

# The equivalence of two linear methods for the improvement of traits expressed as ratios

T. R. Famula

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

Received August 17, 1989; Accepted November 9, 1989  
Communicated by D. Van Vleck

**Summary.** Many traits of interest to animal breeders can be expressed as ratios. Yet there remains no uniquely agreed upon method for the genetic evaluation for ratio traits. To generalize, livestock breeders make direct use of ratios (e.g., feed/gain) or linear approximations to ratios. Dairy breeders, on the other hand, tend to use ratios of linear predictors of genetic merit for the evaluation of ratio traits (e.g., fat percent). In the present note, we demonstrate that the two methods are nearly, though not exactly, identical. The proof relies on the expression of the approximate correlation between two ratios.

**Key words:** Ratios – Selection – Selection Index

## Introduction

Many of the theoretical and statistical techniques developed by animal breeders focus on linear predictors applied to normally distributed phenotypes. The emphasis on normality is justified by the abundance of traits which fit this distribution. The emphasis on linear predictors centers around simplicity of computation. Notwithstanding this traditional approach, not all traits are normally distributed or suited to linear predictors. One such class of traits is that set commonly defined as ratios. One of the most important of these traits is feed efficiency, the ratio of body weight gain to feed consumption. Other examples of ratio traits are percent milk fat, percent milk protein, or percent retail cuts.

Surprisingly, there is no consensus among animal breeders on how animals should be evaluated for ratio traits. For example, dairy bulls are evaluated for milk fat percent (PDF%) by taking the ratio of one-half of their estimated additive genetic value for pounds of fat (PDF) to one-half of their estimated additive genetic value

for pounds of milk (PDM [i.e.,  $\text{PDF}\% = (\mu_F + \text{PDF}) / (\mu_M + \text{PDM})$ ], where  $\mu_F$  and  $\mu_M$  are the breed means for fat and milk; e.g., Wiggans et al. 1988]. However, Lin (1980) advocates the use of a linear approximation of the ratio of two traits (see also Gunsett 1984). Other suggestions have been put forward by Clark and Touchberry (1962), Arboleda et al. (1976), and by Eisen (1977), but none has been as effective as the linear approximation (Lin 1980).

The objective of this paper is to compare two alternatives for the improvement of ratio traits. These two alternatives are: (1) the use of a linear approximation, and (2) the ratio of linearly predicted genetic values. Both methods have been applied to livestock breeding. In effect, dairy breeders, in their evaluation of PDF%, have taken a path different from livestock breeders in their evaluation of feed efficiency. The question that remains is which of the two strategies offers the most effective means of genetic improvement.

## Materials and methods

### Notation

Consider a set of unrelated candidates for selection, each with two phenotypes  $x_1$  and  $x_2$ . For example  $x_1$  could be some measure of body weight gain and  $x_2$  feed intake. The selection objective may be the genetic improvement of feed efficiency. Let a simple model for each phenotype be

$$\begin{aligned} x_1 &= \mu_1 + a_1 + e_1 \\ x_2 &= \mu_2 + a_2 + e_2 \end{aligned} \quad (1)$$

where  $E[x_1] = \mu_1$ ,  $E[x_2] = \mu_2$ ,  $E[a_1] = E[a_2] = 0$ ,  $E[e_1] = E[e_2] = 0$  and

$$\begin{aligned} \text{Var} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \text{Var}[x] = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} = V \\ \text{Var} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} = G \end{aligned} \quad (2)$$

$a_1$  and  $a_2$  can be considered the additive genetic values for traits 1 and 2. Accordingly,  $V$  represents the phenotypic covariance matrix and  $G$  the additive genetic covariance matrix. The objective of the breeding program is, therefore, to improve  $(\mu_1 + a_1)/(\mu_2 + a_2)$ .

*Linear approximation*

A more thorough derivation of the linear index is available in Lin (1980). The purpose here is to identify this index in the notation of model (1). Specifically, the objective of this index is to predict  $T = (\mu_1 + a_1)/(\mu_2 + a_2)$  with a linear function of  $x$ , i.e.,  $b'_L x$  where  $b_L$  satisfies

$$V b_L = G h \tag{3}$$

for  $h = [1 - \mu_1/\mu_2]$ . Using a different approach, Gunsett (1984) arrived at the same index. The vector  $h$  can be thought of as a vector similar to net economic values. However, the values derived are intended to approximate a ratio in the form of a linear index. Accordingly, candidates for selection are chosen on the index;  $I_L = b'_L x = h' G V^{-1} x$ . The subscript 'L' is intended to denote this as the Lin (1980) index.

*Ratio of two linear indices*

As cited above, other investigators evaluate ratio traits by taking the ratio of two linear indices. To present this, consider reparameterizing  $G$  as two column vectors:

$$G = [g_1 : g_2] \tag{4}$$

Best linear predictors of  $a_1$  and  $a_2$  may be derived as

$$\begin{aligned} \hat{a}_1 &= b'_{R1} x = g'_1 V^{-1} x \\ \hat{a}_2 &= b'_{R2} x = g'_2 V^{-1} x \end{aligned} \tag{5}$$

Accordingly, candidates for selection are chosen on the index:  $I_R = (\mu_1 + \hat{a}_1)/(\mu_2 + \hat{a}_2)$ . The subscript 'R' is intended to denote this as the ratio of two linear indices, as distinct from  $I_L$ .

*Example*

Table 1 presents a hypothetical data set of ten individuals scored for two traits,  $x_1$  and  $x_2$ . Also included in Table 1 are values for  $I_L$  and  $I_R$  and the rank of each individual. Indices were computed with the following parameters, taken from Davis (1987):

$$\begin{aligned} \mu_1 &= 190 \\ \mu_2 &= 790 \end{aligned} \quad V = \begin{bmatrix} 235 & 598 \\ 598 & 3708 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 106 & 337 \\ 337 & 1854 \end{bmatrix}$$

The linear approximation index is of the form  $I_L = 0.3068x_1 - 0.0788x_2$ . Predictions of  $a_1$  and  $a_2$  via the usual selection index are

$$\begin{aligned} \hat{a}_1 &= 0.3728x_1 + 0.0308x_2 \\ \hat{a}_2 &= 0.2743x_1 + 0.4558x_2 \end{aligned}$$

and are used in the index  $I_R = (\mu_1 + \hat{a}_1)/(\mu_2 + \hat{a}_2)$ . The observation most noteworthy in Table 1 is the change of rank between animals E and F under the two indices. The conclusion from this small hypothetical example is that indices  $I_L$  and  $I_R$  are not identical. Determining which of the two maximizes the genetic change in  $(\mu_1 + a_1)/(\mu_2 + a_2)$  is the next objective.

**Results**

We begin with the assumption that the objective of each index is to serve as a selection aid in the improvement of

**Table 1.** Example data set of ten animals evaluated with a linear approximation index ( $I_L$ ) and the ratio of two linear indices ( $I_R$ )<sup>a</sup>

Animal	$x_1$	$x_2$	$I_L$ (rank)	$I_R \times 100$ (rank)
A	223	761	12.41 (1)	25.630 (1)
B	208	750	8.68 (2)	25.168 (2)
C	188	761	1.67 (3)	24.266 (3)
D	199	834	-0.71 (4)	23.963 (4)
E	171	781	-5.12 (5)	23.394 (6)
F	176	801	-5.16 (6)	23.398 (5)
G	180	832	-6.38 (7)	23.259 (7)
H	144	756	-11.43 (8)	22.550 (9)
I	162	833	-11.98 (9)	22.557 (8)
J	143	817	-16.55 (10)	21.954 (10)

<sup>a</sup> Phenotypic and genetic parameters taken from Davis (1987)

$T = (\mu_1 + a_1)/(\mu_2 + a_2)$ . When phenotypes and breeding values are jointly, normally distributed, change in  $T$  can be predicted from the expression

$$\Delta T = \frac{\text{Cov}(T, I)}{[\text{Var}(I)]^{1/2}} D \tag{6}$$

where  $I$  is the criterion for truncation selection and  $D$  represents the intensity of selection. Values for  $D$  may be found in many texts (Van Vleck 1983; Falconer 1989).

Our objective is to determine which of the two indices,  $I_R$  or  $I_L$ , leads to greater response in  $T$ . Given that neither  $I$  nor  $T$  is normally distributed, Eq. (6) is not suitable. Nevertheless, we should be able to evaluate which of the indices will yield the greater selection response by evaluating the correlation between  $T$  and each index. Intuitively, this seems a logical approach to evaluating which index will result in the greater response to selection. Thus, to evaluate which index is most efficient we direct our attention to the correlation between each index and  $T = (\mu_1 + a_1)/(\mu_2 + a_2)$ .

Using Eqs. (3), (4), and (5), one can define

$$b_L = h_1 b_{R1} + h_2 b_{R2} = b_{R1} - \frac{\mu_1}{\mu_2} b_{R2} \tag{8}$$

This result is helpful in the evaluation of  $r_{TI}$ .

To approximate  $r_{TI}$  requires a result of Pearson (1897) for the approximate correlation between two ratios. This expression has been cited by Sutherland (1965) and Gunsett (1984), as well as in a variety of other references on selection for ratio traits. As a point of reference, the Pearson (1897) approximation to the correlation of two ratios, e.g.,  $x_1/x_3$  and  $x_2/x_4$  is

$$\frac{r_{12} v_1 v_2 - r_{14} v_1 v_4 - r_{23} v_2 v_3 + r_{34} v_3 v_4}{(v_1^2 + v_2^2 - 2r_{13} v_1 v_3)^{1/2} (v_2^2 + v_4^2 - 2r_{24} v_2 v_4)^{1/2}} \tag{9}$$

where  $r_{ij}$  is the correlation between variables  $x_i$  and  $x_j$  and  $v_i$  is the coefficient of variation for variable  $i$  ( $\sigma_i/\mu_i$ ).

With a bit of algebra, and making use of Eqs. (8) and (9) one can show that for  $T = (\mu_1 + a_1)/(\mu_2 + a_2)$

$$r_{T_{I_R}} = r_{T_{I_L}}$$

and

$$r_{I_R I_L} = 1.0.$$

It should be noted that although the algebra indicates a strict equality in the correlation of the indices with the true value, our example shows this to be untrue for all situations. Obviously, Eq. (9) is only an approximation of the correlation between ratios. Thus, as the example illustrates, the correlation between  $I_R$  and  $I_L$  cannot be identically equal to 1.0. Rather, because  $I_R$  and  $I_L$  differ in the manner in which they rank candidates for selection (Table 1), evaluation of  $r_{TI}$  shows they are nearly, though not strictly, identical. Evidently, the two indices are close enough that the approximate correlation of Pearson (1897) cannot distinguish between the two. This dichotomy, between the algebraic results and the example, can only be explained by a failure of Eq. (9) to exactly define the correlation of two ratios.

**Discussion**

Given that  $I_R$  and  $I_L$  can lead to different rankings of candidates for selection, their approximate equality is an unexpected result. So, too is their equal correlation to the trait to be improved,  $T = (\mu_1 + a_1)/(\mu_2 + a_2)$ . This inconsistency is due to the approximation of the true correlation of the ratios. Logic dictates that, because individuals can rank differently, one index must be superior to the other in producing genetic change. Yet Pearson's (1897) approximation is not sensitive enough to reveal this difference. No doubt the difference in response to using  $I_R$  or  $I_L$  must be small. However, it should be detectable. A small simulation reveals just this result.

The following parameter values, taken from Jara-Almonte and White (1973), were used to generate data on two traits  $x_1$  and  $x_2$ :

$$E \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 156.25 \end{bmatrix}; \quad V = \begin{bmatrix} 3.8554 & 7.1191 \\ 7.1191 & 146.0642 \end{bmatrix}$$

and

$$G = \begin{bmatrix} 0.8608 & 2.1311 \\ 2.1311 & 20.2836 \end{bmatrix}.$$

Genotypic and environmental values for both traits were generated from pseudo-normal random deviates for 40 unrelated individuals. The objective, as in the example cited in Table 1, is to improve the genetic ratio,  $(\mu_1 + a_1)/(\mu_2 + a_2)$ , using the linear approximation index  $I_L$ , the ratio of linear indices  $I_R$ , or direct selection on  $x_1/x_2$ . Of

**Table 2.** Mean genotypic values for traits 1 and 2 after selection on a linear approximation ( $I_L$ ), ratio of linear indices ( $I_R$ ) and direct selection on  $x_1/x_2$

Genera- tion	Trait	Selection on		Direct selection
		$I_L$	$I_R$	
1	1	0.000846	0.000846	0.000846
1	2	0.000896	0.000896	0.000896
2	1	0.374799	0.373546	0.265750
2	2	0.398327	0.391104	-0.268042
Mean change in ratio		0.002062	0.002060	0.001919

the 40 individuals, the top 15 are selected based on  $I_L$ ,  $I_R$ , or  $x_1/x_2$ . Of the 15, 5 are chosen at random to be males, and each is mated at random to 2 of the remaining 10 females. Each mating yields 4 offspring, for a total of 40 individuals in the second generation. This process is repeated for 10,000 data sets of 80 individuals each.

The same random "seed" is used to generate the first generation of each set of data. Thus, the first 40 individuals in each of the 10,000 data sets have the same genotypes and phenotypes. In this way, one can directly compare the second generation values generated from selection on  $I_L$ ,  $I_R$ , or  $x_1/x_2$ , since each first generation was evaluated on precisely the same data. Obviously, because different individuals may have been selected as parents, the second generation phenotypes and genotypes should be different in each set of data.

This result is shown in Table 2, although it is important to note that the second generation means of the genotypes of traits 1 and 2 are nearly, though not exactly, equal. The indication is that these two methods are similar, but not identical, as the approximate correlation suggests. Practically the two indices are identical. After all, the progress in generation 2 from selection on  $I_L$  is greater than the progress with  $I_R$  by only 0.00002. Such differences would not be detectable in any field application of these methods. In both instances,  $I_L$  and  $I_R$  are superior to selection on the phenotypic ratio. This expected result was first presented by Lin (1980).

The relative changes in the two traits from one generation to the next should also be noted. That is, the response to selection is greater in trait 2 (the denominator) than for trait 1, regardless of the fact that trait 1 has a higher heritability. The recent simulation results of Essl (1989) confirm this as a result to be expected.

The difference between these two strategies, however small, does present a result quite different from that of the prediction of quadratic merit (Wilton et al. 1968). The basic difference between  $I_L$  and  $I_R$  is that  $I_L$  attempts to predict the ratio directly, whereas  $I_R$  is based on prediction of the numerator and denominator separately. The two pieces are then brought together in the final index.

For the prediction of quadratic merit, these two procedures yield the same result. That is, predicting a quadratic function directly is equivalent to predicting each term separately and substituting these values into the quadratic function. However, this result is peculiar to quadratic models and not to cubic functions of genetic merit (Mao 1969). The same is evidently true for the nonlinear function discussed in this report, ratios.

If we accept that the two indices are equal, we may yet ask, which is to be recommended? A decision may then be based on ease of computation or interpretation. If these are the criteria, one would favor  $I_R$ . This is particularly true in the more realistic situation, where the phenotypic means are unknown and may vary with sex, season, herd, or age. Because the means may be added to the index rather than be incorporated into the index (as in  $I_L$ ),  $I_R$  offers greater flexibility. Although the simulation considers only one set of parameters, the approximate correlation results should be consistent across all forms of the phenotypic means and  $V$  and  $G$ .

## References

- Arboleda CR, Harris DL, Nordskog AW (1976) Efficiency of selection in layer-type chickens by using supplementary information on feed consumption. 1. Selection index theory. *Theor Appl Genet* 48:67–73
- Clark RD, Touchberry RW (1962) Effect of body weight and age at calving on milk production in Holstein cattle. *J Dairy Sci* 45:1500–1510
- Davis ME (1987) Divergent selection for postweaning feed conversion in beef cattle: predicted response based on an index of feed intake and gain vs feed: gain ratio. *J Anim Sci* 65:886–895
- Eisen EJ (1977) Restricted selection index: an approach to selecting for feed efficiency. *J Anim Sci* 44:958–972
- Essl A (1989) Selection for a ratio of two traits: results of a simulation study. *J Anim Breed Genet* 106:81–88
- Falconer DS (1989) *Introduction to quantitative genetics*, 3rd edn. Longman, Burnt Mill, England, pp 354–355
- Gunsett FC (1984) Linear index selection to improve traits defined as ratios. *J Anim Sci* 59:1185–1193
- Jara-Almonte M, White JM (1973) Genetic relationships among milk yield, growth, feed intake and efficiency in laboratory mice. *J Anim Sci* 37:410–416
- Lin CY (1980) Relative efficiency of selection methods for improvement of feed efficiency. *J Dairy Sci* 63:491–494
- Mao IL (1969) Cubic and quartic selection indices: formula and the equivalency of conventional and substitutional procedure. Notes. Department of Animal Science Cornell University, Ithaca/NY
- Pearson K (1897) Mathematical contributions to the theory of evolution – on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc R Soc London* 60:489–498
- Sutherland TM (1965) The correlation between feed efficiency and rate of gain, a ratio and its denominator. *Biometrics* 21:739–749
- Van Vleck LD (1983) Notes on the theory and application of selection principles for the genetic improvement of animals. Cornell University, Ithaca/NY, pp 108–109
- Wiggans GR, Misztal I, Van Vleck LD (1988) Implementation of an animal model for genetic evaluation of dairy cattle in the United States. In: Schmidt GH (ed) *Proceedings of the Animal Model Workshop*. American Dairy Science Association, Champaign/IL, vol 71, suppl 2, pp 54–69
- Wilton JW, Evans DA, Van Vleck LD (1968) Selection indices for quadratic models of total merit. *Biometrics* 24:937–949