FOREST GENETICS 7(1):39–46, 2000

IMPACT OF ERRORS OF ESTIMATING MULTIVARIATE GENETIC PARAMETERS ON SELECTION EFFICIENCY AND EFFICACY OF ALTERNATIVE METHODS OF OVERCOMING SUCH ERRORS

S. Kumar[1,2], R. D. Burdon[1] & D. J. Garrick[2]

1) New Zealand Forest Research Institute, Private Bag 3020, Rotorua, New Zealand
2) Institute of Veterinary, Animal and Biomedical Science, Massey University, Palmerston North, New Zealand
Correspondence: S. Kumar, Institute of Veterinary, Animal and Biomedical Sciences, Massey University,
Private Bag 11222, Palmerston North, New Zealand. E-mail: S.Kumar@massey.ac.nz

Received June 4, 1998; accepted September 23, 1999

ABSTRACT

Selection index gives optimal index weights and maximises expected genetic gain when variance-covariance parameters are known exactly. However, in practice this is seldom the case; the parameters must be estimated. The present study was conducted to overcome the effects of sampling errors on selection efficiency. A method is proposed which consists of 'regressing' the estimated (least-squares) selection index coefficients ($\hat{b}$) towards the relative economic values (a), which are assumed to be known precisely, as: $\hat{b}^* = \hat{b}\,k + a\,(1\text{-}k)$, for $0 \le k \le 1$. The efficiency of proposed method along with some other index selection procedures (like unmodified index selection, base index and "bending" method) was evaluated for 192 parameter and sample situations with 1000 replicates each, by Monte Carlo simulation. The highest gain was associated with $k < 1$. The mean improvement in % gain obtainable when $k < 1$ (over $k = 1$) was 40.4, 17.7, 6.7 and 2.3 for sample size of 25, 50, 100 and 200 families, respectively. The optimum k-value increased with increase in sample size but decreased as the number of traits in the index increased. The relative efficiency of the proposed method was higher as compared to other procedures. When the heritabilities of index-traits were low and their relative economic values were in opposite order to heritability, the relative efficiency of the proposed method was much higher.

Key words: sampling error, selection index, economic weights, index weights, efficiency.

INTRODUCTION

The theoretical basis of index selection was developed by SMITH (1936) and HAZEL (1943), and involves the indirect selection of an unobserved variable, H, by truncation selection of an observed variable, I, which is jointly distributed with H. The index (I) is a linear function of observations, which aims at ranking the population for aggregate genotype, i.e., H (KEMPTHORNE 1957, FALCONER 1989). Index weights can in principle be found by using a least-squares solution to minimise prediction error, or equivalently maximise the product-moment correlation between index values and aggregate genotype.

An index $I = b'X$, where X is a vector of phenotypic deviates from fixed constants (for example, site means) that are assumed to be known, on p traits, is generally used in order to maximise the correlation with overall aggregate genotype $H = a'g$, where a is a vector of known economic weights and g is vector of true breeding values on the same p traits. Suppose $P = \text{var}(X)$ and $G = \text{var}(g)$ are the phenotypic and genetic covariance matrices, respectively, the optimum index is given by

$$b = P^{-1}Ga \qquad [1]$$

The expected response per generation to selection is:

$$R = i\,(b'Pb)^{1/2} \qquad [2]$$

where i is the selection intensity. By using estimates of P and G, namely $\hat{P}$ and $\hat{G}$, the estimated index weights are:

$$\hat{b} = \hat{P}^{-1}\hat{G}a \qquad [3]$$

and predicted genetic response is

$$\hat{R} = i(\hat{b}\hat{P}\hat{b})^{\frac{1}{2}} \qquad [4]$$

The derivation of the selection index is based on the assumption that the population parameters such as heritability (h^2), genetic and phenotypic correlations, phenotypic standard deviations or alternatively genetic and phenotypic variance-covariance matrices, are known exactly. In practice, however, only estimates of these parameters are usually available for constructing the index, and such an index is less efficient than one computed from the true parameters. The effects of errors in the parameter estimates and the loss in efficiency, in relation to size of the sample used for estimation, have been considered by WILLIAMS (1962a, b), HARRIS (1964), and SALES & HILL (1976a, b). With multi-trait indices, it appears that rankings on index values will be more sensitive to errors in estimating genetic and phenotypic covariance matrices if the traits in the index are adversely correlated than if they are favourably correlated (BULMER 1985).

HAYES & HILL (1981) proposed a technique called 'bending' for modifying parameter estimates for multi-trait individual selection. If phenotypic (***P***) and genetic (***G***) variance- covariance matrices are estimated from between- and within-class covariance matrices, ***B*** and ***W*** respectively, in a one-way multivariate analysis of variance, then according to the method the bent genetic and phenotypic covariance matrices, $\hat{\boldsymbol{G}}^*$ and $\hat{\boldsymbol{P}}^*$ are

$$\hat{\boldsymbol{G}}^* = 4(\boldsymbol{B}^* - \boldsymbol{W})/n$$

$$\hat{\boldsymbol{P}}^* = \boldsymbol{B}^* + (n-1)\boldsymbol{W}/n$$

where: $\boldsymbol{B}^* = (1-\gamma)\boldsymbol{B} + \bar{\nu}\gamma\boldsymbol{W}$, where γ is the 'bending' factor and $\bar{\nu}$ denotes the average root of $\boldsymbol{W}^{-1}\boldsymbol{B}$. For $\gamma = 0, \hat{\boldsymbol{P}}^{*-1}\hat{\boldsymbol{G}}^* = \hat{\boldsymbol{P}}^{-1}\hat{\boldsymbol{G}}$ or $\hat{\boldsymbol{b}}^* = \hat{\boldsymbol{b}}$; and for $\gamma = 1$, $\hat{\boldsymbol{P}}^{*-1}\hat{\boldsymbol{G}}^* = \boldsymbol{I}$ or $\hat{\boldsymbol{b}}^* = \boldsymbol{a}$. HAYES & HILL (1981) propose two alternative procedures: (i) if any roots of $\hat{\boldsymbol{P}}^{-1}\hat{\boldsymbol{G}}$ are negative, bend until the smallest root is zero; (ii) bend on the basis of the sample size alone.

ARNASON (1982) used 'bending' to predict the breeding values for multiple traits in small, non-random-mating (horse) population. MEYER & HILL (1983) extended this bending procedure to the case when both individual and sib-information are available, allowing also for different subsets of traits being represented as characters in the selection criteria (in the index) and traits in the (economic) aggregate genotype, respectively. MEUWISSEN & KANIS (1988) used a bending procedure to make an inconsistent set of contrived population parameters (taken from several sources) consistent. A 'rounding procedure' was proposed by TAI (1989) to improve the efficiency of index selection, which involves performing canonical variate analysis on phenotypic and genetic variances of a group of traits estimated from a progeny test experiment.

The ridge regression technique (HOERLE & KENNARD 1970) was devised to circumvent the problem of an ill-conditioned covariance matrix of independent variables in multiple regression analysis, and has been used to develop a ridge selection index (SAXTON 1986, XU & MUIR 1989, VERRYN 1994). Using this procedure, the modified index weights can also be calculated as

$$\hat{\boldsymbol{b}}^* = [\hat{\boldsymbol{P}} + \delta\, diag(\hat{\boldsymbol{P}})]^{-1}\hat{\boldsymbol{G}}\boldsymbol{a},\ \delta \geq 0$$

SAXTON (1986) applies "ridge regression" and "bending" to prediction in breeding. Bending performed better than the ridge procedures, and ridge procedures performed better than least-squares (LS) selection index (SAXTON 1986).

To circumvent the problem of choosing an optimal bending factor for a given sample situation, ESSL (1991) proposed the use of prior knowledge of the genetic parameters. Also, he suggested to use that bending factor which maximises the correlation between true and estimated aggregate genotype as

$$r_{IH} = \frac{\boldsymbol{a}'\boldsymbol{G}\hat{\boldsymbol{b}}^*}{(\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a})^{0.5}(\hat{\boldsymbol{b}}^{*\prime}\hat{\boldsymbol{P}}\hat{\boldsymbol{b}}^*)^{0.5}}$$

The vector of modified index weights, $\hat{\boldsymbol{b}}^*$, is obtained by bending ***B*** towards ***W*** as suggested by HAYES and HILL (1981). Because the ***G*** matrix is unknown, he suggested to use its prior estimate. This new bending strategy of ESSL (1991) was found to be better than those two suggestions of HAYES and HILL (1981) while dealing with a given sample situation.

An alternative viewpoint, proposed in this paper is to give more weight to the economic information when there is doubt about the accuracy of the estimates of genetic parameters. The extreme approach is to use a so called 'base index' (WILLIAMS 1962a, b). An apparently unexplored refinement of this procedure is to regress the computed index towards the base index. The present study was aimed at deriving satisfactory index weights assuming good economic information. The efficiency of the proposed method was compared with the bending approaches of HAYES & HILL (1981) through Monte Carlo simulation.

MATERIALS AND METHODS

There is assumed to be a one-way classification with f groups or half-sib families each of size n, and p traits are recorded on each individual. It is further assumed that the observations are multivariate normally distributed with among- and within-group effects independent of each other. The multivariate analysis of variance table, in the notation of HAYES & HILL (1981), is as follows:

Source	df	SS	MS	E(MS)
Among groups	f-1	S_B	B	$\Sigma + n\Psi$
Within groups	$f(n$-1)	S_W	W	Σ

The matrices of sums of squares and cross-products follow independent central Wishart distributions, $S_B \sim W_p\,[(f-1), \Sigma + n\,\Psi]$ and $S_w \sim W_p\,[f(n-1), \Sigma]$. An estimate of Σ is $\boldsymbol{W}$, whereas the estimate of Ψ or $0.25\boldsymbol{G} = (\boldsymbol{B} - \boldsymbol{W})/\,n$. The dimension of each matrix is $p \times p$ and W_p signifies a Wishart distribution with p as number of variables. The matrix Σ is positive definite and Ψ is positive semi-definite.

It is assumed that 'mass selection' is practised. The expectation of response that is actually achieved when $\hat{\boldsymbol{b}}$ (equation 3) is used subsequently for making selection decisions in the population is

$$\hat{R}^a = i\hat{\boldsymbol{b}}'Ga(\hat{\boldsymbol{b}}'P\hat{\boldsymbol{b}})^{\frac{1}{2}} \qquad [5]$$

As the index coefficients in (3) are vulnerable to sampling errors in $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{G}}$, the expected gain values in (4 and 5) are themselves sensitive to these errors. The estimated index coefficients were regressed towards the relative economic values (REVs), assuming that the REVs are known precisely, as follows:

$$\hat{\boldsymbol{b}}^* = \hat{\boldsymbol{b}}k + \boldsymbol{a}(1-k),\ 0 \le k \le 1 \qquad [6]$$

where $\hat{\boldsymbol{b}}^*$ is the vector of modified index weights. Clearly, when $k = 1$, the selection is solely based on estimated LS index weights and when $k = 0$, the index is 'base index'. These modified index weights were used in place of $\hat{\boldsymbol{b}}$ in (5) to calculate the expected value of genetic response. In this way, some optimum value of k can be sought in order to maximise the value of expected gain. This strategy of finding the optimum k value is theoretically similar to that of finding optimum γ which maximises the correlation between true and aggregate genotype, proposed by ESSL (1991). The optimum value of k was considered to be that which gives the maximum average expected genetic gain (5).

The true achievable genetic response by using the optimum index weights, $\boldsymbol{b}$ (1), was calculated as

$$R^a = \boldsymbol{b}'Ga\,(\boldsymbol{b}'P\boldsymbol{b})^{-1/2} \qquad [7]$$

which is the maximum achievable gain from the known genetic and phenotypic covariance matrices.

Simulation study

Monte Carlo simulation was used to study the effect of sampling errors on the selection index coefficients. In general, the environmental correlation was assumed to be zero but a few sets of parameters were considered to study the effect of non-zero environmental correlations. The number of traits simulated in the index were 2, 3 or 4. Different sets of economic weights were considered, i.e. equal, in same rank order as heritabilities and in opposite rank order to heritabilities. The sampling intensity and the phenotypic variance of each trait were

Table 1. Parameter sets (heritability, h^2, and genetic correlations, r_g) used in simulation. Figure in parenthesis, if any, are the corresponding enviromental correlations

Set.No	h_1^2	h_2^2	h_3^2	h_4^2	$r_{g(1,2)}$	$r_{g(1,3)}$	$r_{g(1,4)}$	$r_{g(2,3)}$	$r_{g(2,4)}$	$r_{g(3,4)}$	Economic weights
1	.20	.70			-.30						11 12 21
2	.15	.45			-.70						11 12 21
3	.25	.35			.40						11 12 21
4	.05	.25			-.50						11 12 21
5	.25	.30	.35		.10	-.30		-.20			111 123 321
6	.15	.30	.45		.30	-.50		-.40			111 123 321
7	.05	.30	.60		.40	-.70		-.60			111 123 321
8	.05	.15	.20		.30	-.50		-.40			111 123 321
9	.20	.30	.30	.40	.10	0	-.20	0	-.10	0	1111 1233 3211
10	.10	.25	.35	.50	.20	.10	-.50	0	-.40	-.20	1111 1233 3211
11	.05	.15	.30	.70	.40	.10	-.80	0	-.60	-.20	1111 1233 3211
12	.05	.10	.15	.20	.20	.10	-.50	0	-.40	-.20	1111 1233 3211
13	.25	.40	.60		.15 (.08)	.20 (.15)		-.35 (-.10)			111 123 321
14	.15	.30	.45		.50 (.30)	.30 (.10)		-.40 (-.20)			111 123 321
15	.38	.31	.53	.70	.15 (.15)	.62 (.32)	-.29 (-.14)	.28 (.22)	-.16 (-.01)	-.20 (-.01)	1111 2133 2311
16	.21	.14	.26	.33	.16 (.29)	.43 (.34)	.11 (.14)	.13 (.19)	.04 (.01)	-.08 (-.01)	1111 2133 2311

assumed to be unity. Different sample sizes considered were 25, 50, 100 and 200 half-sib families. The number of individuals per family was kept constant at 15. The sets of the assumed parameters used for simulation are given in Table 1. Some of these parameters are as used by ESSL (1991).

The parameter sets where non-zero environmental correlations were assumed are also given in Table 1. The properties of these real (assumed) parameter matrices were examined in terms of their eigenvalues and all eigenvalues of these matrices were positive. The number of replicate simulation runs for each scenario was 1,000. For each replicate run, the among- and within-group matrices of sums of squares and cross-products (for varying sample sizes) were sampled independently from Wishart distributions. The genetic and phenotypic variance-covariance matrices were estimated from the sampled among- and within-group matrices of sums of squares and cross-products.

Different index selection procedures compared for their efficiency were:

1. $k = 0$: Equivalent to the base index of WILLIAMS (1962a, b),
2. $k = 1$: Index based on unadjusted sample estimates,
3. *kR* : Stepwise procedure by progressively regressing the estimated index coefficients towards REVs using (6),
4. γ*R*: Stepwise bending procedure by progressively bending ***B*** towards ***W*** (HAYES & HILL 1981).
5. γ*N*: Bending procedure using fixed γ values (suggestion (ii) of HAYES & HILL 1981).
6. *kN*: regressing the estimated index coefficients towards REVs using fixed *k* values (based on sample size alone).

For the procedure γ*R* the optimum γ value was chosen which maximises expression (5). This is theoretically similar to the proposal of ESSL (1991) to use that bending factor which maximises correlation between the true and the estimated aggregate genotype. It is however different from the first suggestion of Hayes & Hill which states that if any roots of $\hat{G}$ or $(\hat{P}^{-1}\hat{G})$ are negative, bend until the smallest root is zero. The ridge index selection procedure was also evaluated but it was found to be consistently inferior and thus it will not be discussed further.

For each sample run, the procedures, i.e. *kR* and γ*R* were applied using the full range (0 to 1) of constants *k* and γ with increment values of 0.01. The expected genetic gain (5) was calculated for each sample for different values of *k* and γ. After that, these gain values were averaged over all samples (i.e. 1000). All of these comparisons were applied on the average values of (5). For the comparisons, i.e., *kR* and γ*R*, the best 'fixed' value of *k* or γ was considered to be that which maximises average value of expression (5). The recommended γ values of HAYES & HILL (1981) apply to the case of 16 individuals per family. Although the individuals per family were kept at 15 in this study the fixed γ values from HAYES & HILL (1981) were used as such for calculations of procedure γ*N*. The efficiency of all procedures investigated was judged by average expected gain (5) relative to an index with true index weights (7). A computer program to generate Wishart distributions was written in SAS/IML (1989) using the algorithm of ODELL & FEIVESON (1966), as explained in KENNEDY & GENTLE (1980).

RESULTS

The results include replicates where estimated genetic correlations fell outside the theoretical range −1 to 1 and also cases of negative heritability estimates, particularly when the number of families were small i.e. $f = 25$. However, the average estimates of these parameters were almost identical to the true parameters even with a sample size of 25 families with 15 individuals each. For a sample of 100 families, the parameter estimates, i.e. of heritability and genetic correlations, were found to fall consistently well within the theoretical bounds i.e., $0 \le h^2 \le 1$ and $-1 \le r_g \le 1$. Similar to HAYES & HILL (1981), the replicates where genetic parameters fell outside the parameter space were included in further calculations.

The plot of results for parameter set 8 (see Table 1) is shown in Figure 1. The number of families for this case are 50 and REVs of traits were in opposite rank order to the heritabilities. The Figure 1 consists of average (over 1000 replications) expected gain (equation 5) for two different procedures, i.e., *kR* and γ*R*, over full range of *k* and γ. It can be seen from this figure that maximum average expected gains are 0.3018 (at $k = 0.88$) and 0.2947 (at $\gamma = 0.56$) for *kR* and γ*R* procedures, respectively. The optimal genetic gain (7)

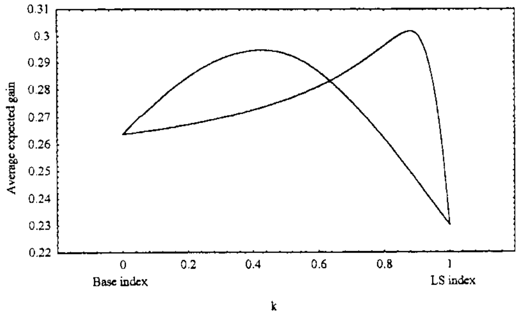

Figure 1. Average (over 1000 replications) expected gain (equation 5) for two index selection procedures *kR* (——) and γ*R* (- - -), $f = 50$, $n = 15$, set 8, REV = 3,2,1. $\gamma = 1$-*k*.

for this case was 0.3527. Thus, the efficiencies are 85.6 and 83.5 % for kR and γR procedures, respectively. For the procedure, γN, the fixed γ value for a sample of 50 families with 16 individuals and three traits, is 0.3 (HAYES & HILL 1981). The value of average expected gain corresponding to $\gamma = 0.3$ is 0.2758 giving efficiency of the procedure, γN, as 78.2 %. The efficiencies of other two procedures ($k = 0$ and $k = 1$) were also calculated in this way.

The simulation results regarding best fixed k values and efficiencies of different index selection methods are given in Tables 2 to 6. The values of k that maximised the average expected gain for various sample situations, are given in Table 2. These values were averaged over all parameter sets given in Table 1. The lower k values were required for small sample size. The overall best k values were 0.65, 0.75, 0.84 and 0.91 for a sample of 25, 50, 100 and 200 families, respectively. This indicates that sample estimates with small number of families are less reliable. Thus, more weight should be given to REVs to overcome the effects of higher sampling bias associated with small sample size. As the sample size increases to 200 families, the estimates are less affected by sampling errors as is evident from higher k value (0.91) required for regressing the least-square index weights towards the REVs. The extent of sampling errors also increased with increase in number of traits included in selection index. For a sample of 25 families, the optimum k values ranged from 0.58 to 0.74 depending upon the number of traits included in index. Thus sampling errors are more important with more traits in an index. However, for a large sample size of 200 families, the number of traits had little impact on the best fixed k values.

Table 2. Average values (which maximises the expectation of genetic gain) of k for different sample sizes. Averaged over all parameter sets in Table 1.

No. of families	Number of traits			Overall
	2	3	4	
25	0.74	0.63	0.58	0.65
50	0.79	0.76	0.70	0.75
10	0.87	0.84	0.81	0.84
200	0.93	0.91	0.90	0.91
Overall	0.83	0.78	0.75	

The relative efficiencies of various index procedures for varying sample sizes are given in Table 3. The % efficiencies presented in this table are averaged over all parameter sets with zero environmental correlations. For the procedure kN, the fixed k values were taken from Table 2 and the efficiency corresponding to the fixed k values were calculated for this procedure. The relative efficiency of $k = 1$ varied from 70.9 to 95.3 % increasing with number of families whereas the efficiency of $k = 0$ was constant at 82.9 % as sample information contributes nothing in this procedure of index selection. The procedure $k = 0$ is better than $k = 1$ only with small sample sizes. The γR procedure was found to be about 2 to 3 % more efficient than γN across all sample sizes. The slightly lower relative efficiency of γN as compared to γR is because γR represents the maximum point on the plot of average expected gain (as shown in Figure 1). The efficiency of the kR procedure was higher than for the other procedures. The procedure kN was found to be less efficient than kR and γR but its efficiency is higher compared to γN.

The results of the effect of number of traits on the efficiency of different procedures are given in Table 4. As the number of trait increases, the efficiency of all procedures also decreases. The rate of declining efficiency is lowest for $k = 0$ whereas highest decline was observed for $k = 1$. The magnitude of the relative efficiency advantage of the procedure kR over γR declines as the number of trait increases but its superiority is maintained throughout.

The relative efficiency of various procedures was also evaluated for different economic weights and

Table 3. Average efficieny (%) of different index procedures for different sample sizes. Averaged over parameter sets with zero enviromental correlations.

No. of families	kR	γR	$k = 0$ (Base index)	$k = 1$ (LS index)	γN	kN
25	91.2	89.8	82.9	70.9	86.6	88.6
50	93.6	93.2	82.9	82.5	90.1	91.0
100	95.6	95.5	82.9	90.5	92.9	93.5
200	97.3	97.3	82.9	95.3	95.1	96.2

Table 4. Average efficieny (%) of different index procedures for different number of traits. Averaged over parameter sets with zero enviromental correlations.

No. of traits	No. of families	kR	γR	$k = 0$	$k = 1$	γN	kN
2	25	94.1	92.2	84.1	80.3	90.2	91.7
	50	96.2	95.6	84.1	88.6	93.4	93.5
3	25	90.5	88.9	83.5	67.2	85.0	88.2
	50	93.1	92.7	83.5	81.7	89.4	90.9
4	25	88.9	88.3	81.3	65.4	84.4	86.0
	50	91.6	91.2	81.3	77.3	87.4	88.6

Table 5. Average efficiency (%) of different index procedures for different relatioship between economic weights and heritabilites. Averaged over parameter sets with zero environmental correlations.

Economic weights heritabilites	No. of families	kR	γR	$k = 0$	$k = 1$	γN	kN
1	25	91.4	90.1	83.5	69.2	86.2	89.0
	50	93.8	93.4	83.5	81.3	88.9	91.1
	100	95.7	95.6	83.5	89.9	91.5	93.6
	200	97.3	97.2	83.5	94.9	93.9	96.3
2	25	94.5	93.5	86.1	80.2	90.7	92.3
	50	96.3	96.1	86.1	89.6	93.6	94.5
	100	97.5	97.4	86.1	94.8	95.7	96.4
	200	98.5	98.4	86.1	97.5	97.1	98.0
3	25	87.5	85.9	79.3	63.4	82.8	84.6
	50	90.7	90.1	79.3	76.6	87.7	87.4
	100	93.7	93.5	79.3	86.7	91.7	91.2
	200	96.2	96.1	79.3	93.3	94.5	94.3

Economic weights: heritabilites = (1) Economic weights are equal for all traits; (2) Economic weights and heritabilites are in same order; (3) Economic weights and heritabilites are in opposite order

Table 6. Average efficiency (%) of different index procedures for different number of traits. Averaged over parameter sets with non-zero environmental correlations.

No. of traits	No. of families	kR	γR	$k = 0$	$k = 1$	γN	kN
3	25	97.2	97.1	96.3	85.9	95.9	97.0
	50	97.9	97.9	96.3	93.9	96.9	97.7
	100	98.5	98.5	96.3	97.1	98.0	98.4
4	25	96.7	96.7	95.9	80.5	95.5	96.5
	50	97.2	97.1	95.9	89.0	95.7	97.0
	100	97.9	97.9	95.9	95.1	96.8	97.8

heritability relationships. Average relative efficiencies for this part of analysis are given in Table 5. The efficiency was higher, for all procedures, when economic weights were in the same rank order as heritabilites. The loss in efficiency, when REVs and heritabilites are in opposite order, is least for kR. Also, the magnitude of higher relative efficiency of kR is maximum when there is opposite relationship between

REVs and heritabilites while it is minimum when REVs and heritabilites are in the same order. As the procedure *kR* regressed the estimated index coefficients towards REVs, therefore, loss in efficiency is lower as compared to other procedures.

Prompted by the results in Table 5, it was decided to compare the efficiency of various procedures in the situation where all the index traits have low heritabilites. For this purpose parameter sets 4, 8 and 12 (see Table 1) were chosen. When economic traits and heritabilities are in opposite rank order, the average efficiency of *kR* was 5.5 and 2.3 % higher than γR with a sample size of 25 and 50 families, respectively.

As mentioned earlier, this study was also designed to examine the impact of non-zero environmental correlations on the relative efficiencies of various procedures. The parameter sets 13 to 16 given in Table 1 were used for this part of analysis. The results for 3 and 4 traits-index are shown in Table 6. It shows that the procedure *kR* has an advantage over all other procedures. One interesting result clearly apparent by comparing Table 4 and Table 6 is that the difference in the relative efficiency of procedure $k = 0$ with others has reduced dramatically in Table 6. Interestingly, $k = 0$ procedures have shown slightly higher relative efficiency as compared to $k = 1$ even with a sample size of 100 families and 4 traits-index.

DISCUSSION

The present study assumes a balanced structure of half-sib families. The effect of sampling errors on the efficiency of index selection, was evaluated using ANOVA estimates of genetic parameters. Different methods of modifying the parameter estimates to increase their reliability, have been proposed in the past. HAYES & HILL (1981) pointed out the possibility of modifying the index weights themselves which has some analogies with the technique of ridge regression (HOERL & KENNARD 1970). The proposed method is a form of index which takes into account the REVs which are assumed to be known precisely. This method lowers the chances of reducing the efficiency of index selection when the breeder is not confident about the reliability of parameter estimates. The proposed method is also an intermediate solution between the two extreme situations, i.e. LS index and 'base index' and maximises the expected genetic gain. This method is similar, although methodologically different, to that of HAYES & HILL (1981). The expectation of response that would be actually achieved, for evaluating the efficacy of the different index selection procedures used in this study, should be theoretically similar to those from maximising the correlation between true and estimated aggregate genotype. The latter criteria have been used in some studies (e.g., ESSL 1991, VERRYN 1994).

Effect of sample size and genetic parameters

The maximum genetic gain was obtained with $k < 1$ when sample estimates of variance components were used. The optimum value of k was found to depend on number of traits, size of experiment and heritabilities. With a two-trait index, the magnitude of sampling errors is comparatively less and thus highest expected gain was obtained with k values closer to 1. With low heritabilities of index traits, the maximum gain was obtained at comparatively low k values. These findings parallel those obtained in other studies (e.g. HAYES & HILL 1981, ESSL 1991) with large γ values. The genetic parameters were found to have influence on the optimum value of bending factor in these studies. The logic behind the γN (i.e., select bending factor on the basis of sample size alone) of HAYES & HILL (1981) was to operate more generally when all roots are positive and genetic parameters are unknown. The relative sub-optimality of procedure γN compared to γR in present study may be because of the effect of genetic parameters on optimum γ.

The calculation of expectation of response in (5) requires that the true parameters be known. However, the improvement of the efficiency of different index selection procedures can only be calculated in Monte Carlo simulation studies (MEYER & HILL 1983). By substituting $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{G}}$ in place of $\boldsymbol{P}$ and $\boldsymbol{G}$ in (5), a roughly linear decline (results not shown) in $\hat{R}$ with decreasing k was obtained. It indicates that the pattern of predicted gains gives no real guidance about optimum k, unless perhaps one is dealing with large matrices and obviously unstable LS index solution which shows clear parallels with multiple regressions.

Effect of environmental correlations

The average expected genetic gain for *kR* procedure was much higher as compared to unmodified index selection (i.e. at $k = 1$). The difference in the relative efficiency of *kR* and $k = 1$ procedures fall drastically when non-zero environmental correlations were taken into account (Table 6). One of the probable reasons for this may be that results in Table 6 are based on parameters set where heritabilities of traits are marginally higher which resulted in the higher relative efficiency of $k = 1$. ESSL (1991) also looked briefly into the aspect of non-zero environmental correlations. Further investigation is required to study the effect of various degrees of environmental correlations on relative efficiency of different index selection procedures.

Effect of different economic weights

Precise knowledge of REVs was assumed in this study. However, the choice of appropriate economic weights can itself be crucial, particularly when adverse genetic correlations are involved. The economic end-product value of observed traits is often difficult to evaluate, especially in tree breeding programmes that involve long generation intervals and uncertain relationships between biological traits and net end-product values. The proposed method proved to be relatively more efficient, under various scenarios, as compared to other procedures considered in this study (Table 4). Its efficiency was much higher particularly when the index involves low-heritabilities traits with REVs in opposite rank order to heritabilities. The results also shows that the proposed method is more efficient than other procedures even when equal REVs have been assigned to different traits. The efficiency of γR was, however, almost identical to that of kR when sample size was more than 50 families.

In this study the efficacy of the proposed method was tested for MANOVA estimation of genetic parameters in a balanced half-sib family structure. Further investigation is required for establishing the efficacy of this method for other genetic parameter estimation methods (like REML) and selection methods (different sources of information: ancestors, individual and progeny). This study does not rule out the possibility that better methods of modifying the parameter estimates exist. Our results offer a simple but very effective procedure which can be further explored.

ACKNOWLEDGEMENTS

S Kumar thanks N. Z. Forest Research Institute, Rotorua, New Zealand, for financial support.

REFERENCES

ARNASON, T. 1982: Prediction of breeding values for multiple traits in small non-random mating (horse) populations. *Acta Agriculturae Scandinavica* **32**:171–176.

BULMER, M. G. 1985: The Mathematical Theory of Quantitative Genetics. Clarendon Press, Oxford, 255 pp.

ESSL, A. 1991: Choice of an appropriate bending factor using prior knowledge of parameters. *Journal of Animal Breeding and Genetics* **108**:89–101.

FALCONER, D. S. 1989: Introduction to Quantitative Genetics. 3rd ed. Longman Scientific & Technical, Essex, England. 438 pp.

HARRIS, D. L. 1964: Expected and predicted progress from index selection involving estimates of population parameters. *Biometrics* **20**:46–72.

HAYES, J. F. & HILL, W. G. 1981: Modification of estimates of parameters in the construction of genetic selection indices ('Bending'). *Biometrics* **37**:483–493.

HAZEL, L. N. 1943: The genetic basis for constructing selection indices. *Genetics* **28**:476–490.

Hill, W. G. & THOMPSON, R. 1978: Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* **34**:429–439.

HOERL, A. E., & KENNARD, R. W. 1970: Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* **12**:55–67.

KEMPTHORNE, O. 1957: An Introduction to Genetic Statistics. Wiley, New York, 545 pp.

KENNEDY, W. J. &, GENTLE, J. E. 1980: Statistical Computing. Marcel Dekker, New York, 591 pp.

MEUWISSEN, T. H. E. & KANIS, E. 1988: Application of bending theory in a pig-breeding situation. *Livestock Production Science* **18**:85–91.

MEYER, K. & HILL, W. G. 1983 A note on the effect of sampling errors on the accuracy of genetic selection indices. *Journal of Animal Breeding and Genetics* **100**:27 –32.

ODELL, P. L. & FEIVESON. A. H. 1966: A numerical procedure to generate a sample covariance matrix. *Journal of American Statistical Association* **61**:199–203.

SALES, J. & HILL, W. G. 1976 a: Effect of sampling errors on the efficiency of selection indices. 1. Use of information from relatives for single trait improvement. *Animal Production* **22**:1–17.

SALES, J. & HILL, W. G. 1976 b: Effect of sampling errors on the efficiency of selection indices. 2. Use of information on associated traits for improvement of a single important trait. *Animal Production* **23**:1–14.

SAS INSTITUTE INC 1989: SAS/IML Software: Usage and reference. Cary, NC: SAS Institute Inc, 501 pp.

SAXTON, A. M. 1986: A comparison of bending and ridge regression in selection index estimation. *In*: Dickerson, GE, Johnson, RK eds. Proc 3rd World Congress on Genetics Applied to Livestock Production, Lincon, 16–22 July 1986 , **12**:449–453.

SMITH, H. 1936: A discriminant function for plant selection. *Annals of Eugenics* **7**:240–250.

TAI, G. C. C. 1989: A proposal to improve the efficiency of index selection by 'rounding'. *Theoretical and Applied Genetics* **78**:798–800.

VERRYN, S. D. 1994: Improving on best linear prediction for tree breeding. PhD Thesis. University of Pretoria, Pretoria.

WILLIAMS, J. S. 1962 a: Some statistical properties of a genetic selection index. *Biometrika* **49**:325–337.

WILLIAMS, J. S. 1962 b: The evaluation of a selection index. *Biometrics* **18**:375–393.

XU, S. & MUIR, W. M. 1990: The application of ridge regression to multiple trait selection indices. *Journal of Animal Breeding and Genetics* **107**:81–88.