

# Effects of Non-Random Mating

L. R. Schaeffer, April 1999

Updated April 2000

The models that have been discussed in this course have assumed an infinitesimal animal model, and that animals are mating randomly. What are the effects when these assumptions are not true.

There are many forms of selection that can nullify or render meaningless any statistical analysis if it is ignored. In this set of notes, only selection of animals to produce the next generation, or non-random mating, will be considered.

Non-random mating produces the following consequences.

1. Causes changes to gene frequencies at all loci.
2. Changes in gene frequencies cause a change in genetic variance. Recall from quantitative genetics that

$$\sigma_G^2 = 2pq[a + d(q - p)]^2 + [2pqd]^2.$$

3. In finite populations, non-random mating causes a reduction in the effective population size, which subsequently causes an increase in levels of inbreeding.
4. Joint equilibrium becomes joint disequilibrium, and therefore, non-zero covariances between additive and dominance genetic effects are created.
5. If the pedigrees of animals are not complete nor traceable to the base generation, then non-random mating causes genetic evaluations by BLUP to be biased, and causes estimates of genetic variances, by any method, to be biased.
6. Because the genetic variance decreases due to joint disequilibrium and inbreeding, response to selection is generally lower than expected by selection index over several generations.

## Effect on Inbreeding

Recall that under random mating and finite population size that the change in inbreeding per generation was defined as

$$\Delta F = \frac{1}{2N_e},$$

where  $N_e$  is the effective population size, or effective number of breeding individuals. Wright (1940) and Crow (1954) showed that if the average family size, was  $v$ , then

$$N_e = 4N/(v + 2),$$

where  $N$  is the number of mating males and females. An idealized population that is able to maintain itself has  $v = 2$ , thus,  $N_e = N$ . Belonsky and Kennedy (1985) simulated a population of 100 females and 5 males per year over ten generations with a single record per animal and discrete generations. Parents were either randomly selected, selected on the basis of phenotype, or selected on the basis of BLUP EBVs. The change in inbreeding under the different selection criteria are given in the table below.

Selection Criteria	Heritability		
	.1	.3	.6
Random	.152	.152	.152
Phenotypic	.169	.206	.215
BLUP EBV	.288	.299	.274

Under random selection of mates, the rate of inbreeding over 10 years is the same for all heritabilities, and is lower than the other forms of mate selection. Inbreeding accumulates due simply to the small population size. Phenotypic selection of mates gives higher rates of inbreeding which increase with increasing heritabilities. Selection of mates using BLUP EBV created the highest rates of inbreeding. Note that there was a decline in inbreeding rate at heritability equal to .6. Why? BLUP EBVs make use of information from relatives. At low heritabilities an animal's EBV is more heavily influenced by the parent average, and many progeny are needed to overcome this influence. Selection on the parent average results in more related animals being selected together, and hence more inbreeding. At the higher heritability of .6, the animal's own record has more influence relative to the parent average, resulting in fewer half-sibs and full-sibs being selected as parents of the next generation, and consequently lower inbreeding levels.

### Effect on Genetic Evaluation

The effects of non-random mating on genetic evaluation are minimal if

- Complete (no missing parent information) pedigrees are known back to a common base population which was mating randomly,
- Data on all candidates for selection are available, and
- Genetic parameters from the base population are known.

If the above conditions hold, then application of BLUP does not lead to bias in EBVs, but selection increases the variance of prediction error over populations that are randomly mating. However, in animal breeding, the practical situation is that complete pedigrees seldom exist. Thus, bias can creep into estimates of fixed effects and EBVs.

Recall that HMME for a simple animal model are

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{nn}k_a & \mathbf{A}^{no}k_a \\ \mathbf{0} & \mathbf{A}^{on}k_a & \mathbf{A}^{oo}k_a \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}}_n \\ \hat{\mathbf{a}}_o \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{0} \end{pmatrix},$$

where  $k_a = \sigma_a^{-2}$ . A generalized inverse of the coefficient matrix can be represented as

$$\begin{pmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xn} & \mathbf{C}_{xo} \\ \mathbf{C}_{nx} & \mathbf{C}_{nn} & \mathbf{C}_{no} \\ \mathbf{C}_{ox} & \mathbf{C}_{on} & \mathbf{C}_{oo} \end{pmatrix}.$$

Then remember that

$$Var \begin{pmatrix} \hat{\mathbf{a}}_n - \mathbf{a}_n \\ \hat{\mathbf{a}}_o - \mathbf{a}_o \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{nn} & \mathbf{C}_{no} \\ \mathbf{C}_{on} & \mathbf{C}_{oo} \end{pmatrix},$$

and that

$$\begin{aligned} Cov(\hat{\mathbf{b}}, \hat{\mathbf{a}}_n) &= \mathbf{0}, \\ Cov(\hat{\mathbf{b}}, \mathbf{a}_n) &= -\mathbf{C}_{xn}. \end{aligned}$$

These results indicate that HMME forces the covariance between estimates of the fixed effects and estimates of additive genetic effects to be null. However, there is a non-zero covariance between estimates of the fixed effects and the true additive genetic values of animals. Hence, any problem with the true additive genetic values, and there will be problems with estimates of fixed effects.

Consider the equation for  $\hat{\mathbf{b}}$ ,

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{a}}_n),$$

and the expectation of this vector is

$$E(\hat{\mathbf{b}}) = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}E(\hat{\mathbf{a}}_n)).$$

The fixed effects solution vector contains a function of the expectation of the additive genetic solution vector. Normally, because the BLUP methodology requires

$$E(\hat{\mathbf{a}}_n) = E(\mathbf{a}_n) = \mathbf{0},$$

then the fixed effects solution vector is also unbiased. Due to selection, however,

$$E(\mathbf{a}_n) \neq \mathbf{0},$$

and therefore, the expectation of the fixed effects solution vector contains a function of  $E(\mathbf{a}_n)$  and is consequently biased. If  $\hat{\mathbf{b}}$  is biased, then this will cause a bias in  $\hat{\mathbf{a}}$ .

## Alternative Methods

Re-state the model (in general terms) as

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

where

$$E \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{u} \\ \mathbf{0} \end{pmatrix},$$

and therefore,

$$E(\mathbf{y}) = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u}.$$

To simplify, assume that  $\mathbf{G} = Var(\mathbf{u})$  and  $\mathbf{R} = Var(\mathbf{e})$  and that neither is drastically affected by non-random mating.

The prediction problem is the same as before. Predict a function of  $\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}$  by a linear function of the observation vector,  $\mathbf{L}'\mathbf{y}$ , such that

$$E(\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}) = E(\mathbf{L}'\mathbf{y}),$$

and such that  $Var(\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u} - \mathbf{L}'\mathbf{y})$  is minimized. Form the variance of prediction errors and add a LaGrange multiplier to ensure the unbiasedness condition, then differentiate with respect to the unknown  $\mathbf{L}$  and the matrix of LaGrange multipliers and equate to zero. The solution gives the following equations.

$$\begin{pmatrix} \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{V}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{V}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{V}^{-1}\mathbf{y} \end{pmatrix}.$$

Because  $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ , and

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{R}^{-1},$$

for  $\mathbf{T} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}$ , then it can be shown that the following equations give the exact same solutions as the previous equations.

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

If a generalized inverse to the above coefficient matrix is represented as

$$\begin{pmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xz} \\ \mathbf{C}_{zx} & \mathbf{C}_{zz} \end{pmatrix},$$

then some properties of these equations are

$$\begin{aligned} Cov(\hat{\mathbf{b}}, \mathbf{u}) &= \mathbf{0}, \\ E(\hat{\mathbf{b}}) &= (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b}, \\ Cov(\hat{\mathbf{b}}, \hat{\mathbf{u}}) &= \mathbf{C}_{xz}. \end{aligned}$$

Firstly, these results suggest that if non-random mating has occurred and has changed the expectation of the random vector, then an appropriate set of equations is the generalized least squares equations. However, we have seen earlier that such equations give a lower correlation with true values and large mean squared errors (when matings are at random). Secondly, the estimates of the fixed effects have null covariances with the true random effects, and the covariances between estimates of the fixed effects and estimates of the random effects are non-zero, which is **opposite** to the results from BLUP. With the least squares solutions, application of the regressed least squares procedure could be subsequently used to give EBVs.

There is another problem with these equations. If  $\mathbf{u} = \mathbf{a}$  as in an animal model, then  $\mathbf{Z} = \mathbf{I}$ , and the generalized least squares equations do not have a solution unless  $\hat{\mathbf{a}} = \mathbf{0}$ . This is not very useful for genetic evaluation purposes.

### An Alternative Model

Earlier in these notes, the Mendelian sampling variance was assumed to be unaffected by non-random mating, but could be reduced by the accumulation of inbreeding. The animal model equation is

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{Z}\mathbf{p} + \mathbf{e}.$$

The animal additive genetic effect can be written as

$$\mathbf{a} = \mathbf{T}_s\mathbf{s} + \mathbf{T}_d\mathbf{d} + \mathbf{m},$$

where  $\mathbf{T}_s$  and  $\mathbf{T}_d$  are matrices of ones and zeros, such that each row has an element that is 1 and all others are 0, and these indicate the sire and dam of the animal, respectively, and  $\mathbf{m}$  is the Mendelian sampling effect. Due to non-random mating then,

$$E(\mathbf{a}) = \mathbf{T}_s\mathbf{s} + \mathbf{T}_d\mathbf{d},$$

which is not a null vector, in general. Let

$$\begin{aligned}\mathbf{Z}_s &= \mathbf{Z}\mathbf{T}_s, \\ \mathbf{Z}_d &= \mathbf{Z}\mathbf{T}_d,\end{aligned}$$

then the model becomes

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_s\mathbf{s} + \mathbf{Z}_d\mathbf{d} + \mathbf{Z}\mathbf{m} + \mathbf{Z}\mathbf{p} + \mathbf{e}.$$

Also,

$$\begin{aligned}E(\mathbf{y}) &= \mathbf{X}\mathbf{b} + \mathbf{Z}_s\mathbf{s} + \mathbf{Z}_d\mathbf{d}, \\ E(\mathbf{m}) &= \mathbf{0}, \\ E(\mathbf{p}) &= \mathbf{0}, \\ E(\mathbf{e}) &= \mathbf{0},\end{aligned}$$

and

$$Var \begin{pmatrix} \mathbf{m} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{B}\sigma_a^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix},$$

where  $\mathbf{B}$  is from

$$\mathbf{A} = \mathbf{TBT}'.$$

If all animals were non inbred then all of the diagonals of  $\mathbf{B}$  would be equal to .5.

Note that the matrix  $\mathbf{A}$  or its inverse are not necessary in this model, and that sires and dams (resulting from selection) are fixed effects in this model. The equations to solve are

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z}_s & \mathbf{X}'\mathbf{Z}_d & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'_s\mathbf{X} & \mathbf{Z}'_s\mathbf{Z}_s & \mathbf{Z}'_s\mathbf{Z}_d & \mathbf{Z}'_s\mathbf{Z} & \mathbf{Z}'_s\mathbf{Z} \\ \mathbf{Z}'_d\mathbf{X} & \mathbf{Z}'_d\mathbf{Z}_s & \mathbf{Z}'_d\mathbf{Z}_d & \mathbf{Z}'_d\mathbf{Z} & \mathbf{Z}'_d\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}_s & \mathbf{Z}'\mathbf{Z}_d & \mathbf{Z}'\mathbf{Z} + \mathbf{B}^{-1}k_a & \mathbf{Z}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}_s & \mathbf{Z}'\mathbf{Z}_d & \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{Z} + \mathbf{I}k_p \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{s}} \\ \hat{\mathbf{d}} \\ \hat{\mathbf{m}} \\ \hat{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'_s\mathbf{y} \\ \mathbf{Z}'_d\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix}.$$

Thus, for each animal with a record we need to know both parents, but we do not need to be able to follow pedigrees back to the base generation, except for calculating inbreeding coefficients of all animals.

This model was applied to the example data which was simulated in Lesson 9. There were 12 animals with records, four sires and four dams. The solutions for the sires and dams are shown below, after forcing their sum to be zero.

$$\begin{aligned} \hat{\mu} &= 49.558333, \\ \hat{s}_1 &= -7.079167, \\ \hat{s}_3 &= 7.2875, \\ \hat{s}_5 &= -16.14583, \\ \hat{s}_7 &= 15.9375, \\ \hat{d}_2 &= 1.7375, \\ \hat{d}_4 &= 0.0875, \\ \hat{d}_6 &= 3.9375, \\ \hat{d}_8 &= -5.7625, \end{aligned}$$

The solutions for sires and dams represent estimated transmitting abilities and should be multiplied by 2 to give EBV. The estimates of the Mendelian sampling effects for animals

5 through 16 were

$$\hat{\mathbf{m}} = \begin{pmatrix} -1.254878 \\ -1.763415 \\ 1.254878 \\ 1.7414634 \\ -0.296341 \\ 1.4634146 \\ 0.2743902 \\ 0.7829268 \\ -0.486585 \\ 0.5085366 \\ 0.0219512 \\ -2.246341 \end{pmatrix}.$$

These solutions sum to zero, automatically, but the general property would be  $\mathbf{1}'\mathbf{B}^{-1}\hat{\mathbf{m}} = 0$ . In this example all of the diagonal elements of  $\mathbf{B}^{-1}$  were equal to 2, but with inbred individuals this would not be the case.

EBV are created by summing sire and dam solutions with the Mendelian sampling estimates. The results for animals 5 through 16 were

$$EBV = \begin{pmatrix} -6.60 \\ 5.61 \\ 10.28 \\ -5.25 \\ -12.50 \\ 11.64 \\ -21.63 \\ 20.66 \\ -3.63 \\ 2.03 \\ -16.04 \\ 7.93 \end{pmatrix}.$$

The correlation of EBV with the true breeding values (shown in Lesson 9) was .8009 which is greater than the correlation obtained with BLUP (.7547). This model appears to be superior to the simple animal model (based on only one example). However, from this model it is possible to have two EBV for some animals. Animal 5, for example, had an EBV of -6.60 based on its own record plus its sire and dam solutions, and based on its progeny as a sire has an EBV of -14.16. Which evaluation is correct or better to use? If the progeny of animal 5 are random progeny, and if animal 5 has a chance to have many more progeny, then EBV=-14.16 is probably the better result to use. If the progeny are not a random sample of progeny, then the other EBV may be better. The correlation of sire and dam solutions with their true breeding values for animals 1 to 8 was .5573. For

animals 1 to 4, the sire and dam solutions are the only information available for these animals because they did not have records.

The simple animal model combines information from data, from the parent average, and from progeny. The above model computes estimated breeding values based on progeny only, and based on parent average plus data. This difference in concept is due to the fact that parents are obtained by selection. Non-random mating is taken into account because the sire and dam of each animal with a record is included in the model. The solutions for sires and dams from this model are valid estimates of transmitting abilities provided that the progeny are a random sample of their progeny. The Mendelian sampling estimates for animals provides a means of estimating the additive genetic variance. Inbreeding should still be taken into account in the matrix  $\mathbf{B}$ . This model also avoids the problem of forming phantom parent groups for animals with parent information. If an animal with a record has an unknown dam (or sire), then a phantom dam (sire) can be created which has this animal as its only progeny. If both parents are unknown, then both a phantom sire and phantom dam need to be assumed, with this animal as their only progeny. Further study on this model is warranted.