individual additive genetic values or sire transmitting abilities, $\mathbf{G} = \mathbf{A}\,\sigma_u^2$ and $\sigma^2\,\mathbf{G}^{-1} = \mathbf{A}^{-1}\lambda$ with $\lambda = \sigma^2/\sigma_u^2$.

1) If $\mathbf{D} = \mathbf{I}\,\lambda$ with corresponding $\mathbf{C}_I^{ZZ+}$, $\mathrm{V}(\boldsymbol{u}^0-\boldsymbol{u}) = \sigma^2\,\mathbf{C}_I^{ZZ+}[\mathbf{I}+\lambda(\mathbf{A}-\mathbf{I})\,\mathbf{C}_I^{ZZ+}]$.
2) If $\mathbf{D} = \lambda\,\mathbf{A}_*^{-1}$ (with $\mathbf{A}_*$ an approximate numerator relationship matrix, e.g., $\mathbf{A}_*^{-1}$ from rules ignoring inbreeding) with corresponding $\mathbf{C}_*^{ZZ+}$,

$$\mathrm{V}(\boldsymbol{u}^*-\boldsymbol{u}) = \sigma^2\,\mathbf{C}_*^{ZZ+}[\mathbf{I}+\lambda\,\mathbf{A}_*^{-1}(\mathbf{A}\,\mathbf{A}_*^{-1}-\mathbf{I})\,\mathbf{C}_*^{ZZ+}].$$

Note that if $\mathbf{A}_* = \mathbf{A}$, then $\mathrm{V}(\boldsymbol{u}^*-\boldsymbol{u}) = \sigma^2\,\mathbf{C}_*^{ZZ+}$ with $\mathbf{C}_*^{ZZ+} = \mathbf{C}^{ZZ+}$.

For incorrect **D**, $\mathrm{diag}\{\mathrm{V}(\boldsymbol{u}^0-\boldsymbol{u})\}_i \neq (1-r_{TI}^2)\,\sigma_u^2$. To calculate r_{TI}^2, diagonals of $\mathrm{Cov}(\boldsymbol{u}^0,\boldsymbol{u}')$ and $\mathrm{V}(\boldsymbol{u}^0)$ are needed:

$$r_{TI}^2 = [\mathrm{diag}\{\mathrm{Cov}(\boldsymbol{u}^0,\boldsymbol{u}')\}_i]^2/[\mathrm{diag}\{\mathrm{V}(\boldsymbol{u}^0)\}_i\,\mathrm{diag}\{\mathrm{V}(\boldsymbol{u})\}_i]$$
$$= [\mathrm{diag}\{(\mathbf{I}-\mathbf{C}_0^{ZZ+}\mathbf{D})\,\mathbf{G}\}_i]^2/[\mathrm{diag}\{(\mathbf{I}-\mathbf{C}_0^{ZZ+}\mathbf{D})\,\mathbf{G}\,(\mathbf{I}-\mathbf{D}\,\mathbf{C}_0^{ZZ+})$$
$$+\sigma^2(\mathbf{I}-\mathbf{C}_0^{ZZ+}\mathbf{D})\,\mathbf{C}_0^{ZZ+}\}_i]\,[(1+F_i)\,\sigma_u^2].$$

If $\mathbf{D} = \sigma^2\,\mathbf{G}^{-1}$ (the correct model with $\mathbf{C}_0^{ZZ+} = \mathbf{C}^{ZZ+}$):

$$r_{TI}^2 = [\mathrm{diag}\{\mathbf{G}-\sigma^2\,\mathbf{C}^{ZZ+}\}_i]^2/[\mathrm{diag}\{\mathbf{G}-\sigma^2\,\mathbf{C}^{ZZ+}\}_i]\,[(1+F_i)\,\sigma_u^2]$$
$$= 1-\mathrm{diag}\{\mathbf{C}^{ZZ+}\}_i\,[\sigma^2/(1+F_i)\,\sigma_u^2]$$

as before because when $\mathbf{D} = \sigma^2\,\mathbf{G}^{-1}$, $\mathrm{Cov}(\hat{\boldsymbol{u}},\boldsymbol{u}') = \mathrm{V}(\hat{\boldsymbol{u}})$.

The consequences of incorrect **D** are not apparent from the general expression for r_{TI}^2. Failure, however, to account for inbreeding in $\mathrm{V}(u_i)$ would appear to overestimate r_{TI}^2. An example will show the consequences for a particular analysis.

Example

The data that stimulated this discussion were those of A.C. Soh (personal communication, 1992) from oil palm breeding. The design of the data will be used to demonstrate the effect on PEV of the failure to account for relationships and inbreeding. The actual records are not necessary for this discussion but would be used in predicting breeding values. The model includes two fixed factors with three and two levels, respectively. Data were progeny of nine related parents (denoted here as sires) that were inbred at two different levels. The numerator relationship matrix for the nine sires is:

$$(1/16)\begin{pmatrix} 20 & 14 & 14 & 14 & 13 & 13 & 13 & 13 & 13 \\ 14 & 20 & 14 & 14 & 10 & 10 & 10 & 10 & 10 \\ 14 & 14 & 20 & 14 & 10 & 10 & 10 & 10 & 10 \\ 14 & 14 & 14 & 20 & 10 & 10 & 10 & 10 & 10 \\ 13 & 10 & 10 & 10 & 19 & 12 & 12 & 12 & 12 \\ 13 & 10 & 10 & 10 & 12 & 19 & 12 & 12 & 12 \\ 13 & 10 & 10 & 10 & 12 & 12 & 19 & 12 & 12 \\ 13 & 10 & 10 & 10 & 12 & 12 & 12 & 19 & 12 \\ 13 & 10 & 10 & 10 & 12 & 12 & 12 & 12 & 19 \end{pmatrix}$$

The least squares coefficients in order of the three levels of factor one, two levels of factor two, and nine sire-progeny groups are:

$$\begin{pmatrix} 17 & 0 & 0 & 17 & 0 & 0 & 10 & 0 & 7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 14 & 0 & 14 & 0 & 1 & 1 & 10 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 18 & 4 & 14 & 0 & 4 & 0 & 0 & 3 & 3 & 2 & 3 & 3 \\ 17 & 14 & 4 & 35 & 0 & 1 & 15 & 10 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14 & 0 & 14 & 0 & 0 & 0 & 0 & 3 & 3 & 2 & 3 & 3 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 10 & 1 & 4 & 15 & 0 & 0 & 15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 10 & 0 & 0 & 0 & 10 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 2 & 0 & 9 & 0 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

For the example, $\sigma^2/\sigma_u^2 = \lambda$ will correspond to a sire model with: $\lambda = (1-h^2/4)/(h^2/4) = 19$ for heritability, $h^2 = 0.20$.

The approximate numerator relationship matrix, $\mathbf{A}_*$, was generated using the Henderson (1976) rules to calculate $\mathbf{A}^{-1}$, assuming no inbreeding. The rules involve only each individual and its sire and/or dam. With selfing, the individual, sire, and dam vectors for the example, where X1–X4, and Y1–Y5 were the sires with outbred progeny, are:

Individual	Sire	Dam
A	–	–
T1	A	A
T2	A	A
T3	A	A
Z	A	–
P1	T1	T2
P2	T3	–
X1–X4	P1	P2
Y1–Y5	X1	Z.

Table 1. Apparent and actual variances of prediction error (divided by residual variance) and accuracy of prediction with incorrect assumptions of no relationships among the sires ($\mathbf{D}=\lambda\mathbf{I}$) and with Henderson's simple rules for calculating the inverse of the relationship matrix ignoring inbreeding ($\mathbf{D}=\lambda\mathbf{A}_*^{-1}$) compared with correct relationship matrix ($\mathbf{D}=\lambda\mathbf{A}^{-1}$) for an example from oil palm breeding for $\lambda=19$

Sire	Number of progeny	Assumptions about relationship matrix					
		$\mathbf{D}=\lambda\mathbf{I}$		$\mathbf{D}=\lambda\mathbf{A}_*^{-1}$		$\mathbf{D}=\lambda\mathbf{A}^{-1}$	
		Apparent[a]	Actual[c]	Apparent	Actual	Apparent	Actual
		Calculated prediction error variance/σ^2					
X1	1	0.0502	0.0663	0.0767	0.0655	0.0654	0.0654
X2	15	0.0430	0.0664	0.0745	0.0642	0.0641	0.0641
X3	10	0.0461	0.0667	0.0755	0.0649	0.0648	0.0648
X4	9	0.0417	0.0663	0.0741	0.0640	0.0639	0.0639
Y1, Y2, Y4, Y5	3	0.0470	0.0628	0.0670	0.0613	0.0612	0.0612
Y3	2	0.0484	0.0628	0.0674	0.0616	0.0615	0.0615
		Calculated accuracy, r_{TI}					
X1		0.2149[b]	0.0731	$(-0.4570)^{0.5}$[b]	0.0731	$(-0.2433)^{0.5}$[b]	0.0731[d]
X2		0.4278	0.1595	$(-0.4146)^{0.5}$	0.1598	$(-0.2181)^{0.5}$	0.1598
X3		0.3524	0.1253	$(-0.4347)^{05}$	0.1253	$(-0.2304)^{0.5}$	0.1253
X4		0.4557	0.1707	$(-0.4070)^{0.5}$	0.1709	$(-0.2135)^{0.5}$	0.1709
Y1, Y2, Y4, Y5		0.3271	0.1369	$(-0.2730)^{0.5}$	0.1417	$(-0.1636)^{0.5}$	0.1418
Y3		0.2835	0.1177	$(-0.2840)^{0.5}$	0.1234	$(-0.1694)^{0.5}$	0.1236

[a] Apparent is from diagonal elements of inverse of coefficient matrix when **D** is added to sire by sire block

[b] Apparent is $1-\text{diag}\{\mathbf{C}_0^{ZZ+}\}_i\lambda$ and not $1-\text{diag}\{\mathbf{C}_0^{ZZ+}\}_i\lambda/(1+F_i)$

[c] Actual is calculated from formulas developed from expected values

[d] Actual with true A is $1-\text{diag}\{\mathbf{C}_0^{ZZ+}\}_i\lambda/(1+F_i)$

The $\mathbf{A}_{**}^{-1}$ for the $7+9=16$ individuals was generated according to the Henderson rules and inverted to find the $\mathbf{A}_*$ corresponding to X1 ... X4 Y1 ... Y5 to be used in $\mathbf{D}=\lambda\mathbf{A}_*^{-1}$.

The calculations for $\mathbf{A}_*$ resulted in a matrix with more apparent inbreeding of the nine sires than the correct numerator relationship matrix, **A**:

$$\begin{bmatrix}
1.4688 & 0.9687 & 0.9687 & 0.9687 & 0.9219 & 0.9219 & 0.9219 & 0.9219 & 0.9219 \\
 & 1.4688 & 0.9687 & 0.9687 & 0.6719 & 0.6719 & 0.6719 & 0.6719 & 0.6719 \\
 & & 1.4688 & 0.9687 & 0.6719 & 0.6719 & 0.6719 & 0.6719 & 0.6719 \\
 & & & 1.4688 & 0.6719 & 0.6719 & 0.6719 & 0.6719 & 0.6719 \\
 & & & & 1.3047 & 0.8047 & 0.8047 & 0.8047 & 0.8047 \\
 & & & & & 1.3047 & 0.8047 & 0.8047 & 0.8047 \\
 & & & & & & 1.3047 & 0.8047 & 0.8047 \\
 & \text{symmetric} & & & & & & 1.3047 & 0.8047 \\
 & & & & & & & & 1.3047
\end{bmatrix}$$

Golden et al. (1991) previously pointed out that inverting $\mathbf{A}_*^{-1}$ computed with the no inbreeding approximation may not yield $\mathbf{A}_*$ with only 1's (no inbreeding) on the diagonals.

For the example data, the generalized inverse of the coefficient matrix was obtained when: (1) $\mathbf{D}=\mathbf{I}\lambda$ was added to the diagonal coefficients for the nine sire equations (i.e., $\lambda=19$ added to each sire diagonal), and (2) $\mathbf{D}=\lambda\mathbf{A}_*^{-1}$ was added to the nine by nine diagonal block of the coefficient matrix corresponding to the nine sires. The apparent and actual prediction error variances (would need to be multiplied by σ^2) are shown in Table 1 for $\mathbf{D}=\mathbf{I}\lambda$, $\mathbf{D}=\lambda\mathbf{A}_*^{-1}$, and $\mathbf{D}=\lambda\mathbf{A}^{-1}$. The columns labelled apparent contain the diagonal coefficients from the inverse of the incorrect coefficient matrix. The actual columns are computed from the algebraic quantities developed earlier from expected values. For $\mathbf{D}=\lambda\mathbf{A}^{-1}$ the apparent and actual computations should be, and are, equal. The approximation $\mathbf{D}=\lambda\mathbf{I}$ resulted in underestimation of PEV. The approximation $\mathbf{D}=\lambda\mathbf{A}_*^{-1}$ for $\mathbf{A}_*$ from the simple rules for $\mathbf{A}^{-1}$ resulted in overestimation of PEV. The actual PEV, however, with $\mathbf{D}=\lambda\mathbf{I}$ and $\mathbf{D}=\lambda\mathbf{A}_*^{-1}$ were only slightly larger than for the correct model with $\mathbf{D}=\lambda\mathbf{A}^{-1}$. The $\mathbf{A}_*^{-1}$ approximation was better than the **I** approximation and was essentially the same as with correct **A**. As theory dictates, the correct model gave the smallest actual PEV although negligibly smaller than from the $\lambda\mathbf{A}_*^{-1}$ approximation and only slightly smaller than from the extreme approximation, the $\lambda\mathbf{I}$ approximation.

The application of the simple formula that ignores inbreeding to obtain the r_{TI} is affected greatly by the incorrect relationship matrix as shown in the bottom half of Table 1. The diagonals of the inverse of the coefficient matrix with incorrect **D** obviously did not provide good approximations of r_{TI}^2 even when $(1+F_i)$ was used in the calculation. For example, with $\mathbf{D}=\lambda\mathbf{I}$, for sire X1 the simple but incorrect calculation is:

$$r_{TI}^* = [1-0.0502(19)]^{0.5} = 0.2149.$$

If F_{X1} is used,

$$r_{TI}^* = [1-(0.0502)(19)/(1.25)]^{0.5} = 0.4868.$$

The actual r_{TI} calculated from $\mathrm{Cov}(u_i^0, u_i)$, $V(u_i^0)$, and $V(u_i)$ is only 0.0731.

The incorrect calculation with $\mathbf{D}=\lambda\mathbf{A}_*^{-1}$ fortunately is obviously not believable. For example, for sire X1, $r_{TI}^* = [1-0.0767(19)]^{0.5} = [-0.4573]^{0.5}$! Adjusting for inbreeding in the simple formula does not help much, $[1-0.0767(19)/(1.25)]^{0.5} = [-0.1658]^{0.5}$!!

With correct $\mathbf{D}=\lambda\mathbf{A}^{-1}$, the formula based on the inverse of the correct coefficient matrix with an adjustment for inbreeding leads to the same result as the definitional formula for a correlation worked out with expected values. Again for sire X1, $r_{TI} = [1-0.065438(19)/(1.25)]^{0.5} = 0.0731$. If the formula ignoring inbreeding is used, the calculation leads to the square root of a negative number for each of the inbred sires and was the stimulus for this discussion.

What stands out from this example with a sire model is how little difference there is between the actual r_{TI} for the two approximations to **A** in the MME as compared to true **A** even though the apparent r_{TI} were misleading and impossible. The same pattern is true for PEV. Whether the similarity in actual r_{TI} would be true for all cases with sire models cannot be stated. This example does suggest that the easy method of calculating the inverse of the relationship matrix by pretending no inbreeding exists may not affect the actual PEV and r_{TI} very much. In fact, the approximate $\mathbf{A}_*$ that resulted in this example implied even more inbreeding than actually existed. With an individual ("animal" to animal breeders) genetic model the similarity between actual PEV and r_{TI} with approximations to **A** and with true **A** in the MME may not be as likely.

References

Golden BL, Brinks JS, Bourdon RM (1991) A performance-programmed method for computing inbreeding coefficients from large data sets for use in mixed-model analyses. J Anim Sci 69:3564–3573

Henderson CR (1975a) Comparison of alternative sire evaluation methods. J Anim Sci 41:760–770

Henderson CR (1975b) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447

Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83

Henderson CR (1977) Best linear unbiased prediction of breeding values not in the model for records. J Dairy Sci 60:783–787

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Ontario, Canada

Quaas RL (1976) Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949

Van Vleck LD (1988) Notes on the theory and application of selection principles for the genetic improvement of animals. Cornell University, Ithaca, N.Y., USA