Theor Appl Genet (1990) 79:289-293

Selection index: economic weights for maximum simultaneous genetic gain

S. Magnussen

Forestry Canada, Petawawa National Forestry Institute, Chalk River, P.O. Box 2000, Ontario K0J 1J0, Canada

Received July 7, 1989; Accepted October 20, 1989
Communicated by A. R. Hallauer

Summary. Selection indices that maximize the correlation between an individual organism's index score and its breeding value frequently require a priori known "economic" weights before the optimum phenotypic weights can be estimated. The long generation intervals and economic uncertainty that surround forest tree breeding can make the choice of weights arbitrary. In this paper an algorithm is introduced for finding "economic" weights that will ensure maximum simultaneous progress in all index traits. At the outset the traits are assumed to be of equal preference. The solutions are functions of the eigenvalues and eigenvectors of a quadratic form of the additive genetic and phenotypic covariance matrices. Examples of applications in tree breeding emphasize the practical aspects of the method.

Key words: Selection index – Economic weights – Optimization – Simultaneous genetic gain

Introduction

Selection index theory was originally formulated on a profit function, defined as a linear function of traits (Hazel 1943; Lush 1947; Osborne 1957; Smith 1936). Essentially, the index is a linear weighted function of observations on an individual or its relatives that aims at ranking the population for breeding values and, thus, expected progeny performance (Falconer 1981; Kempthorne 1957; Lin 1978). Index weights can be found by maximizing the product-moment correlation between index values and the aggregate genotype (see Baker 1986 and Lin 1978 for references). Selection indices have been used effectively in both animal and plant breeding programs as a superior alternative to other alternative selection methods (Cotterill 1985; Hazel and Lush 1942; Turner and Young 1969). Moreover, selection indices may assist the breeder in a structural interpretation of complex multidimensional genetic parameters.

Numerous properties and aspects of selection indices have been investigated, and the main body of theory as applied to animal and plant breeding is well understood. Examples hereof are: restricted selection indices (Cunningham et al. 1970; Harville 1975; Itoh and Yamada 1988 a; Kempthorne and Nordskog 1959; Tai 1977; Tallis 1962), non-linear utility functions (Itoh and Yamada 1988 b; Mather et al. 1988; Tallis 1968), accuracy and errors (Harris 1964; Hayes and Hill 1980; Hill 1984; Sales and Hill 1976; Smith and Pfaffenberger 1970), and indirect selection (Binet 1965; Griffing 1969).

Many selection indices require knowledge of (i) fixed effects, (ii) variance and covariances (phenotypic and genetic), and (iii) a set of relative "economic" values of component traits. The "economic" values may reflect the market situation, preferences, retrospective results, or simply arbitrarily fixed values. However, the assignment of these weights is frequently a difficult task when the breeder is faced with long generation intervals and poorly quantifiable "economic" merits of the traits (Bridgwater et al. 1983; Itoh and Yamada 1988 a; King et al. 1988). Ideally, an economic weight of a single trait should reflect the marginal benefit from a one unit improvement (Muller and Zeddies 1988); a complete cost benefit analysis may be needed to provide the appropriate figures (Bridgwater and Stonecypher 1979; Chollet and Roman-Amat 1987). Simpler indices, like the desired gains index and the weight-free index, permit avoidance of determining economic weights (Baker 1986).

This study provides an algorithm for obtaining the relative "economic" weights when the purpose of selection is to maximize the genetic merits of a composite genotype simultaneously. Under these particular circumstances (i.e., the trait preferences are assumed, conceptually and implicitly, to be equal at the outset), the optimum relative "economic weights" are part of the solution to the indexing problem. Applications of the algorithm are illustrated with examples from forest tree breeding where the simultaneous improvement of many traits is a common objective.

The model

In constructing the index (I) for n traits, a linear function of an individual's phenotypic value x_i is weighted by coefficients (b_i) designed to maximize the correlation between the function and the individual's genetic worth (H), where H is a linear function of the breeding values of each trait (g_i) weighted by the economic importance attributed to each character (a_i). In its most general form, the unrestricted selection index is:

$$I = b' \cdot X \tag{1}$$

where

- X is an $n \times 1$ vector of phenotypic deviates from fixed constants that are assumed to be known,
- b is an $n \times 1$ vector of weighting factors ($'$ denotes a transposed vector).

The overall economically weighted breeding value is given by:

$$H = a' \cdot g \tag{2}$$

where

- a is an $m \times 1$ vector of relative economic weights,
- g is an $m \times 1$ vector of unobservable breeding values.

Weights for the optimum index are given by:

$$b = P^{-1} \cdot G \cdot a \tag{3}$$

where P is the $n \times n$ phenotypic variance-covariance matrix of the phenotypic observations and G is the $n \times m$ matrix of covariances between elements of X and those of g, i.e.:

$$\mathrm{Var}\begin{pmatrix} x \\ g \end{pmatrix} = \begin{pmatrix} P & G \\ G & G \end{pmatrix}. \tag{4}$$

The covariances in G are assumed to be entirely genetic (additive) in origin. Further, we assume that matrices P and G are known and not changed by selection. In practice, sample estimates $\hat{P}$ and $\hat{G}$ are substituted for the theoretical values [this causes a loss of optimum statistical properties (Hayes and Hill 1980; Sales and Hill 1976)]. When the correlation between I and H is maximized, the expected genetic gain (response) due to selection and breeding is:

$$R'(i) = \mathrm{Cov}(I, H)/\sigma_I \tag{5}$$

where $R'(i)$ denotes the response vector per unit selection intensity (i), and σ_I the standard deviation of the selection index ($\sigma_I^2 = b' \cdot P \cdot b$).

Substitution of the expression for b in Eq. (3) into Eq. (5) and rewriting σ_I yields:

$$R'(i) = a' \cdot G' \cdot P^{-1} \cdot G \cdot (a' \cdot G' \cdot P^{-1} \cdot P \cdot P^{-1} \cdot G \cdot a)^{-1/2} \tag{6}$$

or equivalently

$$R'(i) = a' \cdot Q \cdot (a' \cdot Q \cdot a)^{-1/2} \quad \text{for} \quad Q = G \cdot P^{-1} \cdot G.$$

The matrix Q is a projection matrix of the "economic" vector a, whereas the quadratic form $a' \cdot Q \cdot a$ is a measure of distance. In other words $R(i)$ expresses the relative response (location/distance). Our task is now to maximize $R(i)$ for any given Q. This task is considerably easier if Q can be brought to a diagonal form via some linear transformation.

Let U be such a transformation that satisfies $U \cdot D \cdot U' = Q$, where D is a diagonal matrix with r nonzero diagonal elements. By further requiring that the transformation U is orthonormal, i.e., $U \cdot U' = I$ (I is the identity matrix), we achieve the canonical form under orthogonal similarity, i.e., D contains the eigenvalues λ_i of Q (Searle 1982). For P, a positive definite matrix, and G, positive semidefinite, the existence of U has been given by, e.g., Hayes and Hill (1980) and Searle (1982).

Post-multiplying Eq. (6) with U produces:

$$R'(i) \cdot U = a' \cdot U \cdot D \cdot U' \cdot U \cdot (a' \cdot U \cdot D \cdot U' \cdot a)^{-0.5} \tag{7}$$

define $RU'(i) = R'(i) \cdot U$, and $a' \cdot U = au'$ then we obtain:

$$RU'(i) = au' \cdot D \cdot (au' \cdot D \cdot au)^{-1/2}$$

or equivalently:

$$RU'(i) = (\lambda \# au) \cdot \Big(\sum_i \lambda_i \cdot au_i^2\Big)^{-1/2}$$

where λ stands for the vector of eigenvalues, # denotes the Hadamard product, and the subscripts refer to traits in the index I.

The relative economic weights that can be expected to produce the maximum simultaneous response RU in the transformed space can be found by setting the first derivatives of Eq. (7) with respect to au_i to zero, i.e., for clarity, reference to selection intensity i has been dropped:

$$\frac{\partial RU'}{\partial au_i} = \frac{\lambda_i \Big(\sum_j au_j^2 \cdot \lambda_j\Big)^{1/2} - au_i^2 \cdot \lambda_i^2 \cdot \Big(\sum_j au_j^2 \cdot \lambda_j\Big)^{-1/2}}{\Big(\sum_j au_j^2 \cdot \lambda_j\Big)} = 0 \tag{8}$$

multiplying both sides with $\Big(\sum_j au_j^2 \cdot \lambda_j\Big)^{3/2}$ yields:

$$\lambda_i \cdot \Big(\sum_j au_j^2 \cdot \lambda_j\Big) - au_i^2 \cdot \lambda_i^2 = 0$$

the term $\sum_j au_j^2 \cdot \lambda_j$ is common to all derivatives of RU with respect to au and can, therefore, be considered as a constant C (only relative weights are of interest), i.e.:

$\lambda_i \cdot C - au_i^2 \cdot \lambda_i^2 = 0$ which, when solved for au_i, leads to:

$$au_i = \pm (C/\lambda_i)^{1/2}.$$

Hence, in the transformed space U, choosing relative "economic" weights proportional to the inverse square root of the corresponding eigenvalues of Q simultaneously maximizes the genetic progress of all the transformed index traits. To obtain the solution a in the original sample space, it suffices to perform the inverse transformation of U on au, i.e., $a = U' \cdot au$. Note that the solution can be found among all possible permutations of the signs associated with au_i. Phenotypic weights are, as before, obtained via Eq. (3).

It is well known that eigenvalues derived from sample estimates of variance-covariance matrices are biased (Dey and Srinivasan 1985; James 1985); the largest ones are biased high and the smallest ones are biased toward values that are too low. Friedman (1989) recommends regularization and shrinkage of the sample covariance matrices as a way to mitigate this problem. Although intuitively appealing, these measures depend on the selected loss function and, thus, ultimately on a subjective degree-of-belief that is extraneous to this exposition.

Some examples

The illustrations that follow relate to selection problems in tree improvement programs. Assignment of "economic" weights are notoriously difficult in tree breeding, where long rotations and ever-changing market and technological conditions make predictions of future economic values next to impossible (Bridgwater et al. 1983; King et al. 1988; Wilkens 1987; Namkoong et al. 1969; Namkoong 1976; van Buijtenen 1969). In addition to the optimum solution via the eigenvalues of Q and the back-transformation U', a direct search for the optimum simultaneous solution was also implemented in each of the given examples; in no case did the grid-search find a solution with a higher combined response. The sign permutations of the "economic" weights (a) were, in each case, chosen in order to achieve 'best' response (R).

Example 1

Breeding for higher volume production, higher wood density, lower heartwood content, and less stem taper in jack pine (Pinus banksiana Lamb.) by single-tree selection among 54 half-sib families at age 20 (S. Magnussen and C. Keith, unpublished results). From the replicated progeny trial, the following matrices were obtained:

$X_{\text{mean}} = \{\text{volume, heartwood\%, wood density, stem taper}\}$
$= \{371.0, 9.6, 374.4, 100.0\}$

$$\hat{P} = \begin{pmatrix} 24{,}524.10 & -22.14 & -95.04 & 265.51 \\ -22.14 & 23.07 & -15.26 & -2.59 \\ -95.04 & -15.26 & 562.00 & 5.20 \\ 265.51 & -2.59 & 5.20 & 52.93 \end{pmatrix}$$

$$\hat{G} = \begin{pmatrix} 2{,}224.00 & 27.52 & -96.00 & 166.72 \\ 27.52 & 6.60 & -5.20 & -4.92 \\ -96.00 & -5.20 & 228.00 & 4.52 \\ 166.72 & -4.92 & 4.52 & 17.16 \end{pmatrix}$$

$$\hat{Q} = \begin{pmatrix} 677.5188 & -2.2351 & -33.8060 & 52.3769 \\ -2.2351 & 2.3531 & -2.1895 & -2.5418 \\ -33.8060 & -2.1895 & 92.8681 & 2.1856 \\ 52.3769 & -2.5418 & 2.1856 & 6.5597 \end{pmatrix}$$

$\hat{\lambda} = \{683.4949, 91.2896, 4.4905, 0.0247794\}$

$\hat{a} = \{-0.2676, -4.8019, 0.0053083, 4.1787\}$

$\hat{b} = \{0.009608, -0.2039, 0.04401, 0.023361\}$

From these results, the following relative response per unit selection intensity (i) was obtained:

$\hat{R}(i)/\bar{x}_{\text{mean}} \cdot 100 = \{3.59, -6.58, 1.09, 0.60\}$

Note that the decline in heartwood is in agreement with the selection objectives. Independent single-trait selections produced the following expectation of genetic response per unit selection intensity: {3.2%, −11.3%, 1.9%, 2.3%}. As shown, more gain is possible when selecting for just a single trait. However, independent culling cannot be expected to produce a maximum combined gain (Cotterill and James 1981; Hazel and Lush 1942; Turner and Young 1969).

Example 2

Selection of individual trees with superior breeding values of height growth. Data: height growth from age 1 to age 4 (units: cm) in a replicated progeny trial with jack pine (see Magnussen and Yeatman 1987 for details).

$X_{\text{mean}} = \{24.4, 57.9, 115.9, 189.1\}$

$$\hat{P} = \begin{pmatrix} 33.70 & 29.20 & 24.70 & 25.70 \\ 29.20 & 114.20 & 146.70 & 126.90 \\ 24.70 & 146.70 & 373.60 & 229.50 \\ 25.70 & 126.90 & 229.50 & 542.00 \end{pmatrix}$$

$$\hat{G} = \begin{pmatrix} 9.60 & 8.00 & 2.40 & 7.60 \\ 8.00 & 46.40 & 67.60 & 76.00 \\ 2.40 & 67.60 & 129.60 & 126.80 \\ 7.60 & 76.00 & 126.80 & 140.40 \end{pmatrix}$$

$$\hat{Q} = \begin{pmatrix} 2.8460 & 2.1152 & -0.1818 & 1.6215 \\ 2.1152 & 20.8326 & 32.8652 & 35.9421 \\ -0.1818 & 32.8652 & 58.8119 & 60.5255 \\ 1.6215 & 35.9421 & 60.5255 & 64.2611 \end{pmatrix}$$

$\hat{\lambda} = \{141.7563, 4.4979, 0.4345, 0.062883\}$

$\hat{a} = \{0.8539, 2.3116, 1.8859, -2.9672\}$

$\hat{b} = \{0.1924, 0.05945, 0.058995, -0.039329\}$.

The expected genetic gain in percent of the phenotypic mean values is:

$\hat{R}(i)\% = \{4.57, 4.44, 3.08, 2.10\}$.

Index selection for height at age 4 with heights at ages 1, 2, and 3 as "floating" traits (included in the index but given an "economic" value of zero) resulted in an expected 8% genetic gain per unit selection intensity, whereas direct selection for height at age 4 promised no more than 3.2% improvement per unit selection intensity. It may be argued that simultaneous maximum improvement of all heights is more attractive than merely trying to improve the oldest height. After all, rapid early growth is in itself economically important (less tending costs). Simultaneous improvement of growth at all ages promises to shift the growth curve towards an ideotype (Tallis 1968) that may be harder to achieve if selection is based on trait expression at a single age.

Example 3

Selection of individual trees for superior stem thickness (dbh) and wood density at age 12 in a replicated progeny trial with Douglas fir (Pseudotsuga menziesii (Mirb.)

Franco). The phenotypic and additive genetic matrices as reported by King et al. (1988) were:

$$X_{\text{mean}} = \{\text{diameter (mm), wood density (kgm}^{-3}\text{)}\} = \{86.3, \quad 363\}$$

$$\hat{P} = \begin{pmatrix} 134.1 & -134.7 \\ -134.7 & 639.7 \end{pmatrix}, \quad \hat{G} = \begin{pmatrix} 31.4 & -71.0 \\ -71.0 & 572.3 \end{pmatrix}$$

$$\hat{Q} = \begin{pmatrix} 10.439 & -55.817 \\ -55.817 & 535.182 \end{pmatrix}, \quad \hat{\lambda} = \{541.05, 4.5678\}$$

$$\hat{a} = \{\pm 0.4608, \pm 0.0917\}, \quad \hat{b} = \{\pm 0.1146, \pm 0.05503\}.$$

Faced with the strong negative correlation between stem thickness and wood density in Douglas fir, two solutions to the problem of simultaneously maximizing the genetic response are possible. In the first, the expected response is: $R(i)\% = \{-0.25, 4.5\}$ which is obtained by either of the sign permutations $\{+, +\}$ or $\{-, +\}$ when choosing the a-solution. The second solution $R(i)\% = \{0.25, -4.5\}$ is obtained by choosing either of the complementary sign permutations. King et al. (1988) found the desired "economic" weights by direct search among numerous weight combinations. The chosen weights formed a Minimax solution of $a = \{1.00, 0.19\}$ with an expected response of $R(i)\% = \{2.6, 2.6\}$. Note the similarity of the two solutions ($\hat{a}$) and the sensitivity of the expected response to slight changes in the applied "economic weights".

Conclusions

"Economic" end-product values of observable traits are quite often difficult to assess accurately in animal and plant breeding programs. This is especially true for breeding that involves long generation intervals and uncertain relationships between observable traits and product. Forest tree breeding is a prime example hereof. Shifting market conditions and technological changes in the industry make it hard to predict future economic values when trees are selected for breeding. The method of simultaneously maximizing the gain in all observed traits may, therefore, prove a valuable concept that reflects on structural and functional correlations among traits.

Acknowledgements. Valuable and helpful critique of an earlier version of this paper was kindly provided by Dr. E. B. Godshalk.

References

Baker RJ (1986) Selection indices in plant breeding. CRC Press, Boca Raton/FL

Binet FE (1965) On the construction of an index for indirect selection. Biometrics 21:291–299

Bridgwater FE, Stonecypher RW (1979) Index selection for volume and straigthness in a loblolly pine population. Starkville/MA, pp 132–139

Bridgwater FE, Talbert JT, Jahromi S (1983) Index selection for increased dry weight in a young loblolly pine population. Silvae Genet 32:157–161

Buijtenen HP van (1969) Controlling wood properties by forest management. Tappi 52:257–259

Chollet F, Roman-Amat B (1987) Determination of economic coefficients for multi-trait selection on maritime pine (*Pinus pinaster* Ait.). In: Proc Conf Breed Theory/Progeny Testing/ Seed Orchards, Oct. 13–17, 1986. Williamsburg/VA. IUFRO, North Carolina State University, pp 567–590

Cotterill PP (1985) On index selection. II. Simple indices which require no genetic parameters or special expertise to construct. Silvae Genet 34:64–69

Cotterill PP, James JW (1981) Optimising two-stage independent culling selection in tree and animal breeding. Theor Appl Genet 59:67–72

Cunningham EP, Moen RA, Gjedrem T (1970) Restriction of selection indices. Biometrics 26:67–74

Dey DK, Srinivasan C (1985) Estimation of a covariance matrix under Stein's loss. Ann Stat 13:1581–1591

Falconer DS (1981) Introduction to quantitative genetics, 2nd edn. Longman, London

Friedman JH (1989) Regularized discriminant analysis. J Am Stat Assoc 84:165–175

Griffing B (1969) Selection in reference to biological groups. IV. Application of selection index theory. Aust J Biol Sci 22:131–142

Harris DL (1964) Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20:46–72

Harville DA (1975) Index selection with proportionality constraints. Biometrics 31:223–225

Hayes JF, Hill WG (1980) A reparameterization of a genetic selection index to locate its sampling properties. Biometrics 36:237–248

Hazel LN (1943) The genetic basis for constructing selection indices. Genetics 28:476–490

Hazel LN, Lush JL (1942) The efficiencies of three methods of selection. J Hered 33:393–399

Hill WG (1984) On selection among groups with heterogeneous variance. Anim Prod 39:473–477

Itoh Y, Yamada Y (1988 a) Selection indices for desired relative genetic gains with inequality constraints. Theor Appl Genet 75:731–735

Itoh Y, Yamada Y (1988 b) Linear selection indices for non-linear profit functions. Theor Appl Genet 75:553–560

James M (1985) Classification algorithms. William Collins, London

Kempthorne O (1957) An introduction to genetic statistics. Wiley, New York

Kempthorne O, Nordskog AW (1959) Restricted selection indices. Biometrics 15:10–19

King JN, Yeh FC, Heaman JCH, Dancik BP (1988) Selection of wood density and diameter in controlled crosses of coastal Douglas-fir. Silvae Genet 37:152–157

Lin CY (1978) Index selection for genetic improvement of quantitative characters. Theor Appl Genet 52:49–56

Lush JL (1947) Family merit and individual merit as basis for selection. Am Nat 81:241–261, 362–379

Magnussen S, Yeatman CW (1987) Four years' height growth of 25 jack pine (*Pinus banksiana* Lamb.) families in an accelerated nursery trial. In: Demeritt ME J (ed) Proc 30th NE Forest Tree Improv Conf, Orono, Maine, July 22–24, 1986. School of Forest Research, University of Maine, pp 47–64

Mather DE, Gunsett FC, Allen OB, Kannenberg LW (1988) Estimation of phenotypic selection differentials for predicting genetic responses to ratio-based selection. Genome 30:838–843

Muller P, Zeddies J (1988) Economic valuation of varieties and breeding objectives. Plant Breed 100:59–70

Namkoong G (1976) A multiple-index selection strategy. Silvae Genet 25:199–201

Namkoong G, Barefoot AC, Hitchings RG (11969) Evaluating control of wood quality through breeding. Tappi 52:1935–1938

Osborne R (1957) The use of sire and dam family averages in increasing the efficiency of selective breeding under a hierarchial mating system. Heredity 11:93–116

Sales J, Hill WG (1976) Effects of sampling errors on efficiency of selection indices. Anim Prod 22:1–17

Searle S (1982) Matrix algebra useful for statistics. Wiley, New York

Smith HF (1936) A discriminant function for plant selection. Ann Eugen 7:240–250

Smith WB, Pfaffenberger RC (1970) Selection index estimation from partial multivariate normal data. Biometrics 26:625–634

Tai GCC (1977) Index selection with desired gains. Crop Sci 17:182–183

Tallis GM (1962) A selection index for optimum genotype. Biometrics 18:120–122

Tallis GM (1968) Selection for an optimum growth curve. Biometrics 24:169–177

Turner HN, Young SSY (1969) Quantitative genetics in sheep breeding. Cornell University Press, Ithaca

Wilkens J (1987) Review of the significance of variations in wood structure in the utilisation of *Pinus radiata*. Aust For Res 17:215–232