

Chapter 6

Dekkers and Gibson

Economic Selection Indexes

In economically oriented breeding programs, the trait that we want to improve could be called economic merit. The **breeding objective** of our program is then to maximize improvement of economic merit. Economic merit might be defined in different ways, e.g. as profit per animal, profit per enterprise, economic efficiency, or something else. We will return to this problem in later Chapters. For the present, it is only necessary to recognize that the **breeding objective** is a general statement of the economic genetic goal of the breeding program.

For a given definition of the breeding objective, there will likely be several or many traits that contribute to the objective. The **aggregate genotype** is then defined as a function of the additive genetic values of the traits of interest of an individual, which if selected upon would achieve the breeding objective. The function need not necessarily be linear, but in many cases an approximate linear relationship can be found which adequately defines aggregate genotype over the range of genetic values encountered (see Chapter 7). If the function is a linear function, then the **aggregate genotype**, H , can be written as

$$H = v_1 g_1 + v_2 g_2 \dots v_n g_n = \mathbf{v}' \mathbf{g} \quad (6.1)$$

where g_i is the additive genetic value of the i^{th} trait and v_i is the economic value of genetically improving that trait. Note that v_i is a partial economic weight, that is, it is the economic value of genetically improving the i^{th} trait, when all other traits remain unchanged.

In vector notation, $\mathbf{v}' = [v_1, v_2, \dots, v_n]$ and $\mathbf{g}' = [g_1, g_2, \dots, g_n]$.

In practice, additive genetic values are not known. However we can record each individual's performance for a number of traits. Observations on these traits can then be combined into a **selection index**, I of the form,

$$I = b_1 x_1 + b_2 x_2 \dots b_m x_m = \mathbf{b}' \mathbf{x} \quad (6.2)$$

where x_i is an observation on the i^{th} trait and b_i is the selection index coefficient (or weight) for that trait. In vector notation: $\mathbf{b}' = [b_1, b_2, \dots, b_m]$ and $\mathbf{x}' = [x_1, x_2, \dots, x_m]$.

General Derivation of Optimal Weights
Selection Index Theory

Economic Breeding Objective/Goal

$H = v_1 g_1 + v_2 g_2 + \dots + v_n g_n = \mathbf{v}' \mathbf{g}$

g_i = true bv for trait i
 v_i = economic value trait i $H=g$ for single trait EBV

Selection Index

$I = b_1 x_1 + b_2 x_2 + \dots + b_m x_m = \mathbf{b}' \mathbf{x}$

x_i = phenotypic record
 b_i = index weight

Derive weights to maximize
- response in H: $R_H = b_{H,i}(I - \bar{I})$
- accuracy of I: $r_{i,H}$

The problem is then to estimate the selection index coefficients, b_i , such that selection of individuals on their **selection index** value, I , maximizes response in the **aggregate genotype**, H . The selection index methods described in Chapter 4 can be used to derive such indexes based on:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gv} \quad (6.3)$$

It is worth noting at this point that the traits recorded and which appear in the index do not need to be, and often are not, the same traits as those that appear in the aggregate genotype. As a crude example, consider a terminal sire line of pigs. Assume that the profitability of this line is (approximately) a linear function of carcass weight and lean percentage at 120 days of age. Thus the aggregate genotype would be written as:

$$H = v_1 g_{(\text{carcass weight})} + v_2 g_{(\text{lean percentage})}$$

Neither carcass weight nor lean percentage can be recorded directly in live pigs, but we could record live weight at 120 days and ultrasonically estimated back fat depth at 120 days. The selection index would then take the form,

$$I = b_1 x_{(\text{live weight})} + b_2 x_{(\text{ultrasonic back fat depth})}$$

In this case, neither of the two traits recorded actually appear in the aggregate genotype, but both can be expected to be closely related to the traits in the aggregate genotype.

6.1 Predicting Response to Selection and Related Parameters

If we were interested in a practical problem of what weight to give a series of observations on a particular population of animals for selection, and we were certain of our phenotypic and genetic parameters contained in \mathbf{P} and \mathbf{G} and were sure of our economic weights in \mathbf{v} , then we might stop here. But curiosity alone would likely prompt us to ask what the predicted variance of the index would be, what would be the predicted response of the aggregate genotype to selection on the index, and what would be the predicted correlated response of each of the traits in the aggregate genotype? And, mere curiosity aside, such predictions are essential when comparing different possible indexes, for assessing whether predicted responses of individual traits are likely to be acceptable to the users of the index or to their customers, and for determining whether it is worthwhile to record data on a given trait for inclusion in the index.

In section 4.4.1.2 the variance of the index, σ_I^2 , was derived as $\sigma_I^2 = \mathbf{b}'\mathbf{P}\mathbf{b}$ (6.4)

It is important to realize that this equation holds whatever values of \mathbf{b} are used, i.e. not only for the optimal \mathbf{b} derived using selection index theory, but for any arbitrary vector \mathbf{b} . However, if \mathbf{b} is the optimal set of index coefficients, then $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gv}$, which when substituted into (6.4) gives:

$$\sigma_I^2 = \mathbf{b}'\mathbf{P} \mathbf{P}^{-1} \mathbf{Gv} = \mathbf{b}' \mathbf{Gv} \quad (6.5)$$

This second form is most often quoted as the variance of a selection index. But it should always be remembered that this holds only for the "optimal" selection index. Since you may wish to explore the consequences of using sub-optimal indexes, it is probably safer to use the form given in equation (6.4), which holds for any selection index, optimal or sub-optimal.

Applying the general equation for response to selection, to selection on I to improve H , predicted genetic superiority for the aggregate genotype, H , to selection on the index, I , is given by:

$$S_H = i r_{HI} \sigma_H \quad (6.6)$$

As given in 4.4.1.2, the variance of the breeding goal can be derived as:

$$\sigma_H^2 = \mathbf{v}' \mathbf{C} \mathbf{v} \quad (6.7)$$

where \mathbf{C} is an $n \times n$ matrix of genetic covariances among the traits in the aggregate genotype.

Similarly, it follows that $\sigma_{HI} = \mathbf{b}' \mathbf{G} \mathbf{v}$ (6.8)

Hence, $r_{HI} = \frac{\sigma_{HI}}{\sigma_I \sigma_H} = \frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b} \mathbf{v}' \mathbf{C} \mathbf{v}}} \quad (6.9)$

and $S_H = i \frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}} \quad (6.10)$

Note again, all the above derivations, (6.6) to (6.10), apply to any selection index (that is any set of b values) not just the optimum index. However, if we are dealing with the optimal index, where $\mathbf{b} = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}$, then as noted at equations (6.4) and (6.5) $\mathbf{b}' \mathbf{P} \mathbf{b} = \mathbf{b}' \mathbf{G} \mathbf{v}$ and substituting into equation (6.10) gives:

$$S_H = i \sqrt{\mathbf{b}' \mathbf{G} \mathbf{v}} \quad (6.11)$$

We could have also obtained this optimal selection response directly, as $S_H = i b_{HI} \sigma_I$

and recalling that for the optimal selection index: $b_{HI} = 1$, hence: $S_H = i \sigma_I$

and substituting in the variance of the optimal index defined at (6.5): $S_H = i \sqrt{\mathbf{b}' \mathbf{G} \mathbf{v}}$ (6.12)

The accuracy of an index is defined as its correlation with the aggregate genotype, i.e. r_{HI} , as estimated at (2.17). For an optimal index, noting the equivalence of equations (2.13) and (2.14), the expectation for the accuracy of the index can also be written as

$$r_{HI} = \sqrt{\frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\mathbf{v}' \mathbf{C} \mathbf{v}}} \quad (6.13)$$

The expected change in the additive genetic value of the i^{th} trait in the aggregate genotype due to selection on the index, S_{g_i} , can be found as a correlated response to selection on I as:

$$S_{g_i} = b_{g_i I} S_I = b_{g_i I} i \sigma_I = i \frac{\sigma_{g_i I}}{\sigma_I^2} \sigma_I = i \frac{\sigma_{g_i I}}{\sigma_I} \quad (6.14)$$

Where

$$\sigma_{g_i I} = \text{cov}(g_i, \mathbf{b}' \mathbf{x}) = \mathbf{b}' \text{cov}(g_i, \mathbf{x}) = \mathbf{b}' \mathbf{G}_i \quad (6.15)$$

where \mathbf{G}_i denotes the i^{th} column of \mathbf{G} :

Hence for all indexes:

$$\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_i, \dots, \mathbf{G}_n]$$

$$S_{g_i} = i \frac{\mathbf{b}' \mathbf{G}_i}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}} \quad (6.16)$$

$$\text{For an optimum index, the solution is also } S_{g_i} = i \frac{\mathbf{b}' \mathbf{G}_i}{\sqrt{\mathbf{b}' \mathbf{G} \mathbf{v}}} \quad (6.17)$$

It also follows directly from (6.17), that the vector of genetic responses of each of the traits in the aggregate genotype, can be found as: $\mathbf{S}_g = [S_{g_1}, \dots, S_{g_i}, \dots, S_{g_n}] = i \frac{\mathbf{b}' \mathbf{G}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}} \quad (6.18)$

These derivations of the principal parameters defining a selection index for all indexes and for optimal indexes are summarized in Table 6.1

Table 6.1 Summary of selection index formulae for any index and for optimal indexes.

Parameters	Derivation	
	Any Index	Optimal Index
b	Arbitrary	$\mathbf{P}^{-1} \mathbf{G} \mathbf{v}$
σ_I^2	$\mathbf{b}' \mathbf{P} \mathbf{b}$	$\mathbf{b}' \mathbf{G} \mathbf{v}$
σ_H^2	$\mathbf{v}' \mathbf{C} \mathbf{v}$	$\mathbf{v}' \mathbf{C} \mathbf{v}$
σ_{HI}	$\mathbf{b}' \mathbf{G} \mathbf{v}$	$\mathbf{b}' \mathbf{G} \mathbf{v}$
r_{HI}	$\frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b} \mathbf{v}' \mathbf{C} \mathbf{v}}}$	$\sqrt{\frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\mathbf{v}' \mathbf{C} \mathbf{v}}} = \sqrt{\frac{\mathbf{b}' \mathbf{P} \mathbf{b}}{\mathbf{v}' \mathbf{C} \mathbf{v}}} = \frac{\sigma_I}{\sigma_H}$
S_H	$i \frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}}$	$i \sqrt{\mathbf{b}' \mathbf{G} \mathbf{v}} = i \sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}} = i \sigma_I$
S_{g_i}	$i \frac{\mathbf{b}' \mathbf{G}_i}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}}$	$i \frac{\mathbf{b}' \mathbf{G}_i}{\sqrt{\mathbf{b}' \mathbf{G} \mathbf{v}}}$
S_g	$i \frac{\mathbf{b}' \mathbf{G}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}}$	$i \frac{\mathbf{b}' \mathbf{G}}{\sqrt{\mathbf{b}' \mathbf{G} \mathbf{v}}}$

The correlation between index values of two relatives, i and j , which is needed to, e.g., compute selection intensities as in section 3.6, can be computed as:

$$t = \text{corr}(I_i, I_j) = \text{corr}(\mathbf{b}' \mathbf{x}_i, \mathbf{b}' \mathbf{x}_j) = \frac{\mathbf{b}' \text{cov}(\mathbf{x}_i, \mathbf{x}_j) \mathbf{b}}{\mathbf{b}' \mathbf{P} \mathbf{b}} = \frac{\mathbf{b}' \mathbf{R} \mathbf{b}}{\mathbf{b}' \mathbf{P} \mathbf{b}}$$

where \mathbf{R} is a matrix with covariances between information sources on the two relatives (De Boer and Van Arendonk, 1989).

6.2 Example of Economic Index Selection

A beef breed is to be used as a terminal sire and economic analysis has shown that three key traits are post-weaning gain (PWG) with an economic value of 370 \$/kg/d, ultrasonic back fat depth (BF), with an economic value under the local payment system of -20 \$/mm, and feed intake, with an economic value of -50 \$/kg/d. Only PWG and BF are recorded in the bull-testing program. Genetic and economic parameters are summarized in Figure 6.1.

Figure 6.1. Example Derivation of Economic Index Beef Cattle Terminal Sire Line			
	Post Weaning Gain (PWG)	Back Fat (BF)	Feed Intake (FI)
Econ value	370 \$/kg/d	-20 \$/mm	-50 \$/kg/d
h^2	0.40	0.44	0.20
σ_p	0.23 kg/d	0.15 mm	0.5 kg/d
r_g			
r_p	PWG	BF	FI
PWG		0.32	0.70
BF	0.18		0.48
FI	0.23	0.40	
			Select on index of own phenotype for PWG + BF

Matrices \mathbf{P} , \mathbf{G} , and \mathbf{C} have the following elements:

$$\mathbf{P} = \begin{bmatrix} \sigma_{p_{PWG}}^2 & \sigma_{p_{PWG}, p_{BF}} \\ \sigma_{p_{PWG}, p_{BF}} & \sigma_{p_{BF}}^2 \end{bmatrix} \quad (6.19)$$

$$\mathbf{G} = \begin{bmatrix} \sigma_{p_{PWG}, g_{PWG}} & \sigma_{p_{PWG}, g_{BF}} & \sigma_{p_{PWG}, g_{FI}} \\ \sigma_{p_{BF}, g_{PWG}} & \sigma_{p_{BF}, g_{BF}} & \sigma_{p_{BF}, g_{FI}} \end{bmatrix} \quad (6.20)$$

$$\mathbf{C} = \begin{bmatrix} \sigma_{g_{PWG}}^2 & \sigma_{g_{PWG}, g_{BF}} & \sigma_{g_{PWG}, g_{FI}} \\ \sigma_{g_{BF}, g_{PWG}} & \sigma_{g_{BF}}^2 & \sigma_{g_{BF}, g_{FI}} \\ \sigma_{g_{FI}, g_{PWG}} & \sigma_{g_{FI}, g_{BF}} & \sigma_{g_{FI}}^2 \end{bmatrix} \quad (6.21)$$

The elements of \mathbf{P} , \mathbf{G} , and \mathbf{C} can be found by recalling that

$$r_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_1^2 \sigma_2^2}}$$

hence

$$\sigma_{12} = r_{12} \sigma_1 \sigma_2$$

and that

$$\sigma_g^2 = h^2 \sigma_p^2$$

giving

$$\mathbf{P} = \begin{bmatrix} 0.0529 & 0.00621 \\ 0.00621 & 0.0225 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 0.02116 & 0.004632 & 0.02277 \\ 0.004632 & 0.0099 & 0.01068 \end{bmatrix}$$

and

$$\mathbf{C} = \begin{bmatrix} 0.02116 & 0.004632 & 0.02277 \\ 0.004632 & 0.0099 & 0.01068 \\ 0.02277 & 0.01068 & 0.05 \end{bmatrix}$$

The vector of economic weights is: $\mathbf{v} = \begin{bmatrix} 370 \\ -20 \\ -50 \end{bmatrix}$

Hence, the index weights would be $\mathbf{b} = \mathbf{P}^{-1} \mathbf{G} \mathbf{v} = \begin{bmatrix} 123.6 \\ 9.52 \end{bmatrix}$

The index accuracy would be

$$r_{HI} = \sqrt{\frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\mathbf{v}' \mathbf{C} \mathbf{v}}} = 0.6214$$

The expected genetic superiority for the breeding goal to 1 standard deviation selection on the index is:

$$S_H = \frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}} = 28.722$$

Expected genetic superiorities for PWG, BF, and FI to one standard deviation selection on the

index are:

$$S_g = \frac{\mathbf{b}'\mathbf{G}}{\sqrt{\mathbf{b}'\mathbf{G}\mathbf{v}}} = \begin{bmatrix} 0.0926 \\ 0.0232 \\ 0.1015 \end{bmatrix}$$

Thus, one standard deviation of selection on the index would be expected to yield bulls with an average breeding value of +0.0926 kg/day for PWG, + 0.0232 mm for BF, and +.1015 kg/day more FI. Also, their expected breeding value for profit is +\$28.72.

6.3 Economic Indexes Based on Estimated Breeding Values

So far we have dealt with economic selection indexes that were based on phenotypic records. In these cases, the main interest is to predict the individual's breeding value for the aggregate genotype. In practice, however, the sources of information that are available to develop the economic index are EBV for individual traits, rather than phenotypic records:

$$I = b_1 \hat{g}_1 + b_2 \hat{g}_2 + \dots + b_m \hat{g}_m = \mathbf{b}' \hat{\mathbf{g}}_I$$

This leads to a step-wise procedure for development of economic selection indexes, in which the first step consists of predicting BV for individual traits and the second step of combining the resulting EBV into an economic index. An advantage of this step-wise approach is that it allows different breeders to put different emphasis on traits in the aggregate genotype, while utilizing the most accurate EBV for the component traits.

To derive the optimal weight that should be placed on EBV, consider a vector of available phenotypic records \mathbf{x} that can be subdivided into subvectors \mathbf{x}_i which correspond to phenotypic records on trait i for $i=1, \dots, m$:

$$\mathbf{x}' = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_m]$$

and a breeding goal with n traits: $H = v_1 g_1 + v_2 g_2 + \dots + v_n g_n = \mathbf{v}' \mathbf{g}_H$

Then, an economic selection index based on the full set of phenotypic records \mathbf{x} can be derived as:

$$I_F = \mathbf{b}_F' \mathbf{x}$$

with the vector of index weights derived using standard selection index theory from:

$$\mathbf{b}_F = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}$$

Note that this index maximizes accuracy of predicting H given the available phenotypic records.

In the two-step approach, the objective is to develop the following type of index:

$$I_S = b_1 \hat{g}_1 + b_2 \hat{g}_2 + \dots + b_m \hat{g}_m = \mathbf{b}_S' \hat{\mathbf{g}}_I$$

Consider first the situation where the traits in the index and the breeding goal are the same. Thus, $m=n$ and $\mathbf{g}_H = \mathbf{g}_I = \mathbf{g}$.

The first step is to derive EBV for each individual trait i based on all available data \mathbf{x} : $\hat{g}_i = \mathbf{b}_i' \mathbf{x}$

Index weights for this EBV can be derived as: $\mathbf{b}_i = \mathbf{P}^{-1} \mathbf{G}_i$

where \mathbf{G}_i is the i^{th} column of matrix \mathbf{G} , i.e. a vector representing the genetic covariances between the observations in \mathbf{x} and the i^{th} trait in the aggregate genotype g_i :

$$\mathbf{G} = [\mathbf{G}_1; \mathbf{G}_2; \dots; \mathbf{G}_n]$$

Because all EBV are based on the same data vector \mathbf{x} , the selection index equations for the individual EBV can be combined to directly estimate the vector of EBV as:

$$\hat{\mathbf{g}} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n]' = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n]' \mathbf{x}$$

Combining the selection index equations that are used to derive each set of index weights \mathbf{b}_i we get:

$$\hat{\mathbf{g}} = \{\mathbf{P}^{-1}[\mathbf{G}_1; \mathbf{G}_2; \dots; \mathbf{G}_n]\}' \mathbf{x} = (\mathbf{P}^{-1} \mathbf{G})' \mathbf{x}$$

Using these EBV to develop the economic index we get:

$$I_S = \mathbf{b}_S' \hat{\mathbf{g}} = \mathbf{b}_S' (\mathbf{P}^{-1} \mathbf{G})' \mathbf{x} = (\mathbf{P}^{-1} \mathbf{G} \mathbf{b}_S)' \mathbf{x}$$

Note that, if we set $\mathbf{b}_S = \mathbf{v}$ then $I_S = (\mathbf{P}^{-1} \mathbf{G} \mathbf{v})' \mathbf{x} = \mathbf{b}_F' \mathbf{x} = I_F$ because $\mathbf{b}_F = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}$ based on the one-step approach

Note that this proves that the optimal index can be obtained by weighting the EBV by the economic weight for that trait in the breeding goal.

Note, however, that this only holds if the same traits are included in the breeding goal and the index and if the EBV are based on all available data on all traits, i.e. multiple-trait EBV are used.

If the traits in the breeding goal and index are not the same, i.e. $m \neq n$ and $\mathbf{g}_H \neq \mathbf{g}_I$, then the index weights can be derived by partitioning the breeding goal into a component that is related to the traits in the index and an uncorrelated residual:

$$H = \mathbf{v}' \mathbf{g}_H = \mathbf{v}' (\mathbf{b}_{\mathbf{g}_H \mathbf{g}_I}' \mathbf{g}_I + \mathbf{e})$$

where $\mathbf{b}_{g_H g_I}$ is a vector of regression coefficients for the regression of vector \mathbf{g}_H on \mathbf{g}_I . Vector $\mathbf{b}_{g_H g_I}$ can be derived using standard multiple regression methods as:

$$\mathbf{b}_{g_H g_I} = \mathbf{C}_I^{-1} \mathbf{C}_{IH}$$

where \mathbf{C}_I is the genetic variance/covariance matrix among the traits which appear in the index and \mathbf{C}_{IH} is the genetic covariance matrix between traits in the index and traits in the aggregate genotype.

The residual vector \mathbf{e} is by definition uncorrelated to traits in the index and does, therefore, not need to be considered when deriving the index. Thus, considering only the first term, the new breeding goal is one for which the assumption $\mathbf{g}_H = \mathbf{g}_I$ holds but with a new set of economic

values:

$$H^* = \mathbf{v}^* \mathbf{g}_I = \mathbf{v}' \mathbf{b}_{g_H g_I} \mathbf{g}_I'$$

Thus

$$\mathbf{v}^* = \mathbf{b}_{g_H g_I} \mathbf{v}$$

and the optimal weights on the EBV for index $I_S = \mathbf{b}_S' \hat{\mathbf{g}}$ can be derived as:

$$\mathbf{b}_S = \mathbf{C}_I^{-1} \mathbf{C}_{IH} \mathbf{v}$$

An alternative derivation of this same result was given by Schneeberger et al. (1992).

Note that when $\mathbf{g}_H = \mathbf{g}_I$, i.e. when the same traits are included in H and I , $\mathbf{C}_I = \mathbf{C}_{IH}$ and \mathbf{b}_S simplifies to the original result: $\mathbf{b}_S = \mathbf{v}$.

It is interesting to note that the index weights for indexes that contain EBV depend only on the economic values and genetic parameters. They do not depend on the accuracy of the individual EBV. Thus, the same index can be used for all animals.

It is important to realize, however, that these results only hold if all observations on all traits are used to estimate the BV for all traits. This would be the case when multiple-trait genetic evaluation models are used. The previous equations do not hold when trait EBV are derived from single-trait evaluation models, for which only data on trait i are used to estimate \hat{g}_i . The reason is that single-trait evaluation methods do not consider covariances between trait records, whereas covariances are considered in the multiple-trait evaluation methods. However, if the single trait EBV have high accuracy, index weights can still be approximated by the economic values because, with high accuracy, correlated trait information has little impact on EBV and, thus, single-trait EBV approximate multi-trait EBV.

If single trait EBV do not have high accuracy, approximate weights on EBV must be derived using selection index procedures using: $\mathbf{b}_S = \mathbf{P}_S^{-1} \mathbf{G}_S \mathbf{v}$

where \mathbf{P}_S = matrix with (co-)variances among single trait EBV, \hat{g}_i

\mathbf{G}_S = matrix with covariances of EBV \hat{g}_i with true BV of traits in the breeding goal

Elements of \mathbf{P}_S and \mathbf{G}_S must now be approximated by specifying a vector \mathbf{x}_i of sources of phenotypic records for each EBV \hat{g}_i . Standard selection index theory can then be used to derive

each EBV as:

$$\hat{g}_i = \mathbf{b}_i' \mathbf{x}_i \quad \text{with} \quad \mathbf{b}_i = \mathbf{P}_{ii}^{-1} \mathbf{G}_{ii}$$

where \mathbf{P}_{ii} is the (co-)variance matrix for records in \mathbf{x}_i , and \mathbf{G}_{ii} is the vector of covariances of g_i with \mathbf{x}_i .

Then, the covariance between two EBV can be computed as:

$$\sigma_{\hat{g}_i, \hat{g}_j} = \text{cov}(\mathbf{b}_i' \mathbf{x}_i, \mathbf{b}_j' \mathbf{x}_j) = \mathbf{b}_i' \text{cov}(\mathbf{x}_i, \mathbf{x}_j) \mathbf{b}_j = \mathbf{b}_i' \mathbf{P}_{ij} \mathbf{b}_j$$

where \mathbf{P}_{ij} is the (co-)variance matrix between records in \mathbf{x}_i and \mathbf{x}_j

and the covariance between an EBV and a true BV can be computed as:

$$\sigma_{\hat{g}_i, g_j} = \text{cov}(\mathbf{b}_i' \mathbf{x}_i, \mathbf{g}_j) = \mathbf{b}_i' \text{cov}(\mathbf{x}_i, \mathbf{g}_j) = \mathbf{b}_i' \mathbf{G}_{ij}$$

where \mathbf{G}_{ij} is a vector with covariances of \mathbf{x}_i with g_j .

Note that in contrast to indexes based on multi-trait EBV, weights on single-trait EBV depend on the accuracy of the EBV and on the sources of information that contribute to each animal's EBV. Index weights will, therefore, differ from animal to animal. Depending on parameters and the range of accuracies, it may however be possible to use a single index for all animals, with the index derived based on average amounts of information. Methods to evaluate the loss in accuracy when using a single index will be described in the next section.

One issue that we have not yet discussed is how to derive the accuracy of economic selection indexes that are based on EBV. When the index is based on single-trait EBV, accuracy can be

derived using standard selection index theory as: $r_{HI} = \sqrt{\frac{\mathbf{b}_S' \mathbf{G}_S \mathbf{v}}{\mathbf{v}' \mathbf{C} \mathbf{v}}}$

For the index based on multi-trait EBV, although index weights do not depend on the amount of information that goes into each EBV, the accuracy of the index does. For multi-trait EBV indexes, accuracy must, therefore, be derived by specifying the sources of information in the multiple-trait vector of observations \mathbf{x} . Then, accuracy can be obtained by deriving the accuracy

of the full index: $r_{HI} = \sqrt{\frac{\mathbf{b}_F' \mathbf{G} \mathbf{v}}{\mathbf{v}' \mathbf{C} \mathbf{v}}}$

This is illustrated in the following example where selection is based on observations on growing pigs for growth rate (GR, g/d) on the individual (x_{GR}) and the mean of 5 full sibs (for \bar{x}_{GR}) and own performance for feed intake (x_{FI} , g/d). The phenotypic standard deviation is 100 for GR and 200 for FI, and the h^2 for both traits is 0.25. The phenotypic correlation is 0.6 while the genetic correlation is 0.8.

The aggregate genotype is $H = 0.2 g_{GR} - 0.05 g_{FI}$

The selection index weights for prediction of the aggregate genotype using the full index with all three sources of information (x_{GR} , \bar{x}_{GR} , x_{FI}) can be calculated as:

$$\mathbf{b}_F = \mathbf{P}^{-1}\mathbf{Gv} = \begin{bmatrix} 10000 & 1250 & 12000 \\ 1250 & 3000 & 2000 \\ 12000 & 2000 & 40000 \end{bmatrix}^{-1} \begin{bmatrix} 2500 & 4000 \\ 1250 & 2000 \\ 4000 & 10000 \end{bmatrix} \begin{bmatrix} 0.2 \\ -0.05 \end{bmatrix} = \begin{bmatrix} 0.02874 \\ 0.04011 \\ -0.00313 \end{bmatrix}$$

Weights \mathbf{b}_1 for the prediction of the breeding value for growth rate using all data are:

$$\mathbf{b}_1 = \mathbf{P}^{-1}\mathbf{G}_1 = \begin{bmatrix} 10000 & 1250 & 12000 \\ 1250 & 3000 & 2000 \\ 12000 & 2000 & 40000 \end{bmatrix}^{-1} \begin{bmatrix} 2500 \\ 1250 \\ 4000 \end{bmatrix} = \begin{bmatrix} 0.1702 \\ 0.3239 \\ 0.0327 \end{bmatrix}$$

Similarly, weights for feed intake using all data, \mathbf{b}_2 , are:

$$\mathbf{b}_2 = \mathbf{P}^{-1}\mathbf{G}_2 = \begin{bmatrix} 10000 & 1250 & 12000 \\ 1250 & 3000 & 2000 \\ 12000 & 2000 & 40000 \end{bmatrix}^{-1} \begin{bmatrix} 4000 \\ 2000 \\ 10000 \end{bmatrix} = \begin{bmatrix} 0.1061 \\ 0.4935 \\ 0.1935 \end{bmatrix}$$

As demonstrated previously, the multiple-trait EBV can be combined into an economic index to predict the aggregate genotype with index weights equal to the economic values:

$$I_S = 0.2 \hat{g}_{GR} - 0.05 \hat{g}_{FI}$$

Substituting the multiple-trait index weights \mathbf{b}_1 and \mathbf{b}_2 gives:

$$I_S = 0.2 \begin{bmatrix} 0.1702 \\ 0.3239 \\ 0.0327 \end{bmatrix} \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} - 0.05 \begin{bmatrix} 0.1061 \\ 0.4935 \\ 0.1935 \end{bmatrix} \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} = \begin{bmatrix} 0.02874 \\ 0.04011 \\ -0.00312 \end{bmatrix} \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} = \mathbf{b}_F \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix}$$

proving that the index based on multiple-trait EBV is equivalent to the full index.

Let us now look at a situation where breeding values for GR and FI are predicted with a univariate model, i.e. the EBV for GR is based on observations for GR only and similarly the EBV for FI is based entirely on observations for FI. The following index weights result:

$$\mathbf{b}_1 = \mathbf{P}_{11}^{-1} \mathbf{G}_{11} = \begin{bmatrix} 10000 & 1250 \\ 1250 & 3000 \end{bmatrix}^{-1} \begin{bmatrix} 2500 \\ 1250 \end{bmatrix} = \begin{bmatrix} 0.2088 \\ 0.3297 \end{bmatrix}$$

Since the single-trait EBV for FI is based only on own performance, the index weight is equal to heritability: $\mathbf{b}_2 = \mathbf{P}_{22}^{-1} \mathbf{G}_{22} = [40000]^{-1} [10000] = [0.25]$

Combining these single-trait EBV into an economic index to predict H results in:

$$\begin{aligned} \mathbf{b}_S &= \mathbf{P}_S^{-1} \mathbf{G}_S \mathbf{v} = \begin{bmatrix} \sigma_{\hat{g}_{GR}}^2 & \sigma_{\hat{g}_{GR}, \hat{g}_{FI}} \\ \sigma_{\hat{g}_{GR}, \hat{g}_{FI}} & \sigma_{\hat{g}_{FI}}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{\hat{g}_{GR}, g_{GR}} & \sigma_{\hat{g}_{GR}, g_{FI}} \\ \sigma_{\hat{g}_{FI}, g_{GR}} & \sigma_{\hat{g}_{FI}, g_{FI}} \end{bmatrix} \mathbf{v} \\ &= \begin{bmatrix} \mathbf{b}_1' \mathbf{P}_{11} \mathbf{b}_1 & \mathbf{b}_1' \mathbf{P}_{12} \mathbf{b}_2 \\ \mathbf{b}_2' \mathbf{P}_{21} \mathbf{b}_1 & \mathbf{b}_2' \mathbf{P}_{22} \mathbf{b}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{b}_1' \mathbf{G}_{11} & \mathbf{b}_1' \mathbf{G}_{12} \\ \mathbf{b}_2' \mathbf{G}_{21} & \mathbf{b}_2' \mathbf{G}_{22} \end{bmatrix} \mathbf{v} \\ &= \begin{bmatrix} 934.184 & 791.25 \\ 791.25 & 2500 \end{bmatrix}^{-1} \begin{bmatrix} 934.125 & 1494.6 \\ 1000 & 2500 \end{bmatrix} \begin{bmatrix} 0.2 \\ -0.05 \end{bmatrix} = \begin{bmatrix} 0.1292 \\ -0.0109 \end{bmatrix} \end{aligned}$$

These weights on single-trait EBV differ from those on multiple-trait EBV, which are equal to the economic values. Multiplying the weights out gives:

$$I_S = 0.1292 \begin{bmatrix} 0.2088 \\ 0.3297 \\ 0 \end{bmatrix} \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} - 0.0109 \begin{bmatrix} 0 \\ 0 \\ 0.25 \end{bmatrix} \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} = \begin{bmatrix} 0.02698 \\ 0.04261 \\ -0.00272 \end{bmatrix} \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} = \mathbf{b}_F^{**} \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix}$$

These weights are different from those for the optimal index and, thus, the index derived from single-trait EBV is not optimal. The reason for the suboptimality is that, when deriving the single-trait EBV, relative weights put on records are determined while only considering a single trait and not the whole aggregate genotype. In our present example, this resulted in a weight of 0.2088 on x_{GR} and of 0.3297 on \bar{x}_{GR} , a ratio of $0.2088/0.3297=0.633$. This same ratio of weights is still present in the overall index ($0.02698/0.04261=0.633$). In the optimal index, however, the ratio of weights on x_{GR} versus \bar{x}_{GR} is $0.02874/0.04011=0.7165$.

Although the single-trait EBV index is not optimal, an important question is how much we would lose in accuracy when using this index instead of the optimal index. This can be studied by evaluating the accuracy of both indexes.

Accuracy of the multiple-trait EBV index must be derived from the full index with the three sources of information:

$$r_{HI} = \sqrt{\frac{\mathbf{b}_F^* \mathbf{Gv}}{\mathbf{v}' \mathbf{Cv}}} = 0.5518$$

The variance of the aggregate genotype $\mathbf{v}' \mathbf{Cv}$ equals 45 in this example.

The accuracy of the single-trait EBV index can be derived in two ways:

- 1) as the accuracy of a sub-optimal index with the three information sources:

$$r_{HI} = \sqrt{\frac{\mathbf{b}_F^* \mathbf{Gv}}{\mathbf{b}_F^* \mathbf{P} \mathbf{b}_F^* \mathbf{v}' \mathbf{Cv}}}$$

- 2) or as the accuracy of an optimal index based on the two single-trait EBV

$$r_{HI} = \sqrt{\frac{\mathbf{b}_S^* \mathbf{G}_S \mathbf{v}}{\mathbf{v}' \mathbf{Cv}}}$$

Both result in accuracy to be equal to 0.5511, which is only slightly smaller than the accuracy of the optimal index, which was 0.5518. Thus, in this case, use of single-trait EBV had limited impact on efficiency.

As an example of the use of principles similar to those outlined above, Veerkamp et al. (1995) investigated sensitivity to economic values and genetic parameters of a sire selection index of type and production traits for a breeding goal with production traits and longevity. They also investigated the efficiency of using index weights derived for an index with multiple-trait EBV in indexes that included single-trait rather than multiple-trait EBV and in indexes that included EBV from two multiple-trait evaluations, one for production traits and one for type traits. Results showed that losses in efficiency were less than 1% for EBV based on at least 50 daughter records, which confirms the robustness of selection indexes for progeny-tested dairy sires.

6.4 Sensitivity of Selection Indexes to Estimates of Variances and Covariances

As described so far, the selection index provides a method to maximize selection response for a given aggregate genotype when a given set of observations are available. It is assumed that the variances and covariances that make up the elements of \mathbf{P} , \mathbf{G} and \mathbf{C} are known without error.

In practice, elements of \mathbf{P} , \mathbf{G} and \mathbf{C} are estimated with error, and are often obtained from several different sources. Sales and Hill (1976) deal in detail with the ways in which errors in the estimates of variances and covariances affect efficiency of selection. For the present purposes, it is sufficient to note that, when the elements of \mathbf{P} and \mathbf{G} are not known without error, the selection index is often relatively insensitive to errors in estimates of these elements. However, it is always wise to examine how sensitive the resulting index is to the assumed elements of \mathbf{P} and \mathbf{G} . (Note that the elements of \mathbf{C} do not affect selection index coefficients nor the prediction of response to selection.)

Bouchez and Goffinet (1990) propose a method for eliminating traits (or other sources of information) from the index based on maximizing an approximate estimate of the mean square error of prediction of the index, allowing for the inaccuracy of estimates of variances and covariances among traits. Their method deals with the situation where parameters are estimated from the same population as is being selected. Their method does not appear to increase selection response but involving fewer traits, is claimed to be more robust than a full index. An alternative approach would be to use Bayesian methods to simultaneously derive genetic parameters and selection indexes. Bayesian methods allow uncertainty about parameters to be included in development of the index. Literature estimates of genetic parameters can be included through specification of proper priors. Bayesian analyses are, however, typically complex in nature and an alternative series of simpler tests and modifications is proposed here.

6.4.1 Testing Consistency of Variance/Covariance Matrixes

Whenever a selection index is constructed, the **P** and **C** matrixes should be tested for being positive definite. A non-positive definite variance/covariance matrix may indicate that one or more traits in the index are linear functions of combinations of other traits, or that certain elements of the matrix are not possible given that the other elements are correct. For example, consider the phenotypic covariances among three traits with standardized variances such that, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. Then, if $\sigma_{12} = 0.9$ and $\sigma_{13} = 0.9$, common sense would tell us that σ_{23} must also take a relatively high value. A value of $\sigma_{23} = -0.5$ would clearly be impossible. It can be

shown in this case that if

$$\mathbf{P} = \begin{bmatrix} 1 & .9 & .9 \\ .9 & 1 & \sigma_{23} \\ .9 & \sigma_{23} & 1 \end{bmatrix}$$

P becomes non-positive definite when $\sigma_{23} \leq 0.62$. At this point **P** also becomes singular and an inverse to **P** (i.e. \mathbf{P}^{-1}) cannot be found.

All matrix programming languages carry sub-routines to calculate the eigenvalues of a matrix. There are as many eigenvalues as there are rows and columns of the matrix (i.e. an $n \times n$ matrix has n eigenvalues). If one or more of the eigenvalues is negative, the matrix is non-positive definite.

Matrices which are close to being non-positive definite should also be examined closely because such matrixes are very close to being singular and are only just within the allowable parameter space, indicating possible inconsistencies in parameter values.

As noted by Sales and Hill (1976) and Hayes and Hill (1980), matrixes are more likely to become inconsistent (non-positive definite) as the number of traits considered increases, the size of the data base used to obtain estimates decreases, and when estimates are obtained from different sources. This in turn can lead to increasing instability and uncertainty over the resulting indexes and their responses. In general, this argues for keeping the number of traits in both the

index and the aggregate genotype down to the minimum number consistent with achieving effective genetic progress.

The **P** and **C** matrixes referred to above are those for observations in the index and traits in the aggregate genotype respectively. Often the elements of **P** are themselves derived from population phenotypic and genetic parameters. For example, observations in the selection index might be full-sib mean performance for a series of traits. Then, the variances and covariances among these observations would be derived from the phenotypic and genetic variances and covariances among single phenotypic records. We can denote \mathbf{P}_O and \mathbf{G}_O as the phenotypic and genetic variance covariance matrixes among single phenotypic records of traits appearing in the selection index, which we used to derive the elements of **P**. Then, if our records are to be biologically meaningful, \mathbf{P}_O and \mathbf{G}_O should also be within the allowable parameter space and should thus be tested for being positive-definite.

Similarly, if our model is

$$y = g + e$$

then

$$\mathbf{E}_O = \mathbf{P}_O - \mathbf{G}_O$$

is the implied environmental variance/covariance matrix among the traits in the index, given \mathbf{P}_O and \mathbf{G}_O . Again, to be consistent and biologically meaningful, \mathbf{E}_O should also be positive definite.

Variance/covariance matrixes can be biologically and economically inconsistent despite being positive definite. For example, a matrix for the three traits, milk yield (M), fat yield (F) and fat concentration (f) could be positive definite but not conform to the expectation that, since $f = \frac{F}{M}$, the variance of f and its covariances with F and M are entirely dependent on the

variances and covariances among F and M . This is because the test of positive definiteness examines possible linear combinations of traits and does not recognize the possibility of there being a pre-specified linear or non-linear relationship, such as f being a ratio of F/M . In general, any trait that is a direct function of other traits in the index (or the aggregate genotype) will carry little or no additional information and as such is usually best omitted.

The economic weight of a trait is defined as the marginal change in economic value given a genetic change in that trait while holding all other traits in the aggregate genotype constant. Thus any trait in the aggregate genotype which is entirely a function of other traits in the aggregate genotype has an indeterminate economic value since it is impossible to make genetic change in that trait without making genetic change in other traits. As an example, again considering the traits M , F and f , where $f = \frac{F}{M}$, it is quite clear that there cannot be a change in f without a change in either F or M or both. Thus, if F and M appear in the aggregate genotype, f should not appear.

It is surprising how often these basic checks for consistency of variance/covariance components are not made in examples appearing in the literature, sometimes causing quite misleading results and conclusions. It is thus best not only to check over your own parameters for consistency, but

also to check parameters in published papers, unless the authors have quite clearly stated that the appropriate checks have been made.

In many cases, by elimination of unnecessary and less useful traits in the index, matrixes will become positive definite. Where matrixes remain non-positive definite, Hayes and Hill (1981) present a mathematical method ("bending") for altering parameters to obtain positive definite matrixes. This method assumes there is no additional information available about parameter estimates and as such is unlikely to be generally appropriate. An alternative is to adjust individual parameter estimates based on an assessment of reliability of the estimates. An example and further discussion is given in the following section. But there is certainly room for further investigation of this problem.

6.4.2 Tests of Sensitivity of Indexes to Assumed Variances and Covariances

If the various variance/covariance matrixes pass the test of being positive definite and being consistent in other ways, there is still a concern that the index derived may be sensitive to the assumed values of the variance/covariance components. A test of sensitivity would be to ask what proportion of the maximum response would we expect in the aggregate genotype if we used one set of variances to derive our index coefficients when another set of covariances was the correct set.

If we use the subscript u to describe the result of a particular set of parameters used and t to describe the set defined as true, then

$$\mathbf{b}_u = \mathbf{P}_u^{-1} \mathbf{G}_u \mathbf{v}$$

and

$$\mathbf{b}_t = \mathbf{P}_t^{-1} \mathbf{G}_t \mathbf{v}.$$

If \mathbf{P}_t and \mathbf{G}_t are the true variance/covariance matrixes, then, by definition, selection using index coefficients, \mathbf{b}_t , will give the maximum expected response in the aggregate genotype H . Thus use of \mathbf{b}_u will give less response than \mathbf{b}_t . We can then define the efficiency of using \mathbf{b}_u instead of \mathbf{b}_t as the ratio of genetic superiority of selected individuals in H when selecting using \mathbf{b}_u

compared to when selecting using \mathbf{b}_t , i.e. $E_{ut} = \frac{S_{H_u}}{S_{H_t}}$

S_{H_u} can be found from the equation in Table 6.1 that defines response to selection from an arbitrary index and S_{H_t} from the equation which defines response to selection from the optimal index, so that

$$S_{H_u} = i \frac{\mathbf{b}'_u \mathbf{G}_t \mathbf{v}}{\sqrt{\mathbf{b}'_u \mathbf{P}_t \mathbf{b}_u}} \quad \text{and} \quad S_{H_t} = i \sqrt{\mathbf{b}'_t \mathbf{G}_t \mathbf{v}}$$

and thus

$$E_{ut} = \frac{\mathbf{b}'_u \mathbf{G}_t \mathbf{v}}{\sqrt{\mathbf{b}'_u \mathbf{P}_t \mathbf{b}_u}} \cdot \frac{1}{\sqrt{\mathbf{b}'_t \mathbf{G}_t \mathbf{v}}}$$

The process would be to define a parameter set to be our best possible estimate based on the evidence available and to use this as \mathbf{G}_t and \mathbf{P}_t . Alternative sets would then be defined and the question asked of each alternative set, "if the alternative set were indeed true, how well would the index based on our best estimates perform?" There will likely be some difficulty in deciding what are realistic alternative sets of parameters. Clearly, if parameters are well estimated, there would be little relevance in examining widely different parameter sets. Whereas, if we are very uncertain of our estimated parameter set, we may well wish to examine widely different alternative sets. Also, the problem becomes worse the larger the number of elements in \mathbf{P} and \mathbf{G} , since the number of possible combinations of alternative parameter sets rapidly expands with increasing dimensions of \mathbf{P} and \mathbf{G} .

If there is insufficient evidence to say that the parameters used are the "best available" or "most likely", it may be worth considering an alternative set of parameters if one can be found that gives an index that is less sensitive to the parameters being incorrect than use of the original set of parameters. This could be judged by creating a matrix of relative efficiencies where each parameter set was tested against all other parameter sets, i.e.

$$\mathbf{E} = \begin{bmatrix} E_{11} & E_{12} & E_{1n} \\ E_{21} & E_{22} & E_{2n} \\ E_{n1} & E_{n2} & E_{nn} \end{bmatrix} \begin{matrix} \text{'true' parameters} \\ \text{used parameters} \end{matrix}$$

where the rows correspond to parameter sets used and the columns to the parameter sets assumed to be correct, so that: $E_{ij} = \frac{S_{H_{u_i}}}{S_{H_{t_j}}}$ where i and j are the used and true sets of parameters, resp.

The diagonal elements of \mathbf{E} are 1.0, all other elements are <1.0 . The parameter set could then be chosen for which the elements of the corresponding row of \mathbf{E} were as close to 1.0 as possible, i.e. the parameter set which showed the least sensitivity to being incorrect in relation to the alternative parameter sets examined.

These processes of investigation are rather subjective. Indeed there is scope for research into the general area of how best to obtain estimates of population parameters, how to perform sensitivity analyses and how to modify parameters for particular applications in animal breeding. Bayesian methods provide opportunities to address these issues.

There is also the question of what action to take if our results are indeed sensitive to parameter values and what constitutes sensitivity? For example, if our best estimated parameter sets give response efficiencies E_u in the range of 0.85-1.0 for alternative parameter sets, would this be considered as sensitive or insensitive? There is no hard and fast rule, though by analogy with advances in methods of genetic evaluation (which improve accuracy of evaluation and hence response), potential losses in response of one or two percent may be of little concern whereas losses of 5% or more at the very least warrant careful investigation.

6.4.2.1 A Numerical Example of Testing Index Sensitivity to Variance/Covariance Estimates

Imagine a species such as swine or beef cattle, selected for growth performance, perhaps as a terminal sire. Assume that the traits in the aggregate genotype have been determined to be slaughter weight (SW) and feed intake during growth to slaughter (FI), that both these traits are recorded, and that a phenotypic selection index is used. When both traits are scaled so that $\sigma_p = 1$, economic weights of SW and FI were found to be 2 and -1 monetary units per phenotypic s.d. (Note that FI has a negative economic weight, reflecting that feed costs money; increased FI decreases profit.) Heritabilities of SW and FI have been estimated from several data sets and found to be 0.4 for both traits. Phenotypic and genetic correlations have been estimated only once and found to be 0.8 and 0.8. Thus, since selection is on phenotypic records of the

$$\text{individual, } \mathbf{P} = \begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} .4 & .32 \\ .32 & .4 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

Note that the off-diagonal element of G , $\sigma_{g_{SW,FI}}$, is obtained by recalling that $r_{g_{ij}} = \frac{\sigma_{g_{ij}}}{\sqrt{\sigma_{g_i}^2 \sigma_{g_j}^2}}$,

$$\text{hence } \sigma_{g_{SW,FI}} = 0.8 \sqrt{0.4 * 0.4}.$$

From these parameters and selection intensity i we obtain, $\mathbf{b} = \begin{bmatrix} .8 \\ -.4 \end{bmatrix}$ and $S_H = 0.5367 i$.

In the present case, we might decide that heritabilities are well estimated, coming as they do from several published data sets, but we are uncertain about the estimated correlations. In general, phenotypic correlations are much more accurately estimated than genetic correlations. So we conclude that we are most uncertain about our genetic correlation, which we believe may in reality lie anywhere between 0.65 and 0.95.

We then investigate the following four situations for the correlations, all other parameters remaining constant:

1. $r_{g_{SW,FI}} = 0.80$ and $r_{p_{SW,FI}} = 0.8$
2. $r_{g_{SW,FI}} = 0.65$ and $r_{p_{SW,FI}} = 0.8$
3. $r_{g_{SW,FI}} = 0.95$ and $r_{p_{SW,FI}} = 0.8$
4. $r_{g_{SW,FI}} = 0.80$ and $r_{p_{SW,FI}} = 0.7$.

To calculate E_{21} we will need to determine the index weight corresponding to situation 2.

The \mathbf{G} matrix in that case is: $\mathbf{G} = \begin{bmatrix} .4 & .26 \\ .26 & .4 \end{bmatrix}$

yielding new estimates of $\mathbf{b} = \begin{bmatrix} 1.233 \\ -0.867 \end{bmatrix}$

Response to selection when using the index weights for situation 2 while parameters for situation

$$1 \text{ are the true parameters is: } S_{H_{U_2,t_1}} = i \frac{\mathbf{b}'_2 \mathbf{G}_1 \mathbf{v}}{\sqrt{\mathbf{b}'_2 \mathbf{P}_1 \mathbf{b}_2}} = 0.5122 i$$

The efficiency of using alternative parameter set 1 while the original set 1 is true is estimated as:

$$E_{21} = \frac{S_{H_{U_2,t_1}}}{S_{H_{t_1}}} = 0.5122 / 0.5367 = 0.9545$$

Similarly, efficiencies for all the other combinations of parameter sets can be calculated, which

results in:
$$\mathbf{E} = \begin{bmatrix} 1 & 0.9545 & 0.8480 & 0.9738 \\ 0.9545 & 1 & 0.6513 & 0.8746 \\ 0.8480 & 0.6513 & 0 & 0.9258 \\ 0.9751 & 0.8646 & 0.9444 & 1 \end{bmatrix}$$

Two things are worth pointing out in these results. In the first place, uncertainty over the true value of the genetic correlation between SW and FI seems to be important in terms of response to selection since an index using a genetic correlation of 0.8 is only 84.8% effective if r_g is as high as 0.95, although it is 95.5% efficient if r_g is as low as 0.65. We conclude that, in terms of finding an optimal index for this selection program, we need to have a more accurate estimate of the genetic correlation between SW and FI than we have at present. In particular, it will be important to know whether this genetic correlation is higher than our current estimate of 0.8. A reduction of the phenotypic correlation from 0.8 to 0.7 has a smaller effect on efficiency than a similar change in the genetic correlation.

Secondly, the efficiency is symmetric about the diagonal for the first three-parameter sets but not for the last set. In the first three parameter sets \mathbf{G} changed but \mathbf{P} remains constant. For those cases it has been shown that the efficiency criteria is equal to the correlation between indices. Given that similarity, this symmetry is no surprise. Parameter set 4 resulted in a different \mathbf{P} compared to the other three sets and as a consequence, for example, efficiency E_{41} is not the same as E_{14} .

Predicted optimal economic responses (S_H) for the first three parameter sets corresponding to r_g = 0.65, 0.8, and 0.95 were 0.749, 0.537, and 0.422 economic units per standard deviation of selection intensity. Predicted absolute economic responses to selection increased by 77.5% as r_g decreased from 0.95 to 0.65.

From experience, it seems to be a general phenomenon that altering variances and covariances has a much larger effect on the prediction of absolute economic response to selection than on the efficiency of one index versus another. Since the prediction of absolute economic response is used in assessment of the economic cost-benefits of selection programs, having good estimates of variances and covariances will often be of even more importance for cost-benefit assessments than for the optimization of a particular program in terms of the index to be used.

6.4.3 Uncertainty Over Economic Weights

As will be discussed later, economic weights are rarely known without error. Indeed, there is often considerable uncertainty of what the true economic weight might be, arising from uncertainties of biological and management models used and uncertainty about the values of different traits in future production systems and markets. In many cases we will therefore have a "best estimate" set of economic weights and several, perhaps many, alternatives, covering alternative present and future scenarios.

As with investigations of uncertain variances and covariances, we can carry out analogous investigations for uncertain economic weights. In this case the subscripts u and t refer to the economic weights used and those deemed to be "true". Then, efficiency of a set of index weights derived for the used set of economic weights when compared to that for the "true" set of

economic weights is given by:
$$E_{ut} = \frac{S_{H_u}}{S_{H_t}} = \frac{\mathbf{b}'_u \mathbf{G} \mathbf{v}_t}{\sqrt{\mathbf{b}'_u \mathbf{P} \mathbf{b}_u}} \frac{1}{\sqrt{\mathbf{b}'_t \mathbf{G} \mathbf{v}_t}}$$

where

$$\mathbf{b}_u = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}_u$$

and

$$\mathbf{b}_t = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}_t.$$

As with uncertainties over variances and covariances, E_{ut} could be derived for a single set of "most likely" or "best estimate" economic weights, compared to a variety of alternative economic weights. Or, a matrix of efficiencies, \mathbf{E} , could be determined, comparing every set of economic weights with every other set. Economic weights do not affect \mathbf{P} and, as shown in the previous section, \mathbf{E} is symmetric about the diagonal, so that $E_{ij} = E_{ji}$.

6.5 The Value of Including Traits in the Selection Index

Recording of animals takes time and effort and consequently costs money. Some traits may be considerably more difficult and costly to record than others. Thus a key question in the design of breeding programs is which traits and types of relatives should be recorded? It is often relatively straightforward to identify a number of traits and types of relatives potentially useful to record. A selection index could then be constructed and the question asked, how much does each observation contribute to response in the aggregate genotype? Then, the economic benefits of including that observation in terms of enhanced response can be evaluated against the cost of recording and a decision taken on whether or not to collect that information. Consider the usual selection index problem with n observations in the index and m traits in the aggregate genotype, but define the index as including all traits that might possibly be included. Then, the efficiency of a reduced index without observation i can be defined as the ratio of economic (aggregate

genotype) response for the reduced index to that with the full index, or:
$$E_i = \frac{S_{H_i^*}}{S_H}$$

where $S_{H_i^*}$ is the response in the aggregate genotype for selection on the index without observation i , $S_{H_i^*}$, and S_H is the response with the full index I ,

$$S_{H_i^*} = i \sqrt{\mathbf{b}_i^* \mathbf{G}_i^* \mathbf{v}} = i\sigma_{I_i^*}$$

where \mathbf{b}_i^* is

$$\mathbf{b}_i^* = \mathbf{P}_i^{*-1} \mathbf{G}_i^* \mathbf{v}$$

and \mathbf{P}_i^* and \mathbf{G}_i^* are reduced forms of \mathbf{P} and \mathbf{G} corresponding with the reduced index. Matrix \mathbf{P}_i^* is found by deleting the i^{th} row and column from \mathbf{P} , and \mathbf{G}_i^* by deleting the i^{th} row from \mathbf{G} .

Similarly, the efficiency of indexes with more than one observation deleted can easily be found.

For an index with observations i and j missing: $E_{ij} = \frac{S_{H_{ij}^*}}{S_H}$

where

$$S_{H_{ij}^*} = \sqrt{\mathbf{b}_{ij}^* \mathbf{G}_{ij}^* \mathbf{v}}$$

and

$$\mathbf{b}_{ij}^* = \mathbf{P}_{ij}^{*-1} \mathbf{G}_{ij}^* \mathbf{v}$$

and \mathbf{P}_{ij}^* is \mathbf{P} with the i^{th} and j^{th} rows and the i^{th} and j^{th} columns removed and \mathbf{G}_{ij}^* is \mathbf{G} with the i^{th} and j^{th} rows removed.

The efficiency, given a certain aggregate genotype and set of population parameters, is directly related to changes in r_{HI} . The contribution of each individual observation in the index to the accuracy can be calculated. Cunningham (1969) has described a method to derive the decrease in accuracy which is given below.

Assume an index with n sources of information, $I = \mathbf{b}'\mathbf{x}$ and an aggregate genotype with m traits, $H = \mathbf{v}'\mathbf{g}$. Solving the set of equations, $\mathbf{P}\mathbf{b} = \mathbf{G}\mathbf{v}$, gives the vector \mathbf{b} . The variance of the index is $\mathbf{b}'\mathbf{P}\mathbf{b}$. When, for instance, the first observation is ignored, we obtain a new index, I_1^* , with a vector with $n-1$ weighting factors, β . The weighting factors are a solution of solving: $\mathbf{P}_1^*\beta = \mathbf{G}_1^*\mathbf{v}$ where, as before, \mathbf{P}_1^* is the $n-1$ by $n-1$ variance-covariance matrix of the observations in I_1^* , and \mathbf{G}_1^* is the $n-1$ by m matrix obtained from \mathbf{G} by ignoring the first row.

Based on the original equation, this can be visualized as:

$$\mathbf{P}\mathbf{b} = \mathbf{G}\mathbf{v}$$

$$\begin{bmatrix} P_{11} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P}_1^* \end{bmatrix} \begin{bmatrix} b_1 \\ \mathbf{b}_0 \end{bmatrix} = \begin{bmatrix} G_1 \\ \mathbf{G}_1^* \end{bmatrix} \mathbf{v} = \begin{bmatrix} R_1 \\ \mathbf{R}_0 \end{bmatrix}$$

where P_{11} , b_1 , and R_1 are scalars representing the first elements in \mathbf{P} , $\mathbf{G}\mathbf{v}$, and \mathbf{b} , respectively. Matrix \mathbf{Q} is a column vector representing the remaining elements of the first column in \mathbf{P} . In most cases, $\beta \neq \mathbf{b}_0$.

The variance of the reduced index I_1^* is:

$$\sigma_{I_1^*}^2 = \beta' \mathbf{P}_1^* \beta$$

and its correlation with the aggregate genotype H is: $r_{H_i^*} = \sqrt{\frac{\beta' \mathbf{P}_1^* \beta}{\sigma_H^2}}$

Efficiency of the index ignoring observation i is equal to $E_{I_i^*, I} = \frac{S_{H_i^*}}{S_H} = \frac{r_{H, I_i^*}}{r_{H, I}} = \frac{\sigma_{I_i^*}}{\sigma_I}$, which

depends entirely on the difference in variances of the indices, $\beta' \mathbf{P}_1^* \beta$ and $\mathbf{b}' \mathbf{P} \mathbf{b}$. To determine the

magnitude of this difference, let \mathbf{W} be the inverse matrix of \mathbf{P} :

$$\mathbf{W} = \mathbf{P}^{-1} = \begin{bmatrix} W_{11} & \dots & W_{1n} \\ \dots & \dots & \dots \\ W_{n1} & \dots & W_{nn} \end{bmatrix}$$

It can be shown (Cunningham, 1969) that:

$$\sigma_I^2 - \sigma_{I_i^*}^2 = \frac{\mathbf{b}_i^2}{W_{ii}}$$

In general, ignoring the i^{th} observation in the index: $\sigma_I^2 - \sigma_{I_i^*}^2 = \frac{\mathbf{b}_i^2}{W_{ii}}$

The advantage of this method is that in computing the decrease in variance, no new index has to be derived; information for the computation is available from computations for the original

index. Variance $\sigma_{I_i^*}^2$ can be derived from the equation above as: $\sigma_{I_i^*}^2 = \sigma_I^2 - \frac{\mathbf{b}_i^2}{W_{ii}}$

Efficiency of the index ignoring observation i can also be derived from the above results as:

$$E_{I_i^*, I} = \frac{S_{H_i^*}}{S_H} = \frac{r_{H, I_i^*}}{r_{H, I}} = \sqrt{\frac{\sigma_I^2 - \frac{\mathbf{b}_i^2}{W_{ii}}}{\sigma_I^2}}$$

The method can be extended to compute the reduction of variance of the index when simultaneously ignoring more than one trait. The variance of an index, from which the i^{th} until the j^{th} observation is ignored, is $\sigma_{I_{i,j}}^2 = \sigma_I^2 - \mathbf{b}_{i,j}' \mathbf{W}_{i,j}^{-1} \mathbf{b}_{i,j}$

where $\mathbf{b}_{i,j}$ is the vector including the i^{th} until the j^{th} weighting factor of the original index, and $\mathbf{W}_{i,j}$ is the diagonal submatrix of \mathbf{P}^{-1} corresponding with the observations that were removed from the index.

6.5.1 Example of A Reduced Index

Consider the example in section 6.5.2.1 for selection on slaughter weight (SW) and feed intake (FI). How much do inclusion of SW and FI in the index contribute to selection response?

The full index was given by $\mathbf{P} = \begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix}$, $\mathbf{G} = \begin{bmatrix} .4 & .32 \\ .32 & .4 \end{bmatrix}$ with $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

Giving $\mathbf{b} = \begin{bmatrix} .8 \\ -.4 \end{bmatrix}$, $\sigma_I^2 = .288$, and $S_H = 0.537 i$

An index without SW would be found from

$$\mathbf{P}_I^* = [1], \quad \mathbf{G}_I^* = [.32 \quad .4], \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$$

giving $\mathbf{b}_I^* = 0.24$ and $S_{H_I^*} = 0.24 i$

$$\text{and hence } E_1 = \frac{S_{H_I^*}}{S_H} = 0.447$$

Similarly, removing FI from the original index, gives

$$\mathbf{P}_2^* = [1], \quad \mathbf{G}_2^* = [.4 \quad .32] \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

giving $\mathbf{b}_2^* = 0.48$, $S_{H_2^*} = 0.48 i$

and $E_2 = 0.894$.

These results could also have been obtained from \mathbf{b} and \mathbf{P}^{-1} , which is equal to:

$$\mathbf{P}^{-1} = \mathbf{W} = \begin{bmatrix} 2.778 & -2.222 \\ -2.222 & 2.778 \end{bmatrix}$$

$$\text{Thus: } E_1 = \sqrt{\frac{\sigma_I^2 - \frac{b_1^2}{W_{11}}}{\sigma_I^2}} = \sqrt{\frac{.288 - \frac{.8^2}{2.778}}{.288}} = 0.447$$

$$\text{and: } E_2 = \sqrt{\frac{\sigma_I^2 - \frac{b_2^2}{W_{22}}}{\sigma_I^2}} = \sqrt{\frac{.288 - \frac{(-.4)^2}{2.778}}{.288}} = 0.894$$

Thus, not recording SW would reduce economic response by $(1 - 0.447) 100 = 55.3\%$, while not recording FI would reduce economic response by only 10.6% compared to recording both traits. Since recording FI is likely to be more expensive than SW, it would seem well worthwhile to ask in detail how much it will cost to record FI in the breeding program and what the expected extra 10.6% economic response is worth. The final answer will depend on such parameters as the number of animals in our breeding program, the number of animals benefiting from the extra genetic response, the cost of recording individual animals, and the time between incurring costs (recording) and obtaining returns (selling or producing from genetically improved animals).

6.6 The Value of Traits in the Aggregate Genotype

With equation (6.11): $S_H = i \sqrt{\mathbf{b}' \mathbf{G} \mathbf{v}}$, response in the aggregate genotype as a result of selection on the index can be calculated. Response in the aggregate genotype is an important criterion in comparing different indexes. In addition, it is also interesting to look at responses in the individual traits in the aggregate genotype. This will give information on which traits have contributed to response in the aggregate genotype and one might want to look at the direction and size of change in each of the traits. In the example of beef cattle, it is interesting to see whether response to selection is due to a change in slaughter weight, feed intake, or a combination of these two traits. In this section, we will illustrate the relationship between changes in the aggregate genotype and in individual traits. Furthermore, we will look at comparing response to selection for various situations and try to give some guidelines on the criteria to use in such a comparison.

The expected change in individual traits (g_j) of the aggregate genotype as a result of selection on the selection index (I) can be computed using the covariance between g_j and I , as shown in (6.18). Recall that the expected change in g_j (expressed in units of measurement) as a

consequence of selection on I is:

$$S_{g_j} = i \frac{\mathbf{b}' \mathbf{G}_j}{\sigma_I}$$

where \mathbf{G}_j denotes the j^{th} column of \mathbf{G} , and that the vector of responses in each trait is equal to:

$$\mathbf{S}_g = i \frac{\mathbf{b}' \mathbf{G}}{\sigma_I}$$

Response in the aggregate genotype, S_H in monetary units (assuming that elements of \mathbf{v} are expressed in monetary units), can be calculated as:

$$S_H = i \frac{\mathbf{b}' \mathbf{G} \mathbf{v}}{\sigma_I}$$

which is equal to:

$$S_H = i \frac{\mathbf{b}' \mathbf{G}}{\sigma_I} \mathbf{v} = i \mathbf{S}_g \mathbf{v} = i \sum \frac{\mathbf{b}' \mathbf{G}_j}{\sigma_I} v_j$$

The above equation demonstrates that response in the aggregate genotype is the sum of responses in individual traits multiplied by their economic weights. The contribution of response in g_j to

overall response (C_j) can be calculated as: $C_j = \frac{S_{g_j} v_j}{S_H}$

With this criterion we can evaluate the relative contribution of each trait in the aggregate genotype. To illustrate this, we return to the example of 6.5.2.1, use parameter set 1 and set selection intensity equal to 1. We have seen that response in the aggregate genotype, S_H , equals 0.537. Using (6.18), response in slaughter weight is 0.358 and response in feed intake is 0.179, both expressed in phenotypic standard deviations. The economic weight is 2 and -1 for SW and FI, respectively. Using these weights, response in monetary units in SW and FI is equal to \$0.716 and -\$0.179, respectively. The contribution of SW to overall response (C_{SW}) is equal to $0.716/0.537 = 1.33$. In words, response in SW accounts for 133% of the monetary response in the aggregate genotype. At first sight, it looks strange that the contribution exceeds 100%. When we have a closer look at the example, we see that we are trying to improve two traits

which are positively correlated, of which one has a positive and the other has a negative economic weight. Given the parameters we have used, the index results in an increase in slaughter weight, which is associated with a change in feed intake. The latter change results in a reduction in the financial gain. This is reflected by the negative contribution of 33% of FI to the overall response.

From the example it is clear that C_j can take values that are smaller than zero or bigger than one. The advantage of looking at values for C_j instead of S_{g_j} is particularly the unit that is used, i.e. money vs units of measurement.

Different indexes can be compared using accuracy of selection, as long as the parameters for the aggregate genotype (e.g. genetic variances and co-variances) and the economic weights have not changed. Comparison of indexes in other cases is less straightforward. Let us, for example, return to the example and consider a second index for a situation where feed intake was excluded from the aggregate genotype and for two different values of r_g , all other factors including the index remaining unaffected. Results for that and the original aggregate genotype are summarized in Table 6.2.

Table 6.2 Results of selection using a selection index including slaughter weight (SW) and feed intake (FI) for two different values of the genetic correlation between SW and FI r_g and economic weight of FI v_2 using a selection intensity of 1^{a)}.

	$r_g = 0.8$		$r_g = 0.65$	
	$v_2 = -1$	$v_2 = 0$	$v_2 = -1$	$v_2 = 0$
r_{HI}	0.632	0.632	0.765	0.652
σ_I	0.537	0.800	0.750	0.825
σ_H	0.849	1.265	0.980	1.265
S_H	0.537	0.800	0.750	0.835
S_{SW}	0.358	0.4	0.358	0.412
S_{FI}	0.179	0.32	-0.035	0.175
C_{SW}	1.33	1	0.953	1
C_{FI}	-0.33	0	0.047	0
Set	1	2	3	4

^a The index in all cases consists of observations on FI and SW on the individual animal.

Let us first look at the situation where r_g is 0.8. The accuracy of r_{HI} was identical for both sets of economic weights but the variance of the aggregate genotype differed largely. We are looking at a situation with two highly correlated traits with opposite signs for their economic weights. Setting the economic weight equal to zero for one of the traits resulted in a higher variance of the aggregate genotype. The response in the aggregate genotype and slaughter weight increased when v_2 was set equal to zero. Correlated response in FI for $v_2 = 0$ amounted to 0.32 phenotypic standard deviations, which was much larger than for $v_2 = -1$. Changing the economic weight of FI did not affect r_{HI} . This is not a result that one would find in general, but one which is specific for the parameter set used in this example. There are two factors that contributed to this. In the first place, both traits had the same heritability and standard deviation. Schaeffer (1984) showed that the benefit of including a correlated trait in prediction of the EBV for a trait depends on the absolute difference between the phenotypic and genetic correlation. The correlated trait does not

contribute when the phenotypic correlation is equal to the genetic correlation, which is the case for $r_g = 0.8$. This means that $v_2 = 0$ refers to an index with observations on SW only for the correlation of 0.8. In that case one also expects to find $r_{HI} = \sqrt{0.4} = 0.632$.

Different results are found when r_g is lowered to 0.65. For $v_2 = -1$, a positive response of 0.358 units in SW is associated with a reduction in FI of 0.035 units, i.e. a change in the desired direction for both traits in the aggregate genotype. As to be expected, correlated response in FI is positive when $v_2 = 0$ but it is smaller in size than when the genetic correlation was 0.8.

In comparing the results in Table 6.2, it is obvious that there is not a single criterion which can be used. The criteria to look at very much depend on the question one wants to answer. It is important to realize that variation in the aggregate genotype differs between all four cases studied. This is different from a situation where one varies the observations to be included in the index. In that case one generally works with a constant σ_H^2 and can use the accuracy r_{HI} . In comparing the four situations in Table 6.2, the efficiency criteria introduced in section 6.5.2 might be very useful. Efficiencies for the different combinations of parameters sets, using the numbering of sets as given in Table 6.2, can be summarized as:

$$E = \begin{bmatrix} 1 & 0.954 & 0.894 & 0.976 \\ 0.954 & 1 & 0.720 & 0.867 \\ 0.894 & 0.720 & 1 & 0.970 \\ 0.976 & 0.867 & 0.970 & 1 \end{bmatrix}$$

This illustrates that when r_g is 0.8, loss in selection response amounts to $(1 - 0.954)*100\% = 4.6\%$ when FI was ignored in the aggregate genotype, while the true economic value of FI is -1.

6.7 Non-Linear Indexes

The discussion of selection indexes so far has assumed that the aggregate genotype is a linear function of additive genetic values. In practice, the aggregate genotype will often not be linear. In general, however, linear indexes can be found which closely approximate most non-linear descriptions of economic value. These will be discussed in relation to derivation of economic weights in Chapter 7.

6.8 Constrained and Desired Gains Indexes

In the selection indexes discussed so far, response to selection of traits in the aggregate genotype is determined entirely by the economic weight of that trait, the information available in the index, and the phenotypic and genetic variances and covariances among traits. It is, however, possible to construct indexes in which the rate of genetic change in one or more traits is pre-determined. For example, it is possible to maximize an index for genetic change in one set of traits subject to other traits being constrained to zero genetic change. Alternatively, the change

in one trait might be desired to be twice the change in another, with all other traits being allowed to change as the index dictates.

Methods for achieving these types of constrained or desired gain indexes were reviewed by Brascamp (1984). Linear programming (see later Chapter) can also be used to achieve constrained and desired gain indexes (Keller and Gibson, 1990; Toro, 1992).

Following Brascamp (1984), separate the vector of traits in the aggregate genotype $H = \mathbf{v}'\mathbf{g}$ into

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \end{bmatrix}$$

where \mathbf{g}_0 includes traits for which progress will be maximized according to $\mathbf{v}_0'\mathbf{g}_0$, i.e. the economic part of the breeding goal

and \mathbf{g}_1 includes traits for which changes are constrained to relative changes = δ .

Correspondingly, partition matrix \mathbf{G} as: $\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \end{bmatrix}$

Let the constrained index be: $I^* = \mathbf{b}^*\mathbf{X}$ with variance $\sigma_{I^*}^2 = \mathbf{b}^*\mathbf{P}\mathbf{b}^*$

For simplicity, and without loss of generality, set $\sigma_{I^*}^2 = 1$

Then, from Table 6.1, response in \mathbf{g}_1 to selection on this index is proportional to $\mathbf{b}^*\mathbf{G}_1$

Thus, the problem of finding the constrained index can be formulated as the following constrained optimization problem:

$$\underset{\mathbf{b}^*}{\text{Max}} \mathbf{b}^*\mathbf{G}_0\mathbf{v}_0 \quad \text{subject to } \mathbf{b}^*\mathbf{G}_1 = \alpha\delta \quad \text{and} \quad \mathbf{b}^*\mathbf{P}\mathbf{b}^* = 1$$

This problem can be solved using the Lagrange multiplier method, as described in Brascamp (1984) and Weller (1994).

Although there is considerable literature on the construction and properties of such indexes, at no point does there appear in the literature a sound reasoning for applying these indexes in economic animal breeding. Gibson and Kennedy (1991) attempted to provide a rationale for why such indexes should not be used in animal breeding, on the grounds that an economic index can always be found to at least equal and usually out-perform a constrained index. They showed that constrained or desired gains indexes have implied, or **pseudo economic values** for all traits, including for those that are constrained. The vector of pseudo economic values, \mathbf{v}^* , is the set of economic values that would have resulted in the index weights \mathbf{b}^* , using the selection index equations. Values for the pseudo economic values can be derived as follows:

Using the selection index equations, set $\mathbf{P}\mathbf{b}^* = \mathbf{G}\mathbf{v}^*$ and solve for \mathbf{v}^* as follows:

$$\mathbf{G}'\mathbf{P}\mathbf{b}^* = \mathbf{G}'\mathbf{G}\mathbf{v}^* \rightarrow \mathbf{v}^* = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{P}\mathbf{b}^*$$

6.8.1 Example of Desired Gains Index

Figures 6.2 and 6.3 illustrate the potential impact of putting a constraint on genetic progress for a given trait. The example is of selection for an aggregate genotype in pigs consisting of growth, feed intake, and backfat thickness. Economic values are 0.178 \$/g/d for growth, -0.05 \$/g/d for feed intake, and -0.0415 \$/mm for backfat thickness:

$$H = 0.178 g_{growth} - 0.05 g_{feed\ intake} - 0.0415 g_{backfat}$$

The index includes phenotype on the individual and its sire for all three traits. Based on these economic values, unit intensity of selection on the optimal index results in an increase backfat thickness by 0.7316 mm (Figure 6.2). In many breeding programs, such an increase is undesirable and, thus it seems reasonable to develop a constrained index in which genetic gain for backfat thickness is constrained to zero. Results for such an index are in Figure 6.3. For this index, the relative pseudo economic values are:

$$\mathbf{v}^* = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{P}\mathbf{b}^* = \begin{bmatrix} 0.02659 \\ -0.007469 \\ -0.585 \end{bmatrix} \text{ compared to } \mathbf{v} = \begin{bmatrix} 0.178 \\ -0.05 \\ -0.0415 \end{bmatrix} \text{ for the optimal index.}$$

Note that economic values in \mathbf{v}^* are relative economic values. When expressed relative to an economic value of 0.178 for growth rate, which is its true economic value, implied economic

values under the constrained index become: $\mathbf{v}^* = \begin{bmatrix} 0.178 \\ -0.05 \\ -3.916 \end{bmatrix}$

Thus, the implied economic value for fat thickness relative to growth is almost 100 times as large as its true economic value.

Efficiency of the constrained index in terms of response in the true aggregate genotype can be evaluated using the procedures described in sections 6.5.2 and 6.5.3. Alternatively, with genetic superiorities in the individual traits already computed for the desired gains index, response in H can be computed following section 6.7 as:

$$S_H^* = S_g^* \mathbf{v} = [13.65, -67.37, 0] \begin{bmatrix} 0.178 \\ -0.05 \\ -0.0415 \end{bmatrix} = \$5.80$$

In contrast, response for the optimal index is (see Figure 6.2) $S_H = \$7.39$

Thus, efficiency is: $E = \frac{S_H^*}{S_H} = \frac{5.80}{7.39} = 0.785$, a 21.5% loss.

Figure 6.2 Selind program output for optimal economic index for pig selection example

Listing of used matrices

Listing of P-matrix

7744	968	1.144e+04	220	47.52	27.74
968	7744	220	1.144e+04	27.74	47.52
1.144e+04	220	4e+04	5000	56	29.66
220	1.144e+04	5000	4e+04	29.66	56
47.52	27.74	56	29.66	4	1.1
27.74	47.52	29.66	56	1.1	4

Listing of G-matrix

1936	440	55.47
968	220	27.74
440	1e+04	59.33
220	5000	29.66
55.47	59.33	2.2
27.74	29.66	1.1

Listing of C-matrix

1936	440	55.47
440	1e+04	59.33
55.47	59.33	2.2

Listing of P-inverse-matrix

0.0002572	-6.187e-05	-7.397e-05	2.736e-05	-0.001678	-0.0004214
-6.187e-05	0.0002572	2.736e-05	-7.397e-05	-0.0004214	-0.001678
-7.397e-05	2.736e-05	4.751e-05	-1.344e-05	0.0001265	-1.121e-05
2.736e-05	-7.397e-05	-1.344e-05	4.751e-05	-1.121e-05	0.0001265
-0.001678	-0.0004214	0.0001265	-1.121e-05	0.2886	-0.06352
-0.0004214	-0.001678	-1.121e-05	0.0001265	-0.06352	0.2886

Index based on economic weights

	trait	ec weight	genetic sup	
	growth	0.178	30.88	
	feed intake	-0.05	-38.39	
	fat thick	-0.0415	0.7316	

	animal ,	trait	B-value	rvi (%)
	own performance ,	growth	0.08516	30.5
	sire ,	growth	0.01686	1.018
	own performance ,	feed intake	-0.03579	28.88
	sire ,	feed intake	-0.007106	0.9787
	own performance ,	fat thick	1.095	3.882
	sire ,	fat thick	0.1251	0.04967

Standard deviation of index : 7.3861

of aggregate genotype : 8.8283

Correlation between index and aggregate genotype : 0.83663

Figure 6.3 Selind program output for index for pig improvement example with gain for backfat thickness restricted to zero

RESULTS SECOND RUN WITH D AS DESIRED GAINS INPUT

Listing of Gtranspose*P-inverse*G

1252	-136.6	38.44
-136.6	5162	27.62
38.44	27.62	1.395

Listing of S

0.05653	0.011	-1.776
0.011	0.002356	-0.3498
-1.776	-0.3498	56.6

Pseudo relative economic values based on:
 $v^* = (G'G)^{-1}G'Pb^*$

trait	ec weight	genetic sup	TRUE EC. VALUES
growth	0.02659	13.65	0.178
feed intake	-0.007469	-67.37	-0.05
fat thick	-0.585	2.313e-15	-0.0415

ECONOMIC RESPONSE = $(0.178) * (13.65) + (-0.05) * (-67.37) + (-0.0415) * (0) = 5.80$

Compared to 7.3861 for the optimal economic index (21.5% reduction!!!)

animal ,	trait	b-value	rvi (%)
own performance ,	growth	0.009932	30.1
sire ,	growth	0.002312	1.395
own performance ,	feed intake	-0.004965	44.49
sire ,	feed intake	-0.001173	1.95
own performance ,	fat thick	-0.1071	2.685
sire ,	fat thick	-0.04554	0.4801

Standard deviation of index : 0.86603
 of aggregate genotype : 1.139
 Correlation between index and aggregate genotype : 0.76033

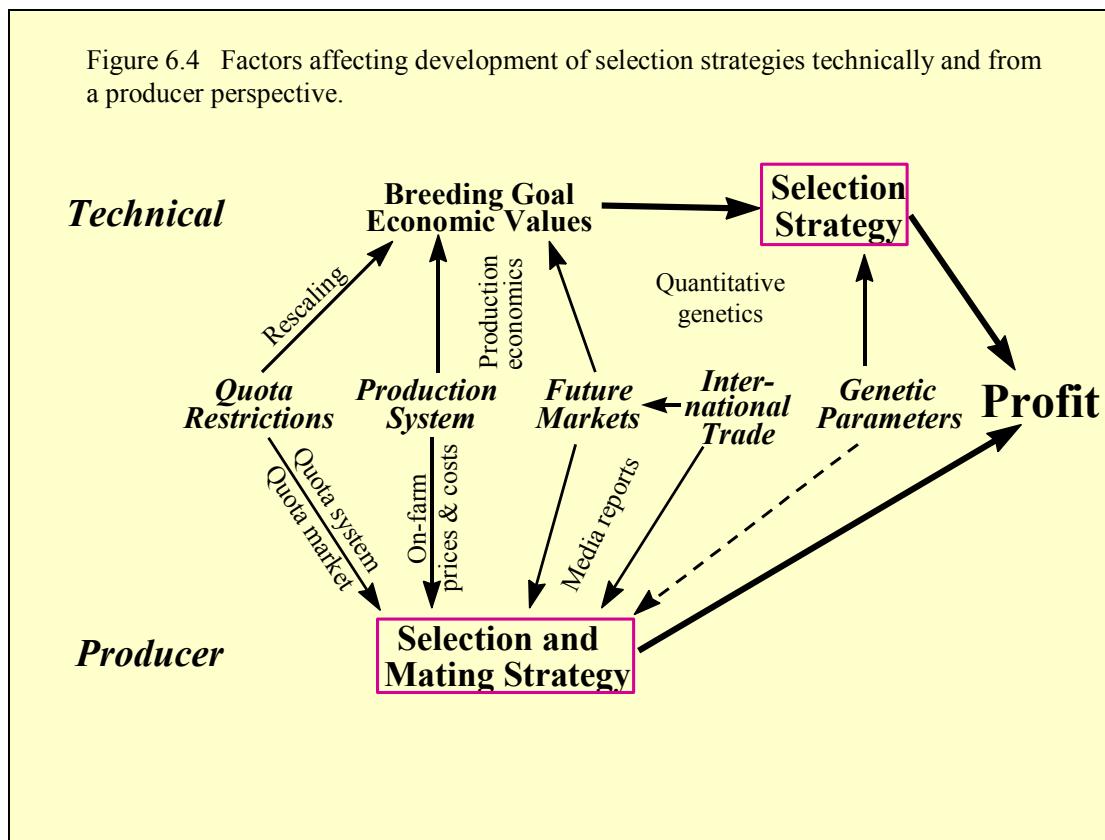
For H with implied economic values

6.9 Some Practical Considerations (Taken from Dekkers and Gibson, 1997, J. Dairy Sci.)

The objective of development of total merit selection indexes is to provide a guide to selection that maximizes selection for the overall breeding objective and to promote selection strategies that minimize misuse of information. Although development of total merit selection criteria can be based entirely on the scientific principles that were summarised above and that will be described further in Chapter 7 with regard to estimation of economic values, implementation of selection indexes in the industry requires careful consideration of acceptance of the index by its target group (i.e., the breeders or producers). An index that is based on the application of sound scientific

principles but not accepted, has much less impact on selection for an overall breeding goal than an index that may not be optimal technically but that receives acceptance by the targeted users.

The main issues involved in the development of selection strategies from a technical versus a practical perspective are summarized and contrasted in Figure 6.4. To overcome these differences in perspective, development of selection indexes for use in the field must take place in close collaboration with the industry. Although this interaction may compromise scientific principles and objectivity, there are several aspects that can be pursued from a technical basis that can aid in development of an index that is technically sound while simultaneously maximizing acceptance of the index by its intended users. Several of these approaches, as well as their theoretical background, are discussed here (taken from Dekkers and Gibson, 1997), while others have been discussed earlier in this Chapter and in Chapter 7.



6.9.1 Development of Selection Criteria

A total merit selection index combines EBV (or ETA or EPD's) for individual traits into a single number that can be used for selection. Although this index facilitates simultaneous selection for multiple traits, most breeders require insight into the index and, in particular, into the index weights. When derived based on selection index principles, weights in a selection index are equivalent to partial regression coefficients in a multiple regression equation. Similar to the difficulty of interpreting partial regression coefficients in a multiple regression analysis when explanatory variables are confounded, selection index weights can be difficult to interpret when

traits are genetically correlated. This complicates extension of the index to producers. There are several avenues that can be pursued in a) derivation of the index, b) expression of the index, or c) extension, to facilitate the understanding and, thereby, implementation of the index by its target audience.

Index derivation. Four approaches, among other, that can be used to develop total merit indexes that facilitates their understanding, acceptance, and implementation of the index are 1) alignment of index weights with economic values, 2) evaluation of the accuracy of alternative index weights, 3) use of customized indexes, and 4) development of indexes based on subindexes. In principle, all four approaches focus on consideration of alternative indexes that are more acceptable or easier to interpret and on exploring the accuracy of these indexes relative to the optimum index. Flexibility in consideration of alternative indexes without compromising accuracy stems from the robustness of genetic change to economic values, which was first explored by Vandepitte and Hazel (1977) and Smith (1983).

Aligning index weights with economic values.

Economic values of traits in the breeding goal are frequently easier to interpret than are selection index weights. Therefore, aligning index weights closely to economic weights in the breeding goal will facilitate understanding of the index. The ability to accomplish this depends on the degree to which traits in the index differ from traits in the breeding goal, whether single- or multiple-trait procedures are used for genetic evaluation of traits in the index, and, in the case of single-trait evaluation procedures, on the accuracy of EBV.

For index traits that are indicators of traits in the breeding goal, index weights can be related to their indirect economic importance based on the relationship between the indicator trait and the economic trait in the breeding goal. For example, for SCC as an indicator of susceptibility to mastitis, an indirect economic value can be derived as the genetic regression coefficient of SCC on mastitis, multiplied by the economic value of mastitis. This derivation was implicit to the economic value of SCC derived by Schutz (1994).

When EBV for traits in the index are based on a joint procedure for multiple-trait evaluation and traits in the index are the same as the traits in the breeding goal, economic values can be used directly as selection index weights, as described previously. When traits in the index are different from traits in the breeding goal, procedures described in previously can be used, and resulting index weights amount to the indirect economic values described.

When EBV are from single-trait genetic evaluation models, index weights can still be approximated based on economic weights if the accuracy of individual EBV is high. The efficiency of indexes derived on this basis relative to the optimum index must be considered in these cases.

Exploring the accuracy of alternative index weights.

Optimum index weights for a given set of economic values can be difficult to accept by producers if those weights do not correspond to the perceived incentives for genetic change, as discussed previously. For example, Gibson et al. (1992) found optimum index weights for Ontario to result in a negative weighting for milk yield, although producers were paid in part for milk volume, in addition to payments for fat and protein yield. The negative weighting on milk yield was a result of

the nature of the dual quota system that was in operation in Ontario at the time. In this system, one quota was based on volume for fluid milk sales, with a substantial premium for milk volume, and the second quota was based on fat sales, with no premium for milk volume. Because of the partial payment for volume, an index with a negative weighting on milk was difficult to accept by producers. An index with a zero weighting on milk volume was explored as an alternative. This index was found to be over 98.5% as efficient as the optimal index and was subsequently implemented by the industry as part of the LPI. Another aspect that helped implementation of this index was the demonstration that selecting on this index was expected to result in substantial improvement in milk yield, despite a zero emphasis on milk in the index.

Customised Indexes.

Customized indexes allow producers to develop a selection index based on economic circumstances that are specific to their herd Bowman et al. (1996). The use of customized indexes is justified technically if economic circumstances differ between herds or if traits are genetically different across herds Visscher and Amer (1996), which would make use of a single index derived based on economic circumstances of an average farm inappropriate. Visscher and Amer (1996) found that, for economic and genetic parameters typical for dairy cattle, customized indexes did not result in substantially greater improvements in profit at the population level than a single index that was based on average parameters. This result was caused in particular by the dominating economic importance of production traits and the high genetic correlations among milk, fat, and protein production. Visscher and Amer (1996) concluded that the main reason for use of customized indexes in dairy cattle is promoting the use of a selection index approach to multiple-trait selection instead of selection on independent culling levels. In development of customized indexes care must be taken, however, that objective economic information is used in their derivation; Visscher and Amer (1996) assumed that customized indexes were based on accurate economic data for individual farms.

Although customized indexes can play an important role in implementation of selection index principles in the field, we do advocate availability and promotion of an overall index that applies to average circumstances in the population. Such an index can provide default values for weights in customized indexes and provide a unified and global focus for selection and marketing decisions for all sectors of the breeding industry.

Development of indexes based on subindexes.

This was described previously.

6.9.1 Expression of the index

Interpretation and implementation of an overall selection index can be facilitated by its expression. Expression of this index includes the name given to the index which must convey the purpose and meaning of the index, expression of the index formula, and the scale on which the index is expressed. The latter two issues are discussed further.

Expression of the index formula.

The magnitude of economic values and index weights is dependent on the scale of expression of EBV and may not reflect the relative emphasis on traits in relation to genetic selection decisions. Multiplying index weights by the genetic standard deviation for the trait provides standardized weights that reflect the emphasis put on each trait in relation to the genetic variability that is present in the population.

Index weights can also be standardized by the standard deviation of EBV. Although the type of standardization (i.e., based on the standard deviation of true versus estimated breeding values) does not affect the eventual index values, there are some significant differences with regard to interpretation of the resulting index weights. These are summarised in the following: 1) index weights standardized by the genetic standard deviation are independent of the accuracy of genetic evaluations and are, therefore, more closely related to the economic importance of traits in the breeding goal; 2) standardization on the basis of the standard deviation of EBV reduces the emphasis that is perceived to be put on traits with low heritability (i.e., low accuracy of EBV). Standardization on the basis of the genetic standard deviation maintains the distinction between the ability to identify genetic differences in the population and the relative emphasis that is put on a given difference in EBV between animals; 3) the relative magnitude of the standard deviation of EBV among traits may differ between, for example, sires and cows. This would result in different indexes if standardization is on the basis of the standard deviation of EBV.

Scale of expression.

A total merit index ranks sires and cows based on genetic merit for the overall breeding goal. Apart from use of the index as a ranking tool, understanding, implementation and use of the index can be enhanced if the index is expressed in meaningful units. For total merit indexes that are intended to rank animals based on genetic merit for profitability, expressing the index in monetary units facilitates and promotes use of the index. The ultimate goal is to express the index in a way that enables its use for investment decisions (e.g., semen purchase). Expression of total merit indexes for sires as a net present value of a dose of semen was explored by McGilliard (1978), Bakker et al. (1980), and others. This requires consideration of the time and frequency of expression of genetic superiority in resulting daughters, discount rate, and conception and survival rates. Based on such indexes, differences between bulls in expected returns from a dose of semen can be compared directly to differences in semen price.

6.9.2 Extension of the index

Effective implementation of any index relies on promotion of the index through extension activities. Extension efforts should focus on the consequences of selection on an index rather than on the composition of the index. Index weights or economic values may not give a clear indication of what can be expected from use of the index. This is because responses to selection on an index are affected by genetic constraints on improvement of individual traits, which are quantified by the genetic parameters, as well as by the emphasis put on each trait in the index through the index weights.

Responses to selection can also be used to illustrate that, although for dairy cattle an increased protein to fat ratio may seem a reasonable selection objective, and selection indexes can be derived that maximise that objective, selection on such an index may have detrimental effects on responses for yield traits and result in reduced fat and protein yields.

Another example of the use of predicted responses to index selection in extension is to alleviate concerns regarding indexes with a negative weight on SCC as indicators of susceptibility to mastitis. Such indexes are often perceived to reduce SCC, which raises concerns about reaching levels of SCC that are too low to manifest an effective resistance to infection. Consideration of responses to selection on an index that includes production and SCC illustrates that such an index does not result in dramatic reductions in SCC but may instead reduce the rise in SCC that is the result of the positive genetic correlation between SCC and production.

Multiple Stage Selection

Jack Dekkers

Multi-trait breeding goal: $H = v_1g_1 + v_2g_2 + v_3g_3 + \dots + v_ng_n = \mathbf{v}'\mathbf{g}$

Information sources: $X_1, X_2, X_3, X_4, \dots, X_m$

Multi-trait selection index: $I = b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_mX_m$

Optimal index weights: $\mathbf{b} = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}$

Selection on I maximizes response to selection in H , but requires
all animals to be measured for all traits.

Multiple-stage selection:

Stage 1: select on $I_1 = b_1X_1 + b_2X_2 + \dots + b_kX_k = \mathbf{b}_1' \mathbf{X}_1$

Stage 2: select on $I_2 = b_1X_1 + b_2X_2 + \dots + b_kX_k + b_{k+1}X_{k+1} + \dots + b_mX_m = \mathbf{b}_2' \mathbf{X}$

Only animals that are selected in stage 1 have to be evaluated for $\mathbf{X}_{k+1}, \dots, \mathbf{X}_m$

→ Cost savings

→ Opportunities to increase population size for early stages

Optimal index weights:

Stage 1: $I_1 : \mathbf{b}_1 = \mathbf{P}_{11}^{-1} \mathbf{G}_1 \mathbf{v} \quad \mathbf{P}_{11} = \text{Var}(\mathbf{X}_1) \quad \mathbf{G}_1 = \text{Cov}(\mathbf{X}_1, \mathbf{g})$

Stage 2: $I_2 : \mathbf{b}_2 = \mathbf{P}^{-1} \mathbf{G} \mathbf{v} \quad \mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$

Optimal weights for index I_2 are not affected by selection on I_1 in stage 1, provided all data included in I_1 is also included in I_2 (Cunningham 1975 Theor. Appl. Genet. 46:55)

But accuracy and response to selection on I_2 are affected by selection on I_1 :

Stage 1: accuracy of I_1 : $r_1 = \sqrt{\frac{\mathbf{b}_1' \mathbf{G}_1 \mathbf{v}}{\mathbf{v}' \mathbf{C} \mathbf{v}}}$ Trait response vector: $\mathbf{S}_{g,1} = i_1 \frac{\mathbf{b}_1' \mathbf{G}_1}{\sqrt{\mathbf{b}_1' \mathbf{P}_{11} \mathbf{b}_1}}$

Stage 2: accuracy of I_2 : $r_2 = \sqrt{\frac{\mathbf{b}_2' \mathbf{G}^* \mathbf{v}}{\mathbf{v}' \mathbf{C}^* \mathbf{v}}}$ Trait response vector: $\mathbf{S}_{g,2} = i_2 \frac{\mathbf{b}_2' \mathbf{G}^*}{\sqrt{\mathbf{b}_2' \mathbf{P}^* \mathbf{b}_2}}$

This assumes multi-variate normality of variables at stage 2 (despite stage 1 selection).

Total response vector across both stages: $\mathbf{S}_g = \mathbf{S}_{g,1} + \mathbf{S}_{g,2}$

Matrices \mathbf{P}^* , \mathbf{G}^* , and \mathbf{C}^* are \mathbf{P} , \mathbf{G} , and \mathbf{C} matrices adjusted for selection on I_1

Matrix equivalent of adjustment of (co-)variance for selection on variable w
(used for Bulmer effect):

$$\sigma_{xy}^* = \sigma_{xy} - k \frac{\sigma_{wx}\sigma_{wy}}{\sigma_w^2} \quad k = i(i-t) \quad t = \text{truncation point}$$

Consider vectors $\mathbf{w}, \mathbf{x}, \mathbf{y}$ Select on index $\mathbf{b}'\mathbf{w}$

$$\begin{aligned} \text{Cov}(\mathbf{x}, \mathbf{y})^* &= \text{Cov}(\mathbf{x}, \mathbf{y}) - k \frac{\text{Cov}(\mathbf{x}, \mathbf{b}'\mathbf{w}) \text{Cov}(\mathbf{b}'\mathbf{w}, \mathbf{y})}{\text{Var}(\mathbf{b}'\mathbf{w})} \\ &= \text{Cov}(\mathbf{x}, \mathbf{y}) - k \frac{\text{Cov}(\mathbf{x}, \mathbf{w}) \mathbf{b} \mathbf{b}' \text{Cov}(\mathbf{w}, \mathbf{y})}{\mathbf{b}' \text{Var}(\mathbf{w}) \mathbf{b}} \end{aligned}$$

With stage 1 selection on $\mathbf{b}'\mathbf{w} = \mathbf{b}_1' \mathbf{X}_1 \rightarrow$ Matrices to use in Stage 2:

$$\begin{aligned} \mathbf{P}^* &= \text{Var}(\mathbf{X})^* = \text{Cov}(\mathbf{X}, \mathbf{X})^* = \mathbf{P} - k \frac{\text{Cov}(\underline{\mathbf{X}}, \underline{\mathbf{X}}_1) \underline{\mathbf{b}}_1 \underline{\mathbf{b}}_1' \text{Cov}(\underline{\mathbf{X}}_1, \underline{\mathbf{X}})}{\underline{\mathbf{b}}_1' \text{Var}(\underline{\mathbf{X}}_1) \underline{\mathbf{b}}_1} \\ &= \mathbf{P} - k \frac{\begin{bmatrix} \underline{\mathbf{P}}_{11} \\ \underline{\mathbf{P}}_{21} \end{bmatrix} \underline{\mathbf{b}}_1 \underline{\mathbf{b}}_1' \begin{bmatrix} \underline{\mathbf{P}}_{11} & \underline{\mathbf{P}}_{21} \end{bmatrix}}{\underline{\mathbf{b}}_1' \underline{\mathbf{P}}_{11} \underline{\mathbf{b}}_1} \end{aligned}$$

$$\begin{aligned} \mathbf{G}^* &= \text{Cov}(\mathbf{X}, \mathbf{g})^* = \mathbf{G} - k \frac{\text{Cov}(\underline{\mathbf{X}}, \underline{\mathbf{X}}_1) \underline{\mathbf{b}}_1 \underline{\mathbf{b}}_1' \text{Cov}(\underline{\mathbf{X}}_1, \underline{\mathbf{g}})}{\underline{\mathbf{b}}_1' \text{Var}(\underline{\mathbf{X}}_1) \underline{\mathbf{b}}_1} \\ &= \mathbf{G} - k \frac{\begin{bmatrix} \underline{\mathbf{P}}_{11} \\ \underline{\mathbf{P}}_{21} \end{bmatrix} \underline{\mathbf{b}}_1 \underline{\mathbf{b}}_1' \underline{\mathbf{G}}_1}{\underline{\mathbf{b}}_1' \underline{\mathbf{P}}_{11} \underline{\mathbf{b}}_1} \end{aligned}$$

$$\begin{aligned} \mathbf{C}^* &= \text{Var}(\mathbf{g})^* = \text{Cov}(\mathbf{g}, \mathbf{g})^* = \mathbf{C} - k \frac{\text{Cov}(\underline{\mathbf{g}}, \underline{\mathbf{X}}_1) \underline{\mathbf{b}}_1 \underline{\mathbf{b}}_1' \text{Cov}(\underline{\mathbf{X}}_1, \underline{\mathbf{g}})}{\underline{\mathbf{b}}_1' \text{Var}(\underline{\mathbf{X}}_1) \underline{\mathbf{b}}_1} \\ &= \mathbf{C} - k \frac{\underline{\mathbf{G}}_1' \underline{\mathbf{b}}_1 \underline{\mathbf{b}}_1' \underline{\mathbf{G}}_1}{\underline{\mathbf{b}}_1' \underline{\mathbf{P}}_{11} \underline{\mathbf{b}}_1} \end{aligned}$$

See 2-stage selection example.xls

Multi-stage selection with availability of multi-trait EBV:

EBV for all m traits available at every stage (but with different accuracies)

- select on complete index at every stage with weights = economic values

$$I = \mathbf{v}_1 \hat{\mathbf{g}}_1 + \mathbf{v}_2 \hat{\mathbf{g}}_2 + \dots + \mathbf{v}_n \hat{\mathbf{g}}_n$$

Optimization of proportions selected at each stage

$$\left. \begin{array}{l} \text{Total proportion selected over } s \text{ stages} = P \\ \text{Proportion selected at stage } i = p_i \end{array} \right\} P = \prod_{i=1}^s p_i$$

\mathbf{a}_i = cost of traits measured at stage i

$$\text{Total cost} = \text{TC} = \mathbf{a}_1 + \sum_{i=2}^s \mathbf{a}_i \prod_{j=1}^{i-1} p_j$$

Proportions selected and measured at each stage can then be optimized based on an overall objective function (e.g. profit) and associated responses to selection.

Selection response in stage 2 predicted by assumes multivariate normality

$$S_{g,2} = i_2 \frac{\mathbf{b}_2' \mathbf{G}^*}{\sqrt{\mathbf{b}_2' \mathbf{P}^* \mathbf{b}_2}}$$

But, selection in stage 1 not only reduces variances but also introduces skewness, depending on the correlation between $I_{1:}$ and $I_{2:}$

Accounting for skewness requires integration of multivariate normal distributions, or Monte Carlo simulation.

See Ducrocq and Colleau (1986 GSE 18:447)
 Jopson et al. (2004, Proc. New Zealand Soc. Anim. Prod. 64: 217)

This has been implemented in SelAction

References:

- Cunningham (1975), Theor. Appl. Genet. 46:55
- Ducrocq and Colleau (1989) Genet. Sel. Evol. 21:185
- Xu and Muir (1991) Genetics 129:963
- Xu and Muir (1992) Theor. Appl. Genet. 83:451
- Xu, Martin, and Muir (1995) J. Anim. Sci. 73:699