

A comparison of restricted selection index and linear programming in sire selection

T. R. Famula

Department of Animal Science, University of California, Davis, CA 95616, USA

Received July 24, 1991; Accepted November 15, 1991

Communicated by L. D. Van Vleck

Summary. The objective of restricted selection index is to enhance genetic change in one trait while restricting to zero change in a second trait. Linear programming is another, yet conceptually different, technique to maximize one function while enforcing limits on others. The objective of this research was to compare restricted selection index and linear programming in ability to maximize performance in one trait while limiting change in a second trait to zero. Results of a numerical study demonstrate that linear programming is a more effective method to limit correlated response than restricted selection index. On average, both methods limited response in a correlated trait to zero. However, the squared deviation of actual response in the restricted trait from zero was smaller with linear programming than with restricted selection index. Response to selection in the unrestricted trait is greater with restricted selection index than with linear programming.

Key words: Linear programming – Selection index – Selection responses

Introduction

Animal improvement has long depended on selection index and related prediction methods. Nevertheless, undesirable genetic change in traits correlated with those characters targeted for improvement is common. For example, selection for milk yield in dairy cattle is usually accompanied by a decline in milk fat percent. To overcome this problem Kempthorne and Nordskog (1959) introduced the restricted selection index. The intent is to improve performance in one or more traits while restricting correlated genetic change in others to zero. The solution is based upon linear equality constraints as part of

the computations involved in finding selection index weights. Comprehensive reviews of restricted selection have been presented by Harville (1975), Niebel and Van Vleck (1983) and Brascamp (1984).

Of course there are other mathematical techniques for maximizing linear functions subject to constraints (or, as phrased here, restrictions). The most common of these is linear programming (LP). Long used by animal nutritionists to develop least-cost rations, LP has received only slight attention by breeders (McGilliard and Clay 1983; Jansen and Wilton 1984; Sivaramasingam et al. 1984; Wilcox et al. 1984; Armstrong et al. 1990). In each of those settings the goal of the LP was associated with economics; to either maximize profit or minimize semen cost. The use of LP to meet exclusively genetic goals has not yet been described.

The objective of the present work was to determine which of the two strategies, restricted selection index or linear programming, is more likely to restrict genetic change in one trait while maximizing change in another. Restricted selection index (or, more generally, restricted best linear unbiased prediction) requires the imposition of restrictions on the mixed model equations (Lin 1990). The inherent computational difficulty of this strategy is prohibitive in large sets of data. Here, we intend to use LP to select individuals that maximize the average genetic value in one trait while constraining the average genetic value of a second trait to zero. The usual (unconstrained) selection index values for the two traits are the basis of these computations. Such a process would require little additional computing effort beyond the computation of genetic proofs and less effort than the construction of restricted index equations. The remaining question is whether or not the genetic goals of constraining genetic change through LP are met as efficiently as with restricted selection.

Materials and methods

Notation

Although much has been written on selection index, restricted selection and linear programming, a brief introduction of notation follows. To satisfy the objectives we simplify the problem to two traits: trait 1 being the trait to improve and trait 2, that to be restricted. Phenotypes, corrected for any nongenetic fixed effects, can be represented with the following simple linear model:

$$y_{ij} = a_{ij} + e_{ij} \quad (1)$$

where y_{ij} is an observation on animal j ($j = 1, \dots, n$) for trait i ($i = 1, 2$), a_{ij} represents the additive genetic value of animal j for trait i and e_{ij} is a random residual for record j of trait i . The mean of y_{ij} is zero. Moreover, for all j ,

$$\text{Var} \begin{bmatrix} a_{1j} \\ a_{2j} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} = \mathbf{G}_0; \quad (2)$$

$$\text{Var} \begin{bmatrix} e_{1j} \\ e_{2j} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix} = \mathbf{R}_0;$$

and

$$\text{Var} \begin{bmatrix} y_{1j} \\ y_{2j} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} = \mathbf{V}_0 = \mathbf{G}_0 + \mathbf{R}_0.$$

Multiple trait sire evaluation

The comparison of restricted selection and linear programming will be made through sire selection. Assume that progeny phenotypes are represented by model (1). The goal is then to select sires, based on their progeny means, to improve performance in trait 1 while restricting correlated change in trait 2 to zero. Our first task is prediction of breeding values of both traits for each of q sires (i.e., predict \hat{S}_{mk} for trait $m = 1, 2$ and $k = 1, \dots, q$) from paternal half-sib progeny means. Define this prediction (denoted by the $\hat{\cdot}$ as distinguished from the true breeding value) as:

$$\hat{S}_{mk} = b_{1m} \bar{y}_{1k} + b_{2m} \bar{y}_{2k} \quad (3)$$

where \bar{y}_{ik} is the phenotypic mean of p paternal half-sib progeny in trait i ($i = 1, 2$) of sire k being evaluated for trait m . The selection index weights (b_i) satisfy

$$\begin{bmatrix} t_{11} & t_{12} \\ t_{12} & t_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} \quad (4)$$

for $t_{ii} = (v_{ii} + 0.25(p-1)g_{ii})/p$ for all i and i' .

Thus, given the two progeny means, we can predict the breeding value of each sire for both traits (i.e., \hat{S}_{1k} and \hat{S}_{2k} for $k = 1, \dots, q$).

Restricted selection index

The objective is to find an index to improve trait 1 while restricting the correlated response in trait 2 to zero. Define this restricted index as:

$$\hat{S}_{1k}^R = b_1^R \bar{y}_{1k} + b_2^R \bar{y}_{2k} \quad (5)$$

where \hat{S}_{1k}^R is the restricted index for trait 1 of sire k evaluated from the same progeny means of the index in (3). The restricted selection index weights (b_i^R) satisfy (Kempthorne and Nordskog 1959):

$$\begin{bmatrix} t_{11} & t_{12} & 0.5 g_{12} \\ t_{12} & t_{22} & 0.5 g_{22} \\ 0.5 g_{12} & 0.5 g_{22} & 0 \end{bmatrix} \begin{bmatrix} b_{11}^R \\ b_{21}^R \\ \lambda \end{bmatrix} = \frac{1}{2} \begin{bmatrix} g_{11} \\ g_{12} \\ 0 \end{bmatrix} \quad (6)$$

where t_{11}, t_{12}, t_{22} are as defined in (4). Thus, among a list of q sires, each with progeny means of both traits, a restricted index value can be computed and used as a selection criterion. This criterion affords improvement in trait 1 and, theoretically, no change in trait 2.

Linear programming

Linear programming is described here as an alternative to restricted selection index. We make use of the predictions of equation (3) to choose sires that will maximize performance in trait 1 but limit change in trait 2 to zero. Specifically, for the q sires available for selection, our objective is to choose weights f_k ($k = 1 \dots q$) which maximize:

$$\sum_k f_k \hat{S}_{1k}$$

subject to

$$\sum_k f_k \hat{S}_{2k} = 0; \quad (7)$$

$$\sum_k f_k = 1.0; \quad \text{and}$$

$$1 \geq f_k \geq 0 \quad \text{for all } k$$

The weights, f_k , can be interpreted as frequencies of use for each of the candidate sires.

A numerical study

To compare restricted selection and linear programming we consider a small simulation study. Four different sire selection settings are considered, defined by population parameters and number of progeny per paternal half-sib group. In each case there are 100 candidate sires and two different selection criterion: select the top 4 of 100 available sires or select the top 8 of 100. Progeny means are simulated based on either 25 or 50 paternal half-sib progeny records. As for the population parameters, several are constant across all four simulation settings. Specifically,

$$\mathbf{V}_0 = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

for all simulated phenotypes. However, two different values of \mathbf{G}_0 are considered, either

$$\begin{bmatrix} 0.25 & -0.08874 \\ -0.08874 & 0.35 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0.35 & -0.08874 \\ -0.08874 & 0.25 \end{bmatrix}$$

both of which correspond to a genetic correlation of -0.3 .

First, let us consider restricted selection. Records of 50 or 25 progeny records, on both traits, are generated for 100 simulated sires with one of the two parameter sets. Next, prediction of \hat{S}_{1k}^R ($k = 1, \dots, 100$) (under index (5)) is made for each sire. Sires are ranked by their restricted index, and the top 4 or top 8 are selected. Next, we compute the average true breeding values for traits 1 and 2 of the top 4 or 8 sires. Ideally, the average true breeding value of trait 2 for the chosen sires is zero, but this will certainly not be the case. This process is replicated for a total of 75 data sets of 100 sires each.

The application of linear programming to restricted sire selection follows in a similar manner. The same simulated data used in restricted selection is analyzed with linear programming. Breeding values for each sire are estimated for each trait using index (3). These 200 values (100 sires \times 2 traits) are then used in a linear program to find the weights (f_k) which maximize $\sum_k f_k \hat{S}_{1k}$ subject to $\sum_k f_k \hat{S}_{2k} = 0$ and $\sum_k f_k = 1$. Computation of the weights was conducted with the spreadsheet Quattro Pro (1989, version 1.0).

The choice of sires under LP is quite different from that of truncation selection with a restricted index. For example, under restricted selection we set the number of selected sires prior to sire ranking (in this case 4 or 8 sires). Once chosen, each sire is used with equal frequency in the breeding of females. Our simple linear program is based on a different principle. Here we permit sires to be used in the breeding program with *unequal* frequency (defined here by f_k). What remains is the limit we may wish to impose on the frequency of sire utilization.

If no limit is imposed on the value of f_k , the LP will always solve by choosing two, and only two sires. Similarly, if we force the LP to return with 4 or 8 sires exactly, with no limits on the f_k , two sires are chosen with 'reasonable' frequencies of use and those remaining sires have frequencies less than 0.001 (i.e., the remaining are not considered for use by the LP). As a compromise, to make the two systems of evaluation nearly comparable, we set the limits of f_k to $0.25 \geq f_k \geq 0$ in the 4-sire case and to $0.125 \geq f_k \geq 0$ in the 8-sire case. Any limits we might set can be interpreted as arbitrary. The values of 0.25 and 0.125 were chosen so as not to advantage LP over restricted selection in maximizing the response to selection in trait 1. Though not identical to restricted selection, imposition of these limits helps to provide equitable comparisons of the two sire ranking procedures.

Having solved for those weights, the weighted average of true breeding values for each trait is computed. That is, we compute

$\sum_k f_k S_{1k}$ and $\sum_k f_k S_{2k}$ for comparison with the similar values determined from restricted index selection. As before, this process is repeated on all of the 75 data sets.

We compare the average true breeding values for restricted selection and LP from the results of the 75 data sets. Principal questions include: which selection scheme offers the greatest genetic change in trait 1 and which scheme is more likely to restrict change in trait 2 to zero? Moreover, we also estimate the average squared difference of change in trait 2 from zero to assess the precision of each method at restricting correlated genetic change.

Results

Tables 1 and 2 present results of the numerical study. A quick survey of these values leads to a general conclusion that both strategies, restricted selection index and LP, effectively limit correlated change in trait 2 as expected. Accordingly, we are concerned more with how consistently each selection method limits correlated response rather than the fact that, asymptotically, both converge toward zero genetic change in trait 2. For this, a closer examination of standard errors and squared deviations is necessary.

First consider Table 1. With 50 progeny per sire, the expected superiority of the selected group for trait 1 under restricted selection is 0.887 and 0.775 with top 4 and top 8 sire selection, respectively. With the accompanying standard errors, the observed response is not significantly different from that expected. Of the two competing methods of selection, restricted selection index is lightly better than LP at increasing the level of trait 1. However, LP is better able to limit genetic change in trait 2 than restricted selection index. This is particularly true when one examines the standard errors and squared deviations. In that case, LP is considerably more consistent (i.e., the correlated change in trait 2 is less variable under LP than with restricted selection) in limiting the correlated change in trait 2 than is restricted selection.

The mean squared deviation for trait 2 (variation in response) is computed as $\left\{ \sum_k f_k S_{2k} - 0 \right\}^2$ for LP solutions and averaged. Note that we are using true breeding values of the 100 available sires. For restricted selection this

Table 1. Response to selection (and standard errors) in traits 1 and 2 and variation in trait 2 response in populations with a heritability of 0.25 in trait 1 and 0.35 in trait 2

	Restricted selection				Linear programming	
	Select top 4		Select top 8		Min 4 sires	Min 8 sires
	Estimated ^a	Expected ^b	Estimated	Expected	Estimated	Estimated
<i>Response to selection</i>						
	50 progeny per sire					
Trait 1	0.948 (0.022)	0.887	0.822 (0.016)	0.775	0.922 (0.020)	0.808 (0.015)
Trait 2	-0.054 (0.030)	0.0	-0.029 (0.019)	0.0	-0.016 (0.012)	-0.001 (0.010)
<i>Variation in response^c</i>						
Trait 2	0.072		0.029		0.013	0.009
<i>Response to selection</i>						
	25 progeny per sire					
Trait 1	0.808 (0.021)	0.806	0.719 (0.015)	0.704	0.779 (0.019)	0.704 (0.014)
Trait 2	-0.043 (0.033)	0.0	-0.023 (0.025)	0.0	-0.021 (0.017)	-0.018 (0.013)
<i>Variation in response</i>						
Trait 2	0.087		0.053		0.032	0.019

^a Estimated genetic change as a function of true breeding values

^b Expected genetic change under truncation selection

^c Average value of $\left\{ \sum_k f_k S_{2k} - 0 \right\}^2$ over the 75 simulated data sets

Table 2. Response to selection (and standard errors) in traits 1 and 2 and variation in trait 2 response in populations with a heritability of 0.35 in trait 1 and 0.25 in trait 2

	Restricted selection				Linear programming	
	Select top 4		Select top 8		Min 4 sires	Min 8 sires
	Estimated ^a	Expected ^b	Estimated	Expected	Estimated	Estimated
<i>Response to selection</i>			50 progeny per sire			
Trait 1	1.102 (0.023)	1.089	0.962 (0.017)	0.952	1.094 (0.023)	0.970 (0.017)
Trait 2	0.045 (0.049)	0.0	0.021 (0.017)	0.0	-0.003 (0.011)	-0.007 (0.007)
<i>Variation in response^c</i>						
Trait 2	0.063		0.025		0.010	0.004
<i>Response to selection</i>			25 progeny per sire			
Trait 1	0.984 (0.028)	1.016	0.875 (0.018)	0.888	0.968 (0.023)	0.861 (0.016)
Trait 2	0.053 (0.029)	0.0	0.032 (0.020)	0.0	-0.001 (0.013)	-0.008 (0.011)
<i>Variation in response</i>						
Trait 2	0.065		0.031		0.012	0.009

^a Estimated genetic change as a function of true breeding values

^b Expected genetic change under truncation selection

^c Average value of $\left\{ \sum_k f_k S_{2k} - 0 \right\}^2$ over the 75 simulated data sets

value is computed as [(average S_{2k} of the selected group) - 0]² averaged over the 75 data sets. The zero is included in this expression to reinforce the idea that the expected change in trait 2 is zero.

In all cases, regardless of the population parameters or sample size of progeny numbers, LP provides a smaller mean squared deviation than restricted selection. A similar observation is found in the comparison of standard errors. In general, the more progeny per sire (and hence, increased accuracy of prediction) the smaller the squared deviation from zero.

Table 2 also illustrates the differences between LP and restricted selection. In this case, the trait to be improved has a higher heritability than the trait to be restricted. Regardless of the number of progeny per sire, LP provides a smaller mean change in trait 2 as well as a much reduced mean squared deviation. Standard errors of trait 2 breeding value means are also much smaller under LP, indicative of a more consistent (i.e., smaller variance of correlated response) ability of LP to limit correlated response than restricted selection.

Discussion

As the results of Tables 1 and 2 demonstrate, restricted selection is less likely to limit correlated response than a linear programming procedure using the usual (unconstrained) selection index. This opinion is based, not on bias in prediction, but on the smaller mean squared deviation of the constrained trait from zero when comparing

LP to restricted selection index. The LP procedure is consistently closer to zero change than its restricted selection counterpart.

Restricted selection does have the advantage of an increased response to selection on the trait of interest. In each of the populations examined in the simulation study, gains in trait 1 under LP selection were less than those with comparable restricted selection. Hence, in the absence of an example with clearly defined economic weights for the two traits, a preference among the two alternative selection schemes may be a matter of opinion. An emphasis on maximal response in trait 1 while imposing a limit on a second trait would direct breeders to remain with restricted selection index. However, if the limit on a correlated trait is to be emphasized, at some reduction in response to selection in the trait of interest, then LP provides a reliable alternative.

Linear programming has other technical advantages in its favor. First among these is computational simplicity. The prediction of restricted selection indices, particularly in models with unknown means that require the use of restricted best linear unbiased prediction (Quaas and Henderson 1976; Henderson 1984), demands the construction of more elaborate equations. Such equations are an augmentation of the multiple trait mixed model equations (Henderson and Quaas 1976) that are already difficult to construct and solve. Moreover, once constructed and solved, this restricted index is, in a sense, 'frozen'. Breeders who do not wish to use restricted selection or who choose to restrict other or different traits could not make use of the published index. These breed-

ers would have to set up and solve their own mixed model equations for the restrictions they wish to make. LP offers a more flexible alternative because the unrestricted predicted genetic values can be used as they are published in any restricted selection program.

In our proposed procedure, LP takes advantage of predictions of genetic value from the usual (or unconstrained) mixed model equations. Such predictions could be based on single or multiple trait equations. Our present simulation study uses the equivalent of multiple trait prediction. The LP is then constructed on the predicted genetic values. This procedure provides flexibility for breeders to customize the restricted selection program. Secondly, the additional computations required by the LP are potentially straightforward and easily accomplished (even, as in this research, on commonly available spreadsheets for personal computers).

In comparing the proposed LP procedure with restricted selection, one must insure that the methods are on equal footing. Recall, the objective function of the LP is to choose weights, f_k ($k = 1, \dots, q$), such that $\sum_k f_k \hat{S}_{1k}$ is a maximum subject to $\sum_k f_k \hat{S}_{2k} = 0$. To insure

that LP and restricted selection index are compared equivalently (though clearly not identically), we limit the upper value of f_k so that no one sire can 'dominate' the LP solution. As mentioned earlier, without limits on the solutions of f_k the LP will only choose two sires. To permit equivalent comparisons, we set limits on the f_k of (0.025) for the selection of 4 sires [and (0, 0.125) for the selection of 8 sires]. This limit has the effect of forcing a minimum of 4 (8) sires into the LP solution, with most solutions containing 5 (9) sires.

In light of this upper limit on sire use, the result that restricted selection achieves higher selection response in trait 1 than LP is easily explained. The use of more sires (e.g., generally 5 for LP against 4 for restricted selection) naturally leads to a decrease in the response to selection in trait 1. Likewise, permitting more sires into the solution also allows LP to keep the correlated response in trait 2 closer to zero than restricted selection, although, this alone is not sufficient to explain the success of LP in restricting genetic change. Variation in response of trait 2 (see Tables 1 and 2) for LP when selecting 4 sires is still smaller than the variation of response of trait 2 for restricted selection with 8 sires (a less narrow selection criterion). Hence, the simple LP strategy proposed remains a more reliable means of restricting correlated genetic change than restricted selection.

This formulation of sire selection through LP can be done quite simply, yet it is not without disadvantages. If the final decision is to be how many units of semen to purchase or how many mates to allocate to each sire, the frequencies can be reinterpreted as fractions of the total purchase or allocation. Of course, one could change the

linear programming procedure to one of integer programming; that is, force sire weights to reflect exact integers of units of semen to purchase. Erba et al. (1991) have shown, however, that the additional effort of integer programming results in little improvement over the use of fractions in LP. The difficulty is in adapting this interpretation to dam selection problems. The selection of females to remain in the herd or flock is usually on a yes or no basis. The frequency solutions of the LP are not amenable to such yes/no decisions. In this setting, restricted selection may be the only alternative for constraining correlated response.

In summary, we have seen that linear programming can be an effective genetic tool for restriction of genetic gain for a specified trait. Previous application of LP to problems in animal breeding has emphasized maximization of profits or minimization of semen costs (Shanks and Freeman 1979; Wilcox et al. 1984). Although not addressed in this research, LP could also be applied to other problems in animal improvement (e.g., the minimization of accumulated inbreeding over the course of a selection program, see Quinton et al. 1991). We now see that LP can be a more effective (as defined by a reduction in the variance of correlated change in trait 2) means of constraining correlated response than traditional restricted selection index. LP is also computationally more attractive than restricted selection index. The somewhat cumbersome process of restricted selection can be replaced with the relatively simple procedure of linear programming when constructing breeding programs under constraints.

References

- Armstrong SL, Wilton JW, Schaeffer LR, Pfeiffer WC (1990) Feed intake, weaning weight and net returns for four breeding systems. *J Anim Sci* 68:1848–1856
- Brascamp EW (1984) Selection indices with constraints. *Anim Breed Abstr* 52:645–654
- Erba EM, Fadel JG, Famula TR (1991) The use of integer programming in dairy sire selection. *J Dairy Sci* 14:3552–3560
- Harville DA (1975) Index selection with proportional constraints. *Biometrics* 31:223–225
- Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Ontario
- Henderson CR, Quaas RL (1976) Multiple trait evaluation using relatives' records. *J Anim Sci* 43:1188–1197
- Jansen GB, Wilton JW (1984) Linear programming in selection of livestock. *J Dairy Sci* 67:897–901
- Kempthorne O, Nordskog AW (1959) Index selection with restrictions. *Biometrics* 15:10–19
- Lin CY (1990) A unified procedure of computing restricted best linear unbiased prediction and restricted selection index. *J Anim Breed Genet* 107:311–315
- McGilliard ML, Clay JS (1983) Selecting groups of sires by computer to maximize herd breeding goals. *J Dairy Sci* 66:647–653
- Niebel E, Van Vleck LD (1983) Optimal procedures for restricted selection indices. *Z Tierzucht Zuchtungsbiol* 100:9–26

- Quaas RL, Henderson CR (1976) Selection criteria for altering the growth curve. *J Anim Sci* 43:221 (abstr)
- Quattro Pro (1989) Users guide. Version 1.0. Borland Int Scotts Valley, Calif.
- Quinton M, Smith C, Goddard ME (1991) Comparison of selection methods at the same level of inbreeding. *J Anim Sci* 69 [Suppl]:187
- Shanks RD, Freeman AE (1979) Choosing progeny tested Holstein sires that meet genetic goals at minimum semen cost. *J Dairy Sci* 62:1429–1434
- Sivarajasingam S, Burnside EB, Wilton JW, Pfeiffer WC, Grieve DG (1984) Ranking dairy sires by a linear programming dairy farm model. *J Dairy Sci* 67:3015–3024
- Wilcox ML, Shumway CR, Blake RW, Tomaszewski MA (1984) Selection of artificial insemination sires to maximize profits. *J Dairy Sci* 67:2407–2419