

Efficient selection rules to increase non-linear merit : application in mate selection (*)

S.P. SMITH (**) and F.R. ALLAIRE (***)

*Department of Dairy Science
The Ohio State University, Columbus, OH 43210 (U.S.A.)*

Summary

Merit is defined to be a non-linear function of an animal's phenotype for various traits. A Bayesian type selection rule to increase merit in hypothetical populations is proposed. The rule is based on the conditional expectation of total merit in the population given data. This rule has similarities to selection index theory. An animal's phenotype for any trait and data are assumed distributed as multivariate normal random variables. Situations are treated when associated population means are known or unknown. When means are unknown and must be estimated, the procedures can take advantage of mixed model methodology. An illustration of its application to a mate selection problem is presented.

Key words : Bayesian methods, mate selection, non-linear merit, selection.

Résumé

*Décisions efficaces de sélection pour une fonction d'objectif non linéaire :
application aux choix des conjoints*

L'objectif de sélection est défini par une fonction non linéaire de la valeur phénotypique d'un animal pour différents caractères. Une décision de sélection de type bayésien est proposée pour accroître la fonction d'objectif dans diverses situations hypothétiques. La décision de sélection correspond à l'espérance conditionnelle de l'objectif dans la population sachant les données recueillies.

Cette règle présente des similitudes avec la théorie des indices de sélection. On suppose notamment que le phénotype de l'animal et les données ont une distribution conjointe multi-normale. On aborde les cas de moyennes connues et inconnues. En situation de moyennes inconnues à estimer, les méthodes proposées peuvent s'inspirer de celle de modèle mixte. Un exemple d'application relatif au choix des conjoints est donné.

Mots clés : Méthode bayésienne, choix des conjoints, objectif non linéaire, sélection.

(*) Approved as Journal Article No. 63-84, the Ohio Agricultural Research and Development Center, The Ohio State University.

(**) A contributing project to North Central Regional Project, NC-2, Improving Dairy Cattle Through Breeding.

(**) Present address : Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia.

(***) Reprint request.

I. Introduction

The goal of artificial selection is typically to increase some quantity (T) in the selected population. When T is a relatively simple quantity, the selection index and linear model procedures are quite powerful aids to selection. T can be considered simple if, for example, it is a linear combination of additive genetic effects. In this case, the linear combination may reflect the relative economic worth of each genetic effect.

To say T is complicated is frequently due to a belief that hypothetical components of T are well described by additive genetic models. In this setting the practitioner is unwilling to use simple additive models to describe T itself (i.e., if T can be measured). Our paper is directed at this situation.

When T is complicated, « optimal » selection rules become complicated and the usefulness of the selection index or linear model procedures are much in doubt. Complicated merit functions have been described by Allaire (1980) in the context of mate selection.

In this paper, T will be an expression that reflects the economic merit or utility of an animal's phenotype (or phenotypes). Assume

$$T = \sum_{i=1}^n f_i (P_i) \quad [1]$$

where P_i is the phenotype for the i^{th} trait, $f_i (\cdot)$ is an arbitrary function that assigns an economic value to P_i . The arbitrary functions will be assumed known *a priori*.

There are a number of observations that should be made about [1] :

a) It has been assumed, rather arbitrarily, that merit is a function of n traits (i.e., $P_1, P_2, \dots P_n$). The choice of which traits is usually a personal one. Merit need not exist independently for any one of the traits. Merit is a subjective quantity assigned to all the traits in concert.

b) We have not used the most general representation of T (i.e., $T = f (P_1, P_2, \dots P_n)$). This is simply a practical requirement and it is theoretically unjustified. It would be harder to estimate a more general function. Moreover, given such a function, application of theory presented in this paper would be made harder. We are not advocating the use of [1] for all applications. However, [1] can be made more general implicitly if we define $P_1, P_2, \dots P_n$ as arbitrary (but known) linear combinations of phenotypic measurements ($M_1, M_2, \dots M_m$). In this setting, $M_1, M_2, \dots M_m$ determines our subjective ideal of merit. This interpretation causes no problems with methods in our paper.

c) T is a function of the phenotypes and not the genotypes directly. This convention is not mandatory for all selection problems. However, we decided to use it because the economic utility of any animal can generally be quantified through phenotypic relationships. Furthermore, if the function $f (\cdot)$ assigns a merit ($f (P)$) to phenotype P then it should not be assumed that $f (G)$ represents the merit of genotype G (where $P = G + E$, and E is an environmental effect). Still, statements related to genotypic worth can be made. For example, the genotypic value of a sire in breeding may be taken to equal the expected phenotypic worth of his progeny. Realization of genotypic

worth is ultimately mediated by a phenotype or phenotypes. Thus, the genotypic worth may be a function of both genetic and non-genetic quantities. Usually the non-genetic quantities will reflect (in some way) the class of all possible environmental happenings. When T is a simple function the above distinction is usually only academic. However, when T is complicated the distinction is critical.

d) The function T can be generalized to accommodate things like sex differences, inbreeding depression and animal dependent investment cost. For example, two functions like [1] can be defined for each sex. Investment cost can be included in [1] by adding an extra term (usually negative). Examples of investment cost are semen cost or the cost of purchasing breeding animals. To accommodate a more general T, methods in this paper can be extended in a straight-forward manner. However, to describe methods for a more general T would only serve to obscure our message.

It is the purpose of this paper to describe practical selection rules that aid in increasing T in hypothetical populations. The rules are designed for the realization of short term response.

II. Bayesian selection theory

Selection requires a decision. Consequently, standard techniques in decision theory can be used to establish useful selection rules. In this section, we will describe Bayesian decision rules (BERGER, 1980, p. 14) in the context of selection.

We will not use words like « optimal » or « best » to describe selection rules. These words foster misconceptions. To call a selection rule best implies a certain objectivity that does not usually exist. Decisions are affected by subjective beliefs or attitudes. Bayesian methods force users to identify their subjectivity.

Despite subjectivity, Bayes decision rules can be justified by strong arguments. If one is to be consistent with « rationality axioms » then his decision rule should be equivalent to some Bayes rule (BERGER, 1980, p. 91). This means that if our decisions are not equivalent to some Bayes rule, then we might be accused of being irrational.

Establishing a useful Bayes rule depends upon the appropriateness of assumptions related to preference and prior information. In practice, needed assumptions may seem arbitrary.

A. Development

The objective of selection is to increase overall merit of a hypothetical population^(a). After selection this population will be called the « selected population ». The selected population need not represent the population that underwent physical selection. For example, given that physical selection involves the formation of mating pairs

(a) This view may be too simplistic for some applications. The objective of selection may be to increase merit in several populations. If populations are defined by the time frames then discounting may need to be considered. In his case T will need to be redefined.

(i.e., sires and dams), the selected population may be the resulting progeny. That is, the objective of selection may be to increase the overall merit of the progeny. The selected population will be understood to be finite. Thus, given the phenotypes of this population, the total merit can be calculated exactly using [1]. However, these phenotypes will generally be unknown when selection decisions are being made.

The selection rule (S) is a function of data (say a column vector y). That is, $S(y)$ defines a signal specifying an action (a) of choosing one of numerous selection alternatives. Thus, a or $S(y)$ will set in motion the stochastic mechanism that will determine the selected population. Every action is associated with a loss determined by a loss function. The loss function is at least a function of $w(a)$, where w is the true state of nature in the selected population. Here, w is simply a vector containing the realized phenotypes.

The opportunity cost can be derived from the definition of T. Define $M(a)$ as the sum of the realized merit or utility (i.e., T) from each individual in a selected population. Hence, $M(a)$ represents the total merit or utility of the selected population resulting from an action a ^(b). Given an alternative action a' , the opportunity cost is then $M(a') - M(a)$. With a' fixed, it is quite natural to take the opportunity cost as the loss function corresponding to action a . Moreover, the loss function may simply be taken as $-M(a)$. This assignment will be used.

It would be nice to choose some action among all acceptable actions (A) so the loss is minimized. Unfortunately, when decisions need to be made the losses resulting from various actions are not usually known. However, given y and a the stochastic behavior of $w(a)$ may be known. If so, the necessary ingredients are available to choose an action by Bayes rule.

BERGER (1980, p. 109) states that the Bayes rule can be found by choosing an action among A , that minimizes the conditional expectation of the loss given data. Thus, the selection rule that will be proposed, is to find an action, a^* included in A , that will minimize

$$E [-M(a) | y] \quad [2]$$

when $a = a^*$. Note that minimizing $E [-M(a) | y]$ is the same as maximizing $E [M(a) | y]$.

In order to find a^* it is sufficient to do the following :

a) Determine the smallest set of individuals containing all individuals in all possible selected populations represented by selection schemes in A . If all selected populations consist of offspring of known animals, this requirement would consist of listing parents or mating pairs.

b) Compute $E [T | y]$ for each uniquely identified individual.

c) Identify a^* by inspection or by comparing a sufficient number of the quantities given by [2], where a is in A . The total of the conditional expectations of the losses for each a (i.e., [2]) can be evaluated by adding together the negative of the appropriate quantities computed in b).

(b) It is technically improper to assume that $M(a)$ represents the utility resulting from action a . That is, the utility of action a need not be representable as a sum of utilities corresponding to individuals in the selected population. We will assume otherwise due to practical considerations. For a discussion of utility theory, see BERGER (1980).

B. Application

The difficulty in finding a^* is a function of the complexity of both A and of stochastic properties of $w(a)$. When these complexities are relatively minor the Bayes selection rule reduces to procedures that are familiar to most animal breeders. For example, consider the use of the selection index in ranking animals for real producing ability. A typical action would be to select a fixed proportion of animals; those corresponding to the highest index values. From a decision theoretic perspective, this corresponds to taking A to be the set of all actions that involve selecting a fixed proportion of animals. Moreover, the utility of individuals in the selected population can be assigned exclusively to animals that are physically selected. With this variety of decision problem, the selection rule proposed here involves computing conditional expectations of T for each animal and selecting animals corresponding to the highest expectations. BULMER (1980, p. 196) developed a similar rule to increase the genetic merit of pure lines.

In mate selection problems, the Bayes selection rule can become complicated. For example, assume that there are 15 sires available (via artificial insemination) to be mated to 20 cows. An attempt will be made to mate each cow only once in the next month. However, any sire will be used once, several times or not at all. Let i index the i -th sire, $i = 1, 2, \dots, 15$. Assume that the i -th sire has only n_i units of semen available. Thus, the i -th sire can not be used more than n_i times. Clearly, the class of acceptable actions is very large and possesses complicated constraints. Moreover, the utility of each individual in the selected population can be assigned to a sire-dam pair rather than just one animal (i.e., for one stage selection).

To solve the mate selection problem it is best to refer to the three rules given earlier. Step c) can be cast as an integer linear programming problem. This fact has been discovered independently by JANSEN & WILTON (1984). Let j index the j -th cow, $j = 1, 2, \dots, 20$, and let c_{ij} equal the expected T for the progeny produced by mating the i -th sire to the j -th cow. The integer linear programming problem is

$$\begin{aligned} \text{Maximize : } & \sum_{ij} x_{ij}c_{ij} \\ \text{Subject to : } & \sum_i x_{ij} = 1, j = 1, 2, \dots, 20 \\ & \sum_j x_{ij} \leq n_i, i = 1, 2, \dots, 15 \\ & x_{ij} = 0 \text{ or } 1, \text{ all } i \text{ and } j \end{aligned}$$

This problem can be solved by using the methods described in PFAFFENBERGER & WALKER (1976). If $x_{ij} = 1$ when the solution is found, then inseminate the j -th cow with semen from the i -th sire.

GODDARD (1983) suggests that non-random mating (alone) should not be used to improve long term genetic gain. We agree; however, all mate selection schemes should not be considered as simply non-random mating or assortative mating. Mate selection is the synthesis of selection and non-random mating. Mate selection can affect reproductive fitness (usually fitness of males).

One stage mate selection can be used sequentially to improve long term merit. Mate selection is similar (but less restrictive) to creating subdivisions in the population where mating (following selection) occurs only within subpopulations. Each subpopula-

tion can be sequentially selected so as to improve long term merit. Yet the direction in which subpopulation means are changed may be quite different. It should be noted that random mating can destroy gains made via mate selection. If mate selection is to be practiced, random mating should never be allowed. It should also be noted that sequential single stage selection may direct some subpopulation to a locally desirable state of nature rather than a globally desirable state. This seems to depend on the shape of the merit function. The last criticism is directed at single stage selection and not mate selection per se. Admittedly, determining mating pairs that maximize long term expected merit is complicated.

It is difficult to say when mate selection is preferable (long term response) to alternative methods. Consider only a univariate merit function (i.e., $T = F(P)$). If $f(\cdot)$ is monotone it may make little difference if mate selection or selection with random mating is used. Alternatively, if $f(\cdot)$ has a global maximum near the population mean, the question of long term response maybe a little ill-posed. In this situation, control of population variance becomes more important. If $f(\cdot)$ is « U » shaped, mate selection should fragment a population into « high » and « low » lines. Mate selection can do this more effectively than approaches that do not allow all animals to contribute genes to both lines (when advantages). This advantage is lost when the lines become so different that migration between them (when advantages) becomes unlikely. Mate selection is probably most valuable as a tool to realize short term gains. For example, mate selection may be useful in controlling calving difficulty in dairy or beef cows.

A third type of selection problem is the gene pool problem. For this case a fixed number of parents are selected and allowed to contribute genes to a hypothetical gene pool (thoroughly mixed by recombination). The object is to select those parents that maximize the expected merit of a randomly selected representative (animal) of the gene pool. Note that each selected population (corresponding to a particular gene pool) can be thought of as having one individual. Thus, only one $E[T | y]$ need be computed for each group of parents (action) considered. Important considerations pertaining to the evaluation of $E[T | y]$ are given in Annex A. The Bayes one stage selection scheme is very similar (but different, see Annex A) to the procedure given by BULMER (1980, p. 197). GODDARD (1983) points out that this kind of problem is very difficult to solve because it is usually not practical to enumerate all possible parent combinations (actions). Thus, step a) of the rules given earlier may be prohibitive. It might be better to approximate a solution to the gene pool problem by using the linear indices described by GODDARD (1983) or MOAV & HILL (1966). The selection rule proposed by GODDARD (1983) is equivalent to the Bayes rule, if a unique Bayes rule exists and given additional assumptions (equal information, infinite population size, selected animals are sufficiently unrelated, population means known). Approximate solutions can be improved as outlined in Annex A.

In this paper the stochastic properties of $w(a)$ will be assumed to be relatively simple. Precisely, the phenotypes associated with $w(a)$ will be taken to have a conditional normal distribution given data. This convention is suitable for one stage selection. Methods presented in this paper are designed only for short term gains.

The selection rules given here can be implemented in a sequential manner. The decisions of the past are usually responsible for the propagation of observations that will be used to make up-to-date decisions. Expectation [2] can be evaluated by ignoring the fact that records (i.e., y) are selected, if the vector y contains all the observations that prior decisions were based on. This result was demonstrated by GOFFINET (1983) and FERNANDO & GIANOLA (1984).

III. Computing the expectation

Let T_k represent the merit of animal k in a selected population. Denote the realized phenotypes for various traits on animal k as P_{ik} , $i = 1, 2, \dots, n$. Using [1] it can be shown that $E [T_k | y]$ is equal to

$$\sum_{i=1}^n E [f_i (P_{ik}) | y] \quad [3]$$

This section is devoted to describing methods that can be used to compute

$$E [f (P) | y] \quad [4]$$

where $f(\cdot)$ is some function (representing $f_i(\cdot)$), P is a phenotype (representing P_i). These methods can be implemented directly, in order to compute the various terms in [3]. Computed terms can be combined in order to obtain $E [T_k | y]$. Thus, $E [T | y]$ can be computed for various individuals and a^* can be determined as outlined in the previous section.

P and y in [4] will be assumed to have a multivariate normal distribution with a known variance-covariance structure. For now we will assume that means associated with P and y are known. In order to evaluate [4], the posterior density of P given y must be determined. This can be done by using standard selection index theory (Van VLECK, 1974). Let

$$E \begin{bmatrix} P \\ y \end{bmatrix} = \begin{bmatrix} U_p \\ U_y \end{bmatrix} \quad \text{and} \quad \text{Var} \begin{bmatrix} P \\ y \end{bmatrix} = \begin{bmatrix} r & \mathbf{d}' \\ \mathbf{d} & \mathbf{V} \end{bmatrix}$$

Then P given y has a normal distribution with mean $U_p + \mathbf{d}'\mathbf{V}^{-1}(\mathbf{y} - U_y)$ and variance $r - \mathbf{d}'\mathbf{V}^{-1}\mathbf{d}$. Denote the mean as $U_{p|y}$ and the variance as $\sigma_{p|y}^2$. Using standard terminology, $U_{p|y}$ is the selection index and $\sigma_{p|y}^2$ is the prediction error variance. The selection index and $\sigma_{p|y}^2$ are necessary ingredients to evaluate [4].

In the next subsection we will describe algorithms that can be used to evaluate [4] given $U_{p|y}$ and $\sigma_{p|y}^2$. The same algorithms can be used when means associated with P and y are unknown. However, $U_{p|y}$ and $\sigma_{p|y}^2$ must be modified as we will see later. The unknown means situation is certainly the most realistic characterization of knowledge pertaining to P and y .

A. Algorithms

One way [4] can be evaluated is by GAUSSIAN quadrature (STOER & BULIRSCH, 1980, pp. 142-151). This method can be used for an arbitrary $f(\cdot)$. Details of this method are given in Annex B.

Method of evaluating [4] may be closely allied with methods of estimating $f(\cdot)$. For example, an attempt might have been made to describe $f(\cdot)$ as a polynomial. In which case $f(P)$ can be taken to equal $\sum_{i=0}^s a_i P^i$ and consequently, [4] can be expressed as

$$\sum_{i=0}^s a_i E [P^i | y]. \quad [5]$$

The terms (i.e., $E [P^i | y]$) in [5] can be computed directly via recursion. That is, $E [P^0 | y] = 1$, $E [P^1 | y] = U_{p|y}$ and for $i \geq 2$, $E [P^i | y] = (i - 1) \sigma_{p|y}^2 E [P^{i-2} | y] + U_{p|y} E [P^{i-1} | y]$. For the situation when $s = 2$ [5] can be written as

$$a_0 + a_1 U_{p|y} + a_2 U_{p|y}^2 + a_2 \sigma_{p|y}^2.$$

Quadratic indices have been described by WILTON *et al.* (1968). These authors ignored terms analogous to $a_2 \sigma_{p|x}^2$ in their indices. Clearly, $a_2 \sigma_{p|y}^2$ should be considered if candidates available for selection have unequal information.

Estimating $f(\cdot)$ by a polynomial may be ill-advised because such a scheme may induce unrealistic fluctuations in the estimate (i.e., if $f(\cdot)$ is not a polynomial). Generally, $f(\cdot)$ can be better estimated as a piece-wise cubic. In addition to being piece-wise cubic, the estimate of $f(\cdot)$ can be made to be continuous and first derivative continuous. Piece-wise estimation can be handled via interpolation by spline function (STOER & BULIRSCH, 1980, pp. 93-106). Alternatively, piece-wise linear regression (NETER & WASSERMAN, 1974) might be useful in estimating $f(\cdot)$. The regression approach can be generalized in a straight-forward manner to piece-wise cubic models. Appropriate continuity constraints can be imposed by the method of Lagrange multipliers (KAPLAN, 1973). A method of evaluating [4] when $f(\cdot)$ is a piece-wise cubic is presented in Annex B.

It should be clear that [4] can be evaluated with the aid of a computer. Moreover, $f(\cdot)$ can be taken to be a very general function.

In the next sub-section we will see how to modify $U_{p|y}$ and $\sigma_{p|y}^2$ when the means of P and y are unknown.

B. Unknown Means

When U_p and U_y are not known the selection rule that minimized loss can not usually be found (i.e., if one insists that U_p and U_y are fixed). Fortunately, it is usually possible to mimic this selection rule when means are unknown. For example, if estimates for U_p and U_y are available, the practitioner might use the estimates as if they were known. However, such a scheme can be criticised on grounds of sensitivity to errors associated with the estimated means. To avoid some of the problems related to sensitivity, it is best to increase $\sigma_{p|y}^2$ so that in some way an accounting is made for the precision of estimated means. It would then be more reasonable for the practitioner to use means as if they were known.

Assume that y contains information that can be used to estimate U_p and U_y . In particular, let $U_p = \mathbf{t}'\mathbf{X}\mathbf{b}$ and $U_y = \mathbf{X}\mathbf{b}$ where \mathbf{t} is a known column vector, \mathbf{X} is a known full column rank matrix and \mathbf{b} is a column vector of unknown fixed effect ^(c).

Consider \mathbf{b} as a vector of normal random variables even though it is not. Let

$$E(\mathbf{b}) = \mathbf{U}_b \text{ and } \text{Var}(\mathbf{b}) = \mathbf{D} \quad [6]$$

(c) It may seem unduly restrictive to assume that the mean of a future observation (U_p) is a linear combination of the means of past observations (U_y). However, if U_p can not be estimated from data then U_p can be thought of as a random effect with its own mean and variance. Thus, appropriate modifications can be made in model specification.

where \mathbf{D} is a diagonal matrix. With \mathbf{U}_b and \mathbf{D} given, the machinery described for known means can be implemented in a straight-forward manner. Because \mathbf{U}_b may not be close to \mathbf{b} , it is best to pick the diagonal elements of \mathbf{D} to be large. In this way the subjective variation we assign to \mathbf{b} reflects our confidence in \mathbf{U}_b . If we have no confidence in \mathbf{U}_b it is reasonable to let the diagonal elements of \mathbf{D} go to infinity. In this case \mathbf{b} can take on any value with equal likelihood. The posterior distribution of \mathbf{P} given \mathbf{y} exists in the limit as the diagonal elements of \mathbf{D} go to infinity. Moreover the limiting distribution does not involve \mathbf{U}_b . Thus, it is reasonable to use the limiting distribution to evaluate [4] via procedures already described. The only new things needed are the mean and variance of \mathbf{P} given \mathbf{y} as diagonal elements of \mathbf{D} go to infinity.

The strategy just described is a common Bayesian method. The limiting distribution used for \mathbf{b} is called an improper prior. Because this prior assigns equal likelihood to all possible realizations of \mathbf{b} , the prior is frequently referred to as noninformative or vague. A formal generalization of the Bayes decision rule for the improper prior is straight-forward and is given in BERGER (1980, p. 116). From the point of view of robustness, use of the improper prior is generally very reasonable. Unfortunately, there are situations where use of an improper prior is not very satisfactory (BERGER, 1980, pp. 152-155).

Using [6], the means and variances given earlier for \mathbf{P} and \mathbf{y} are changed to

$$E \begin{bmatrix} \mathbf{P} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{t}'\mathbf{X}\mathbf{U}_b \\ \mathbf{X}\mathbf{U}_b \end{bmatrix} \text{ and } \text{Var} \begin{bmatrix} \mathbf{P} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} r + \mathbf{t}'\mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t} & \mathbf{d}' + \mathbf{t}'\mathbf{X}\mathbf{D}\mathbf{X}' \\ \mathbf{d} + \mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t} & \mathbf{V} + \mathbf{X}\mathbf{D}\mathbf{X}' \end{bmatrix}$$

Thus, by standard selection index theory

$$U_{ply} = \mathbf{t}'\mathbf{X}\mathbf{U}_b + (\mathbf{d} + \mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t})' (\mathbf{V} + \mathbf{X}\mathbf{D}\mathbf{X}')^{-1} (\mathbf{y} - \mathbf{X}\mathbf{U}_b) \tag{7}$$

and

$$\sigma_{ply}^2 = r + \mathbf{t}'\mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t} - (\mathbf{d} + \mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t})' (\mathbf{V} + \mathbf{X}\mathbf{D}\mathbf{X}')^{-1} (\mathbf{d} + \mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t}). \tag{8}$$

The limiting values of U_{ply} and σ_{ply}^2 are derived in Annex C. The limiting value of U_{ply} is

$$[\mathbf{d}' \{ \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \} + \mathbf{t}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{y}. \tag{9}$$

A generalized least squares estimate of U_y (say \hat{U}_y) is given by $\mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Moreover, an estimate of U_p (say \hat{U}_p) is given by $\mathbf{t}'\hat{U}_y$. Thus, [9] can be written as

$$\hat{U}_p + \mathbf{d}'\mathbf{V}^{-1} (\mathbf{y} - \hat{U}_y).$$

This expression is directly analogous to the standard selection index with known means. However, the limiting value of σ_{ply}^2 is

$$r - \mathbf{d}'\mathbf{V}^{-1}\mathbf{d} + \mathbf{d}'\mathbf{V}^{-1}\mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{d} + \mathbf{t}'\mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{t} - 2\mathbf{t}\mathbf{X} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{d}. \tag{10}$$

Terms other than $r - \mathbf{d}'\mathbf{V}^{-1}\mathbf{d}$ in [10] can be regarded as corrections that were needed due to estimation of unknown means.

In theory, [9] and [10] can be evaluated in order to find the U_{ply} and σ_{ply}^2 that are needed to determine [4]. However, the formulae in their current form are very awkward and actual evaluation of [9] and [10] may be prohibitive. Fortunately, U_{ply} and σ_{ply}^2 can be found using alternative formulae.

If P and y can be described jointly by a suitable linear model, [9] will lead naturally to the mixed model equations (HENDERSON, 1973). Moreover, [10] can be expressed using machinery associated with mixed model methodology. These results are not surprising given the correspondence between mixed model methodology and Bayesian estimation (DEMPFLE, 1977). The mixed model is generally used to estimate genetic quantities. However, the problem at hand requires estimation of a phenotype. Mixed model methodology must be employed with this subtle difference in mind.

Write $P = t'Xb + k'u + e$, where $t'Xb$ was defined earlier, k is a known column vector, u is a column vector of random effects and e is a random variable that is stochastically independent of y , u and b . Assume that the variance of e (say σ_e^2) is known and that $E[e] = 0$. Using the terminology of HENDERSON (1975), U_{ply} is the best linear unbiased predictor of $t'Xb + k'u$ and σ_{ply}^2 is σ_e^2 plus the variance of the error of prediction of $t'Xb + k'u$.

Determining σ_{ply}^2 via mixed model procedures involves computing inverse elements of the coefficient matrix described by HENDERSON (1975). In practice this step may be prohibitive. We acknowledge that approximations for σ_{ply}^2 may be useful.

IV. Example

In this section, theory described earlier will be applied to a mate selection problem. Throughout our example we will assume additive inheritance.

Assume that a dairy farmer wants to mate two bulls (Sire 1 and Sire 2) to two cows (Cow 1 and Cow 2). He decides not to use the same bull twice. Thus, he must choose one of the two mating schemes. These are :

Scheme 1 : Sire 1 \times Cow 1 ; Sire 2 \times Cow 2

Scheme 2 : Sire 1 \times Cow 2 ; Sire 2 \times Cow 1.

Each mating scheme will result in two progeny. The farmer wishes to use the scheme that corresponds to progeny with the highest expectation of total merit.

Merit on female progeny will be taken to be a simple function of the phenotypes for milk yield and rear leg set. No merit will be assigned to male progeny. The merit function for females is

$$.0037 (\text{milk}) + 2.56 (\text{set}) - .017 (\text{set})^2 \quad [13]$$

where milk is the 305 day mature equivalent milk yield measured in Kg, set is linear type trait score (50 to 99) (THOMPSON *et al.*, 1983) depicting the rear leg side view set. The merit expression [13] was constructed from survey data and was provided by GONYON (personal communication, 1984). It can be argued that merit should be a function of more than just milk and set. For simplicity we will ignore this.

Genetic evaluations for Sire 1 and Sire 2 and phenotypic measurements taken from Cow 1 and Cow 2 are provided in Table 1. The herd average for milk and set will be assumed to be 7 258 kg and 76.6, respectively. These quantities are clearly realistic (e.g. EVERETT *et al.*, 1976 ; THOMPSON *et al.*, 1983). The herd averages will be assumed known without error and directly applicable given the information in Table 1. Thus, the expected phenotype for any progeny can be obtained by adding the herd average, sire ETA and dam ETA. An implicit assumption is that the genetic base corresponding to the sire evaluations is assumed to equal the average genetic level of the herd.

TABLE 1

*Information to be used in the mate selection problem.
Données relatives au problème du choix des conjoints.*

Sire information :				
	ETA milk ^(a,b)	r_{TI}^2 ^(c)	ETA set ^(b)	r_{TI}^2 ^(c)
Sire				
1	+ 226	.51	- 2.28	.25
2	+ 210	.79	- .10	.67
Dam information ^(d) :				
Cow				
1	+ 134	.016	+ .98	.006
2	+ 79	.016	- 1.05	.006

(a) Kg.

(b) Estimated transmitting ability (ETA) expressed as a deviation from the genetic base (kg).

(c) Square of the correlation between the transmitting ability and the estimated transmitting r_{TI}^2 .

(d) Observations were + 488 and + 287 for milk ; and + 13 and - 14 for set, respectively.

The heritability (h^2) and phenotypic standard deviation (σ_p) for milk yield will be taken as .25 and 907 kg, respectively. The heritability and phenotypic standard deviation for set will be taken to be equal to estimates published by THOMPSON *et al.* (1983). These values are .15 and 6.7, respectively.

Assume that each sire has equal probability of producing female calves. Then without loss in generality, all calves produced via schemes 1 and 2 can be taken as female. This convention will be used. Thus, the expected merit of any particular progeny can be found by determining the conditional expectation of [13] given the information in Table 1.

In order to determine the expectation of [13], the conditional means and variances for phenotypes expressed on particular progeny must be found. Assume that the phenotypic and genetic correlations between milk and set are null. This assumption is probably wrong (THOMPSON *et al.*, 1983), however it is used only to simplify the discussion and notation. Given the assumption, the conditional expectation of any phenotype (milk, set) for a particular progeny is

$$\text{Herd average} + \text{sire ETA} + \text{dam ETA}$$

where the transmitting abilities of the sire and dam can be found in Table 1. Likewise, the conditional variance of this phenotype is

$$(1 - 1/4r_{T1}^2h^2 - 1/4h^4) \sigma_p^2$$

where r_{T1}^2 is a measure of the precision associated with the transmitting ability of the sire and it can be found in Table 1. The computed conditional means and variances for each progeny produced by schemes 1 and 2 are listed in Table 2.

The expectation of [13] for any progeny can be found by using the quantities given in Table 2 in accordance to the formula

$$.0037 (U_m) + 2.56 (U_s) - .017 (U_s^2 + V_s)$$

where U_m is the conditional mean for milk, U_s is the conditional mean for set and V_s is the conditional variance for set. Note that the conditional variance for milk is not needed. The expectation of [13] for progeny produced by the mating schemes are given in Table 2.

The values in Table 2 suggest that scheme 1 is better than scheme 2. The differences in expected merit are not dramatic. This is due to the relatively flat merit function for set ^(d).

It is possible to incorporate into the decision process information on maternal grandsires. This type of decision is probably more realistic than the example given here. However, information on any maternal grandsire would only contribute in a small way to the corresponding total phenotype.

TABLE 2

*Conditional means and variances for the phenotypes and the expected merit of progeny.
Espérances et variances conditionnelles des valeurs phénotypiques
et fonctions d'objectif de la descendance.*

	U_m ^(a)	V_m ^(b)	U_s ^(c)	V_s ^(d)	Expected merit
Scheme 1 :					
Sire 1 × Cow 1	7 618	7.84	75.3	44.2	123.8
Sire 2 × Cow 2	7 547	7.69	75.5	43.5	123.6
Total					247.4
Scheme 2 :					
Sire 1 × Cow 2	7 563	7.84	73.3	44.2	123.5
Sire 2 × Cow 1	7 602	7.69	77.5	43.5	123.7
Total					247.2

(a) U_m = conditional mean for milk.

(b) V_m = conditional variance for milk $\times 10^{-7}$.

(c) U_s = conditional mean for set.

(d) V_s = conditional variance for set.

(d) This observation is a little artificial. A reasonable measure of utility can be taken as $k_1T + k_2$ for any $k_1 > 0$ and k_2 . Decisions resulting from the use of $k_1T + k_2$ are the same as those resulting from the use of T . Any deviation observed in the expectation of $k_1T + k_2$ can be made to look small by taking k_1 to be small and k_2 to be large.

V. Conclusion

In the previous example the importance of milk in selection decisions was removed because each sire and dam would produce one offspring regardless of the selection alternative (thus the example does not display selection) and because of the linear contribution of milk to merit. However, the value of milk production seems to dominate mate selection rules when merit is a function of milk and several type traits (ALLAIRE *et al.*, 1984). In this study an attempt was made to use realistic genetic parameters and a realistic merit function. This suggests that « corrective mating » as practiced in dairy cows may be improper.

In this paper we have ignored ways of estimating the merit function. However, we have implied that merit is directly related to some monetary measure. Thus, it may be possible to estimate the merit function by a regression equation where the dependent variable is measured in monetary units. Whereas this seems reasonable, it is bending theory. More formally, the total merit function ($M(\cdot)$) for the selected population can be estimated via utility theory (BERGER, 1980). In this setting $M(\cdot)$ reflects an individual's gambling philosophy when phenotypic expressions are at stake. From a theoretical perspective $M(\cdot)$ need not be representable as a sum of identical merit functions (i.e., T) corresponding to individuals in the selected population. However, it seems practical to assume that such a representation exists and that utility theory can be used to estimate the component functions (i.e., T) of $M(\cdot)$. Even with appropriate modifications, estimating T by utility theory can be criticized due to nonobjectivity. However, BERGER (1980, p. 58) claims that such a criticism is « silly » because decisions pertaining to uncertainties are personal choices and thus nonobjective anyway.

Received July 24, 1984

Accepted January 3, 1985

Acknowledgements

This research was supported in part by Holstein Association of America (Brattleboro, VT), Noba, Inc. (Tiffin, OH) and Ohio Dairy Farmers Federation. The authors wish to thank Dr. D. GIANOLA (University of Illinois, U.S.A.), G.B. JANSEN (University of Guelph, Canada) and the reviewers for their useful remarks.

References

- ABRAMOWITZ M., STEGUN I.A., 1972. — *Handbook of mathematical functions with formulas, graphs and mathematical tables*. 1046 pp., U.S. Department of Commerce.
- ALLAIRE F.R., 1980. — Mate selection by selection index theory. *Theor. Appl. Genet.*, **57**, 267-272.
- ALLAIRE F.R., SMITH S.P., SHOOK J.E., JOHNSON L.P., 1984. — Improving an aggregate phenotype in replacements by selecting their sires conditioned on dam phenotypes (Submitted to *J. Dairy Sci.*).
- BERGER J.O., 1980. — *Statistical decision theory*. 425 pp., Springer-Verlag, New York, Inc.

- BULMER M.G., 1980. — *The mathematical theory of quantitative genetics*. 255 pp., Oxford, Oxford University Press.
- DEMPFLE L., 1977. — Relation entre BLUP (best linear unbiased prediction) et estimateurs bayésiens. *Ann. Génét. Sél. Anim.*, **9**, 27-32.
- EVERETT R.W., KEOWN J.F., CLAPP E.E., 1976. — Production and stayability trends in dairy cattle. *J. Dairy Sci.*, **59**, 1505-1510.
- FERNANDO R.L., GIANOLA D., 1984. — Optimum rules for selection (Submitted to *Biometrics*).
- GODDARD M.E., 1983. — Selection indices for non-linear profit functions. *Theor. Appl. Genet.*, **64**, 339-344.
- GOFFINET B., 1983. — Selection on selected records. *Génét. Sél. Evol.*, **15**, 91-97.
- HARRIS D.L., 1964. — Genotype covariances between inbred relatives. *Genetics*, **50**, 1317-1348.
- HENDERSON C.R., 1973. — Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Lush, Jay Dr. Blacksburg, Virginia, July 29, 1972*, 10-41, A.S.A.S.-A.D.S.A., Champaign, Illinois.
- HENDERSON C.R., 1975. — Comparison of alternative sire evaluation methods. *J. Anim. Sci.*, **41**, 760-770.
- HENDERSON C.R., 1976. — A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, **32**, 69-83.
- HENDERSON H.V., SEARLE S.R., 1981. — On deriving the inverse of a sum of matrices. *Siam Review*, **23**, 53-60.
- JANSEN G.B., WILTON J.W., 1984. — Selecting mating pairs with linear programming techniques. *J. Dairy Sci.*, **67**, suppl. 1, 246.
- KAPLAN W., 1973. — *Advanced calculus*. 2nd edition, 184-185. Addison-Wesley Publishing Company.
- MOAV R., HILL W.G., 1966. — Selection sire and dam lines. 4. Selection within lines. *Anim. Prod.*, **8**, 375-390.
- NETER J., WASSERMAN W., 1974. — *Applied linear statistical models: Regression, analysis of variance and experimental design*. 313-315. Richard D. Irwin, Inc.
- PFAFFENBERGER R.C., WALKER D.A., 1976. — *Mathematical programming for economics and business*. 275-311. The Iowa State University Press.
- STOER J., BULIRSCH R., 1980. — *Introduction to numerical analysis*. 609 pp., Springer-Verlag, New York, Inc.
- THOMPSON J.R., LEE K.L., FREEMAN A.E., JOHNSON L.P., 1983. — Evaluation of a linearized type appraisal system for *Holstein* cattle. *J. Dairy Sci.*, **66**, 325-331.
- VAN VLECK D., 1979. — *Notes on the theory and application of selection principles for the genetic improvement of animals*. 248 pp., Cornell University.
- WILTON J.W., EVANS D.A., VAN VLECK L.D., 1968. — Selection indices for quadratic models of total merit. *Biometrics*, **24**, 937-949.

Annex A

A. Evaluating Expected Merit of Gene Pools

The Problem: we will describe how to evaluate [4] given the assumptions of multivariate normality and additive inheritance^(e). In this case \mathbf{y} contains information associated with parents that will contribute genes to a gene pool. Also \mathbf{P} is a

(e) In theory it is possible to use a model that incorporates more complicated gene action. Gene pools will usually be inbred. Thus, determining necessary covariances will be complicated (HARRIS, 1964).

hypothetical phenotype randomly created from additive genetic effects (representing genes from gene pool) and environmental factors. Write P as :

$$P = a_1 + a_2 + E$$

where a_1 equals the average additive genetic effects of the parents, a_2 is the additive genetic effect due to segregation in the random mating population (or gene pool) and E is the environmental effects. The conditional mean and variance of $a_1 + E$ can be found directly using mixed model procedures. Note that a_1 is a linear combination of parental additive genetic effects (these will usually be effects in the linear model). Likewise E will usually be represented as a linear combination of effects in the model plus a random residual (this residual is stochastically independent from y and other relevant terms). The term a_2 has mean zero and it is stochastically independent from y , a_1 and E. Thus, if we know the variance of a_2 we can find the conditional mean and variance of P using mixed model procedures. To find the variance of a_2 it suffices to construct a relationship matrix involving a sufficient number of animals in the analysis plus the hypothetical animal. The diagonal element (times the additive genetic variance) corresponding to the hypothetical animal will equal the unconditional variance (UV) of $a_1 + a_2$. The variance of a_2 can be found by subtracting the UV of a_1 from the UV of $a_1 + a_2$. The UV of a_1 can be found in a straight forward manner. We will only show how to compute the appropriate relationship matrix.

BULMER (1980, p. 197) described a selection rule that can be used to improve nonlinear merit in outbreeding populations. Because he was not very explicit it is difficult to tell whether he attempted to solve the gene pool problem as we defined it. Nevertheless BULMER's procedure is very similar to the one proposed here (select those parent where [2] is minimized). BULMER's procedure is different at least in one way because he seems to assume that the variance of a_2 is constant across all selection alternatives. This may be a minor issue in practice.

B. Genomic Tabular Method

For many cases the tabular method (Van VLECK, 1979) can be used to compute the relationship matrix. However, we will propose a genomic tabular method because this procedure can be adapted to our problem in a conceptually simple manner. Unlike the relationship matrix, every element in the genomic table is a probability. Moreover, when building the genomic table inbreeding can be ignored. The inverse genomic table can be computed (if one wants it) using shortcut procedures very similar to the methods that HENDERSON (1976) described for the relationship matrix. The genomic tabular method can also be adapted to non-diploid individuals (e.g., bees are diploid or haploid). The only disadvantage of the genomic table is that it is usually 4 times larger than the relationship matrix.

We will describe the genomic tabular method by example. Assume that animals A and B are mated to produce animal C. Assume that the genomes that animals A and B received from their parents are unrelated. Animals A, B and C contribute the genes to a gene pool. Animals B and C are of the same sex. Thus, A contributes twice as many genes as B or C. Animals B and C each contribute same amount of genes. Assume that animal D is created from the gene pool. The genomic table is presented in Table 3.

TABLE 3
Genomic table.

		A		B		C		D	
		A ₁	A ₂	B ₁	B ₂	C ₁	C ₂	D ₁	D ₂
A	A ₁	1	0	0	0	1/2	0	5/16	5/16
	A ₂	0	1	0	0	1/2	0	5/16	5/16
B	B ₁	0	0	1	0	0	1/2	3/16	3/16
	B ₂	0	0	0	1	0	1/2	3/16	3/16
C	C ₁ = A	1/2	1/2	0	0	1	0	3/8	3/8
	C ₂ = B	0	0	1/2	1/2	0	1	1/4	1/4
D	D ₁ = 1/2A + 1/2BC .	5/16	5/16	3/16	3/16	3/8	1/4	1	9/32
	D ₂ = 1/2A + 1/2BC .	5/16	5/16	3/16	3/16	3/8	1/4	9/32	1

Observations. Each row or column of Table 3 corresponds to a genomic group. Each animal has two genomic groups. Define A₁ and A₂ to be the first and second genomic groups in animal A. Define similar quantities for animals B, C, and D.

The letters on the top (or on the left hand side) identify the animals. The table is set up such that animal symbols to the left (or top) correspond to older animals than symbols to the right (or bottom). The symbols below (or to the right of) animal symbols identifies the genomic groups. The genomic groups for any animal are adjacent and ordered (first, second). The parentage of genomic groups are identified by the codes following the equal signs in the second column. The genomic group code A for C₁ means that C₁ was derived from animal A. The code 1/2A + 1/2BC for D₁ (or D₂) means that half of D₁ (or D₂) was derived from A and the other half was derived from B and C.

Any element in Table 3 equals the probability that a gene on a particular locus from one genomic group is equal by descent to another gene at the same locus for a different (or the same) genomic group. For example, the probability is 1/2 that genes corresponding to some locus are equal in A₁ and C₁ (this probability can be found in two places, i.e., the genomic table is symmetric). Note that the diagonal elements are all one. This simply says that the probability that genomes are equal to themselves is unity.

The additive relationship matrix is obtained by partitioning the genomic table into 2 by 2 blocks (corresponding to animals) and combining the 4 elements in each block and dividing by 2. Note that animal D is 9/32 inbred.

How the Table was constructed. To construct a genomic table initially add one to all diagonals. Next add zeros to all off diagonals corresponding to genomic groups in the base population. For our example the animals A and B are the base population. The remaining elements are now computed by recursion. The recursion formula uses elements in a row to compute elements to the right in the same row. Thus, the

elements must be determined from left to right. Use the recursion starting with the top row. The recursion is identified by the parentage code. The symbol A corresponding to C_1 indicates that the 2 elements (in the appropriate row) listed under animal symbol A are averaged. This number is put in the table under C_1 . The symbol $1/2A + 1/2BC$ corresponding to D_1 (or D_2) indicates that elements listed under animal symbol A are averaged and in a separate calculation the 4 elements listed under animal symbols B and C are averaged. Finally the computed averages are each weighted by $1/2$ and combined. This number is put in the table under D_1 (or D_2). After the row is completely determined fill in the column that is determined by symmetry. Then return to the row directly below the row that was previously evaluated and compute its elements. Never use a recursion directly to compute elements below the diagonal. These elements should always come from calculations that were made to find elements above the diagonal.

The recursion formulae are easy to derive. Each probability is related back to probabilities that involve the parentage of the youngest genomic group (or of equal age). Consider for example the probabilities associated with A_1 and C_1 . The parentage of C_1 is animal A. Half of the genes in C_1 come from A_1 and the other half come from A_2 . These events are equally likely and are mutually exclusive. If the gene in question from C_1 comes from A_1 , the probability of identity is 1. If the gene comes from A_2 the probability is 0. Thus, the probability we are looking for is $1/2*1 + 1/2*0 = 1/2$.

C. Approximate Solution to the Gene Pool Problem

The gene pool problem is very hard to solve. We will suggest a procedure to find an approximate solution given that we have an initial group of parents that might contribute genes to a gene pool. The initial group can be improved if we substitute one of the parents with some other candidate such that [2] is reduced. We might use that candidate that reduces [2] the most. Next we do the same substitution for a different parent and continue the process to a third parent or a fourth, etc. We should continue in an iterative way until [2] can not be reduced any more by substitution of any individual parent in our solution.

This procedure need not solve the gene pool problem. The solution that we get may depend on the initial group of parents and the order parents are considered for substitution. However, the algorithm will find a choice of parents that reduces [2] relative to the initial group of parents.

Annex B

A. Gaussian Quadrature

With Gaussian Quadrature (STOER & BULIRSCH, 1980, pp. 142-151) [4] is approximated by

$$\sum_{i=1}^s w_i f(x_i \sqrt{2} \sigma_{ply} + U_{ply}) / \sqrt{2\pi} \quad [14]$$

where s is a user selected integer, x_i , $i = 1, 2, \dots, s$, are the roots of the s^{th} order Hermite polynomial and w_i , $i = 1, 2, \dots, s$, are the associated « weights ». The x_i and

$w_i, i = 1, 2, \dots, s$, are tabulated and can be found in ABRAMOWITZ & STEGUN (1972). The difference between [14] and [4] is equal to

$$\frac{d^{2s}f(z)}{d^{2s}z} \Big|_{z=z'} = \frac{(2\sigma_{p|y}^2)^s 2^{s-1/2} s!}{(2s)!} \tag{15}$$

for some z' (i.e., if [15] exists). If the absolute value of [15] is small for all z' in $(-\infty, +\infty)$, Gaussian quadrature will yield a good approximation. However, as an indicator of the precision of Gaussian quadrature, the upper bound of the absolute value of [15] may be too pessimistic (STOER & BULIRSCH, 1980, p. 151).

B. Expectation of Piece-Wise Cubic

Assume that $-\infty = t_0 < t_1 < t_2 \dots < t_s = \infty$ and let $f(P) = a_{0i} + a_{1i}P + a_{2i}P^2 + a_{3i}P^3$, if P is in $[t_i, t_{i+1})$. Then [4] after some simplification equals

$$\sum_{i=0}^{s-1} \sum_{j=0}^3 a_{ij} V_{ji} \tag{16}$$

where $V_{ij} = \int_{t_i}^{t_{i+1}} \frac{P^j \exp[-(1/2)(P - U_{p|y})^2 / \sigma_{p|y}^2]}{\sqrt{2\pi} \sigma_{p|y}} dP$.

The terms, V_{0i}, V_{1i}, V_{2i} and V_{3i} , can be computed together via recursion. Given these terms, evaluation of [16] is straight forward. The formulae are given below :

Define $C_1 = (t_i - U_{p|y}) / \sigma_{p|y}, C_2 = (t_{i+1} - U_{p|y}) / \sigma_{p|y}, \phi(y) = \exp(-(1/2)y^2) / \sqrt{2\pi}$ and $\Phi(u) = \int_{-\infty}^u \exp(-(1/2)y^2) / \sqrt{2\pi} dy$.

Next, compute the quantities

$$\begin{aligned} f_0 &= \Phi(C_2) - \Phi(C_1), \\ f_1 &= \phi(C_1) - \phi(C_2), \\ f_2 &= C_1\phi(C_1) - C_2\phi(C_2) + \Phi(C_2) - \Phi(C_1) \text{ and} \\ f_3 &= C_1^2\phi(C_1) - C_2^2\phi(C_2) + 2\phi(C_1) - 2\phi(C_2). \end{aligned}$$

By convention, $\phi(C) = C\phi(C) = C^2\phi(C) = 0$ if $C = +\infty$ or $-\infty$ and $\Phi(C) = 1$ (or 0) if $C = +\infty$ (or $-\infty$). Finally set,

$$\begin{aligned} V_{0i} &= f_0, \\ V_{1i} &= U_{p|y}f_0 + \sigma_{p|y}f_1, \\ V_{2i} &= U_{p|y}^2f_0 + 2U_{p|y}\sigma_{p|y}f_1 + \sigma_{p|y}^2f_2 \text{ and} \\ V_{3i} &= U_{p|y}^3f_0 + 3U_{p|y}^2\sigma_{p|y}f_1 + 3U_{p|y}\sigma_{p|y}^2f_2 + \sigma_{p|y}^3f_3. \end{aligned}$$

Annex C

A. Limiting Value of Conditional Mean

From HENDERSON & SEARLE (1981)

$$(\mathbf{V} + \mathbf{XDX}')^{-1} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}. \quad [17]$$

Hence, by substitution U_{ply} (i.e., [7]) equals

$$\mathbf{t}'\mathbf{XU}_b + (\mathbf{d} + \mathbf{XDX}'\mathbf{t})' \{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\} (\mathbf{y} - \mathbf{XU}_b).$$

Consider only that part of the U_{ply} given by

$$\mathbf{t}'\mathbf{XDX}' \{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\}. \quad [18]$$

[18] can be written as

$$\mathbf{t}'\mathbf{X} \{\mathbf{D} - \mathbf{DX}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{D} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\}\mathbf{X}'\mathbf{V}^{-1}.$$

Moreover, it is easy to show that

$$\mathbf{D} - \mathbf{DX}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = (\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \quad [19]$$

Thus [18] equals

$$\mathbf{t}'\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \quad [20]$$

and consequently U_{ply} can be written as

$$\mathbf{t}'\mathbf{XU}_b + [\mathbf{d}' \{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\} + \mathbf{t}'\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] (\mathbf{y} - \mathbf{XU}_b) \quad [21]$$

It can be shown that the limiting value of [21] as diagonals of \mathbf{D} go to infinity, can be obtained by dropping \mathbf{D}^{-1} . Thus, in the limit U_{ply} is

$$\mathbf{t}'\mathbf{XU}_b + [\mathbf{d}' \{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\} + \mathbf{t}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}] (\mathbf{y} - \mathbf{XU}_b). \quad [22]$$

Now,

$$\mathbf{d}' \{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\} \mathbf{XU}_b = 0$$

and

$$\mathbf{t}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{XU}_b = \mathbf{t}'\mathbf{XU}_b.$$

Thus [22] can be written as

$$[\mathbf{d}' \{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\} + \mathbf{t}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{y}.$$

B. Limiting Value of Conditional Variance

We can write σ_{ply}^2 (i.e., [8]) as

$$r + \mathbf{t}'\mathbf{XDX}'\mathbf{t} - \mathbf{d}'(\mathbf{V} + \mathbf{XDX}')^{-1}\mathbf{d} - 2\mathbf{t}'\mathbf{XDX}'(\mathbf{V} + \mathbf{XDX}')^{-1}\mathbf{d} - \mathbf{t}'\mathbf{XDX}'(\mathbf{V} + \mathbf{XDX}')^{-1}\mathbf{XDX}'\mathbf{t} \quad [23]$$

Using relation [17] the term $-\mathbf{d}'(\mathbf{V} + \mathbf{XDX}')^{-1}\mathbf{d}$ can be written as

$$-\mathbf{d}' \{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\}\mathbf{d}. \quad [24]$$

Consider now parts of terms in [23] given by

$$\mathbf{t}'\mathbf{X}\mathbf{D}\mathbf{X}'(\mathbf{V} + \mathbf{X}\mathbf{D}\mathbf{X}')^{-1} \quad [25]$$

This term is equal to [18] following substitution with relation [17]. Thus, [25] is equal to [20]. Pre-multiplying [20] by -2 and post-multiplying [20] by \mathbf{d} yields

$$-2\mathbf{t}'\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{d} \quad [26]$$

which is one of the terms in [23].

Post-multiplying [20] by $-\mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t}$ produces

$$-\mathbf{t}'\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t} \quad [27]$$

which is another term in [23]. Combining the term $\mathbf{t}'\mathbf{X}\mathbf{D}\mathbf{X}'\mathbf{t}$ from [23] with [27] yields after rearranging

$$\mathbf{t}'\mathbf{X}\{\mathbf{D} - (\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{D}\}\mathbf{X}'\mathbf{t}.$$

This expression simplifies to

$$\mathbf{t}'\mathbf{X}(\mathbf{D}^{-1} + \mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{t} \quad [28]$$

which was found by substitution as suggested from the transpose of identity [19].

We have shown that [23] is equal to r plus [24] plus [26] plus [28]. The limiting value of $\sigma_{p|y}^2$ if found by dropping \mathbf{D}^{-1} from the term [24], [26] and [28]. In the limit $\sigma_{p|y}^2$ is

$$r - \mathbf{d}'\{\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\}\mathbf{d} - 2\mathbf{t}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{d} + \mathbf{t}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{t}$$