THE EFFICIENCY OF ANCESTOR RECORDS IN ANIMAL SELECTION

S. R. SEARLE *

Cornell University, Ithaca, N.Y.

Received 11.ii.63

1. INTRODUCTION

ANIMAL selection programmes are usually based on estimates of genetic merit obtained from an animal's own performance records, those of its progeny, or a combination of both. Records of parents and other relatives are sometimes utilised also. But when an animal is young and has neither performance records nor progeny, the only information available about its breeding value is the performance of its relatives, in particular of its direct ancestors, and of siblings, if any. In such cases selection programmes have to be based on estimates of breeding value derived from the records of relatives, a procedure which is commonly used in the artificial breeding of dairy cattle when selecting young unproven bulls for the implementation of sire-sampling programmes, as described, for example, in Searle (1962). The efficiency of several combinations of parent and grandparent records has been considered by Kempthorne (1960), who conjectures on "what is the best possible predictor assuming information on the parent and infinite arrays of all possible collateral or ancestral relatives". Osborne (1957), Jardine (1958) and Young (1961) have partially discussed this matter in extensive presentations of formulæ for combining records on parents with those of the individual and/or its relatives and/or its progeny. Briefly discussed also is the use of parent records alone, as is the efficiency of using just parent and grandparent records by Kempthorne (1960), but none of these authors deals with the questions (i) what is the *maximum* efficiency that can be gained from ancestor records alone and (ii) how much is to be gained by going further back in the pedigree than either parents or parents and grandparents? It is the purpose of this paper to consider these and allied matters in some simple situations.

A suitable criterion for discussing the relative merits of different selection indexes is the magnitude of the correlation r , between index and true genetic merit. Reference could be made to several papers for a development of this correlation (*e.g.* Czekanowski, 1933; Hazel, 1943; Comstock 1948; Henderson, 1961) but it is useful to state our own, in matrix notation. Suppose that

$$I = b_1X_1 + b_2X_2 + b_3X_3$$

is an index for the genetic value y , of an individual, using the single

* Paper No. BU-74 of the Biometrics Unit and No. 433 of the Plant Breeding Department.

performance records X_1 , X_2 and X_3 of three relatives. It can be written as

$$I = b'x \quad (1)$$

where b is the vector of b -coefficients and x is the vector of records. The selection index procedure then involves obtaining b as the solution to the equation

$$Vb = w \quad (2)$$

where $V\sigma^2$ is the variance-covariance matrix appropriate to the vector of records x , and $w\sigma^2$ is the vector of covariances of y with x ; and $h\sigma^2$ is the variance of y . In general h is heritability in the "broad sense", Lush (1947), but in practice the "narrow sense" definition is used, y then being defined as the additive genetic merit of the individual. Now from (2) we have

$$b = V^{-1}w. \quad (3)$$

Hence the covariance of I with y is

$$\text{cov}(Iy) = b' \text{cov}(xy) = b'w\sigma^2 = w'V^{-1}w\sigma^2$$

and likewise the variance of I is

$$\sigma_I^2 = b'E(xx')b = b'Vb\sigma^2 = w'V^{-1}w\sigma^2.$$

Thus

$$r = r_{Iy} = \text{cov}(Iy)/\sigma_I\sigma_y = \sqrt{w'V^{-1}w/h}. \quad (4)$$

There is a wide variety of uses for these equations for they are applicable to situations of performance-testing, progeny-testing, pedigree evaluation such as we are primarily interested in here, and combinations of all three. For example, in the simple case of performance-testing, V and w are both scalar, and if just a single record from the individual is used, $V = 1$ and $w = h$, so that $b = h$ and $r = \sqrt{h}$. Likewise in progeny-testing V and w are again scalar, and when using single records of n progeny

$$V = [1 + 0.25(n-1)h]/n, \quad w = \frac{1}{2}h, \quad b = \frac{2nh}{[4 + (n-1)h]} \\ \text{and } r = \sqrt{nh/[4 + (n-1)h]}$$

These are well-known results. Following others we shall refer to r as the efficiency of the index; thus $\sqrt{h}$ is the efficiency of performance-testing with a single record and $\sqrt{nh/[4 + (n-1)h]}$ is the efficiency of progeny-testing with single records of n progeny. Sometimes the term "relative efficiency" is used also, for the efficiency of an index expressed as a fraction of the efficiency of another index.

2. PEDIGREE EVALUATION

We will now use (4) to obtain expressions for the efficiency of ancestor records, and so consider to what limit (if any other than 1.00) the efficiency increases through making use of ancestors progressively further back in the pedigree, and by how much this limit differs from the efficiency of using records on parents alone or on parents and grandparents. An allied question, bearing on the value of ancestor records at all, is to compare their efficiency with that of

performance tests and progeny-tests, for in due course the necessary records will be available for the latter.

Since there can be great variation in the number of records available from ancestors we will assume for simplicity's sake that each ancestor has a single record, and that in evaluating the index the records in each generation are averaged. Thus for three generations X_1 will represent the mean record of the two parents (one record on each), X_2 the mean of the four grandparent records and X_3 the mean of the eight great-grandparent records. Equation (2) is then

$$\begin{Bmatrix} \frac{1}{2} & \frac{1}{4}h & \frac{1}{8}h \\ \frac{1}{4}h & \frac{1}{4} & \frac{1}{8}h \\ \frac{1}{8}h & \frac{1}{8}h & \frac{1}{8} \end{Bmatrix} \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix} = \begin{Bmatrix} \frac{1}{2}h \\ \frac{1}{4}h \\ \frac{1}{8}h \end{Bmatrix}$$

and equation (4) becomes

$$r^2 = \begin{Bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \end{Bmatrix} \begin{Bmatrix} \frac{k}{2} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{4} & k & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & k \end{Bmatrix}^{-1} \begin{Bmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{8} \end{Bmatrix}$$

where $k = 1/h$. If we write this for n generation of ancestors as

$$r_n^2 = \mathbf{t}'_n \mathbf{A}_n^{-1} \mathbf{t}_n$$

the general form of $\mathbf{t}_n$ and $\mathbf{A}_n$ is readily apparent. Thus, putting $p = \frac{1}{2}$ to simplify notation,

$$\mathbf{A}_n = \begin{Bmatrix} kp & p^2 & p^3 & \dots & p^n \\ p^2 & kp^2 & p^3 & \dots & p^n \\ p^3 & p^3 & kp^3 & \dots & p^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p^n & p^n & p^n & \dots & kp^n \end{Bmatrix}$$

and $\mathbf{t}'_n = (p \ p^2 \ p^3 \dots p^n)$. If $\mathbf{A}_n^{-1} = \{a_n^{ij}\}$, a_n^{ij} being the term in the i th row and j th column of $\mathbf{A}_n^{-1}$, then

$$r_n^2 = \sum_{i=1}^n \sum_{j=1}^n p^{i+j} a_n^{ij} \quad (5)$$

But from the elements of the first row of the identity $\mathbf{A}_n \mathbf{A}_n^{-1} = \mathbf{I}$ we have

$$kpa_n^{11} + \sum_{i=2}^n p^i a_n^{i1} = 1$$

and

$$kpa_n^{1j} + \sum_{i=2}^n p^i a_n^{ij} = 0 \text{ for } j > 1.$$

Thus
$$\sum_{i=1}^n p^i a_n^{i1} = 1 + (1-k)pa_n^{11} \quad (6)$$

and
$$\sum_{i=1}^n p^i a_n^{ij} = (1-k)pa_n^{1j} \text{ for } j > 1.$$

Substituting in (5) gives

$$\begin{aligned} r_n^2 &= p[1 + (1-k)pa_n^{11}] + p(1-k) \sum_{j=2}^n p^j a_n^{1j} \\ &= p + p(1-k) \sum_{j=1}^n p^j a_n^{1j}, \end{aligned}$$

and since A is symmetric $a_n^{1j} = a_n^{j1}$, so that equation (6) can be applied again to give

$$r_n^2 = p(2-k) + p^2(1-k)^2 a_n^{11}. \quad (7)$$

Thus the only element of A_n^{-1} affecting r_n is the leading term, which can be obtained as

$$a_n^{11} = (\text{cofactor of the leading term of } A_n) / |A_n|.$$

From the definition of A_n and t_n we see that A_n can be written in partitioned form as

$$A_n = \begin{Bmatrix} kp & pt'_{n-1} \\ pt_{n-1} & pA_{n-1} \end{Bmatrix}$$

and hence

$$a_n^{11} = |pA_{n-1}| / |A_n|.$$

Further, from the well-known result for the determinant of a partitioned matrix

$$\begin{aligned} |A_n| &= [kp - pt'_{n-1}(pA_{n-1})^{-1}pt_{n-1}] |pA_{n-1}| \\ &= p(k - r_{n-1}^2) |pA_{n-1}|. \end{aligned}$$

Hence

$$a_n^{11} = 1/p(k - r_{n-1}^2)$$

and

$$r_n^2 = p(2-k) + \frac{p(1-k)^2}{k - r_{n-1}^2}.$$

Thus we have established a recurrence relationship for r_n^2 in terms of r_{n-1}^2 , which on replacing p and k by their values $1/2$ and $1/h$ respectively, becomes

$$r_n^2 = \frac{2h-1}{2h} + \frac{(1-h)^2}{2h(1-hr_{n-1}^2)}. \quad (8)$$

We proceed to consider the limiting values of this series as n increases.

The maximum value of r_n^2 is 1 because r_n is a correlation; therefore if r_n^2 has a limit as n tends to infinity that is different from 1, r_L^2 say, it is, from (8), given by

$$r_L^2 = \frac{2h-1}{2h} + \frac{(1-h)^2}{2h(1-hr_L^2)}.$$

This is a quadratic in r_L^2 ,

$$2hr_L^4 - (2h+1)r_L^2 + h = 0,$$

having roots

$$r_L^2 = [(h+\frac{1}{2}) \pm \sqrt{(h+\frac{1}{2})^2 - 2h^2}] / 2h.$$

These tend to $\frac{1}{2}(1 \pm 1)$ as h tends to zero, and since r_n is zero for $h = 0$

$$r_L^2 = [h + \frac{1}{2} - \sqrt{(h + \frac{1}{2})^2 - 2h^2}] / 2h \quad (9)$$

and because r_n is always positive

$$r_L = +\sqrt{[h + \frac{1}{2} - \sqrt{(h + \frac{1}{2})^2 - 2h^2}] / 2h}. \quad (10)$$

Values of this expression are given in table 1 for h ranging from 0.1 to 1.0. Also shown are values of $r_1 = +\sqrt{\frac{1}{2}h}$, and

$$r_2 = +\sqrt{h(3-2h)/2(2-h^2)},$$

appropriate to basing the index respectively on parent records alone and on parent and grandparent records. These provide comparison

TABLE 1
Efficiency of single records of ancestors

Heritability (h)	Efficiency (r)		
	Upper limit *	Ancestor records	
		Parents †	Parents and grandparents ‡
0.1	0.29	0.22	0.27
0.2	0.39	0.32	0.37
0.3	0.45	0.39	0.43
0.4	0.50	0.45	0.49
0.5	0.54	0.50	0.53
0.6	0.57	0.55	0.57
0.7	0.61	0.59	0.61
0.8	0.64	0.63	0.64
0.9	0.67	0.67	0.67
1.0	0.71	0.71	0.71

$$* r_L^2 = \frac{1}{2h} [h + \frac{1}{2} - \sqrt{(h + \frac{1}{2})^2 - 2h^2}].$$

$$† r_1^2 = \frac{1}{2}h.$$

$$‡ r_2^2 = \frac{1}{2}h(3-2h)/(2-h^2).$$

of the efficiency of using just one or two generations of ancestors relative to the maximum efficiency possible from ancestors. We may note the following:

- (1) Maximum efficiency obtainable from ancestor records is $\frac{1}{2}\sqrt{2}$, which occurs when there is perfect heritability, *i.e.* $h = 1$. In this case parent records alone are sufficient for maximum efficiency.
- (2) For heritabilities 0.1 and 0.2 the efficiency of parent records is 80 per cent. of the maximum obtainable from ancestors in such cases and that of parent and grandparent records together

is approximately 90 per cent. of maximum; for heritability equal to 0.3, 0.4 and 0.5, parent records have efficiency 90 per cent. of maximum and when grandparent records are used as well this increases to approximately 97 per cent.

- (3) When heritability exceeds 0.5, parent records have almost maximum efficiency and parent and grandparent records used together do have, for practical purposes, maximum efficiency.

In general, therefore, if selection is being based on ancestor records, parental records provide maximum efficiency if the trait concerned has heritability greater than 0.5, and although grandparent records increase the efficiency when heritability is 0.5 or less, they do so to any appreciable extent only when heritability is less than 0.3. In any case, parent records on their own are never less efficient than 80 per cent. of maximum, provided heritability exceeds 0.1.

These results have been derived by assuming one record per ancestor, whereas in practice there are likely to be varying numbers of records available on different ancestors. All the possible variations cannot be considered here, but were there to be m records on each ancestor h would be replaced in the expression for r_n by $mh/[1 + (m-1)\rho]$ where ρ is repeatability. This simply increases the effective value of heritability in the calculations. For example if the heritability of a trait was 0.2 and $mh/[1 + (m-1)\rho]$ equalled 0.3 the relevant values in table 1 would be those corresponding to $h = 0.3$. Likewise, if ancestor records are not available but records on g of their progeny are (single records on g progeny of each ancestor), then in the expression for r_n , h is replaced by $gh/[4 + (g-1)h]$.

3. SIB RECORDS

Sometimes records may be available on sibs but not on ancestors, as for example on full sibs in poultry breeding or pig breeding, or on half-sibs on just one side of the pedigree, as frequently occurs in dairy cattle breeding. We will therefore discuss, relative to the efficiency of ancestor records, the efficiency of single records on N sibs, namely $\sqrt{Nh/s[s + (N-1)h]}$ where $s = 2$ for full sibs and $s = 4$ for half-sibs. Thus N sibs are more efficient than ancestor records when

$$\frac{Nh}{s[s + (N-1)h]} > r_L^2 \text{ i.e. } N > \frac{s(s/h - 1)}{1/r_L^2 - s}. \quad (11)$$

The latter expression becomes infinitely large for values of $1/r_L^2$ close to s and when $1/r_L^2$ exceeds s , is negative: but since (11) defines a lower limit on N , a number of animals, it must be finite and positive to be of use, and therefore since its numerator is positive, its denominator must be also. (This reasoning will be used repeatedly in what follows.) Thus (11) holds true for $r_L < \frac{1}{2}\sqrt{2}$ for full sibs and for $r_L < \frac{1}{2}$ for half-sibs, corresponding to heritabilities 1.0 and 0.4 respectively, from table 1. Hence full sibs are more efficient than pedigree evaluation for any situation other than one of perfect heritability ($h = 1$), the r_L -values

in table 1 applied to (11) with $s = 2$ indicating that 5 of them are more efficient for heritability 0.6 or less, whereas at least 6, 7 and 13 are needed for heritabilities of 0.7, 0.8 and 0.9; and half-sibs are more efficient than pedigree evaluation for heritabilities less than 0.4, at least 20, 29 and 54 being required for heritabilities of 0.1, 0.2 and 0.3 respectively.

Sib records can also be compared in this manner with parent or parent and grandparent records by replacing r_L^2 in (11) by $\frac{1}{2}h$ and $\frac{1}{2}h(3-2h)/(2-h^2)$ respectively. It may also be noted that for an infinitely large number of sibs the upper limit of efficiency for sib records is $1/\sqrt{s}$, namely $\frac{1}{2}\sqrt{2}$ for full sibs and $\frac{1}{2}$ for half-sibs.

4. COMPARISONS WITH PERFORMANCE AND PROGENY TESTING

As mentioned earlier the efficiency of performance-testing using a single record is $\sqrt{h}$. Since it can be shown from expression (9) that h exceeds r_L^2 , performance testing on even only a single record is more efficient than pedigree evaluation, regardless of the heritability value.

Progeny-testing using single records of g progeny has efficiency $\sqrt{gh/[4+(g-1)h]}$ which exceeds that of pedigree evaluation when

$$\frac{gh}{4+(g-1)h} > r_L^2, \text{ i.e. when } g > \frac{4/h-1}{1/r_L^2-1}.$$

From the r_L -values of table 1 it can be shown that so long as g exceeds 4 (for heritability not less than 0.1) this inequality holds true. Thus for heritability 0.1 and greater, a progeny-test on 4 or more progeny is more efficient than pedigree evaluation, no matter how many ancestors one uses in the latter. In fact for heritability exceeding 0.3 three progeny are sufficient. Hence pedigree evaluation as defined here is never as good as either performance testing, or progeny-testing with four or more progeny. Thus although pedigree evaluation is the only procedure available for estimating the genetic merit of an individual before records are available for performance-testing or progeny-testing, these latter procedures are more informative when they do become available. Nevertheless, Young (1961) has pointed out the value of combining parent records with performance and progeny test data; the latter are more efficient than the former but less efficient than using a combined index. For example, Young shows that for low heritability values two or more records on each parent in addition to the individual's own record, can increase the efficiency of the index by 30 per cent. or more.

Sib records are compared with performance-testing using a single record by replacing r_L^2 with h in (11). Thus N sibs are more efficient than performance-testing with one record when

$$N > \frac{s(s-h)}{1-sh}. \quad (12)$$

This holds true only when the denominator is positive. Hence sib records can be more efficient than performance testing only for heritability less than $1/s$, *i.e.* less than 0.5 in the case of full sibs and 0.25 for half-sibs, the minimum values of N for N sibs to be more efficient than performance testing with one record being given by (12); they

TABLE 2

Minimum number of sibs (with single records) that are more efficient than performance testing with 2, 3 or 4 records

Heritability (<i>h</i>)	No. of records, <i>n</i> , in performance test	Repeatability									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
		Full sibs									
0.1	2	11	10	9	8	7	7	6	6	6	
0.2		25	19	15	13	11	10	9	8	7	
0.3		...	...	69	35	23	18	14	12	10	
0.4		...	...	...	...	...	...	65	33	22	
0.1	3	20	15	12	10	9	8	7	6	6	
0.2		...	55	37	19	14	11	10	8	7	
0.3		...	...	...	...	52	26	18	13	11	
0.4		...	...	...	...	...	...	...	49	25	
0.1	4	31	20	14	11	9	8	7	7	6	
0.2		...	...	49	25	17	13	10	9	7	
0.3		...	...	...	...	137	35	20	14	11	
0.4		...	...	...	...	...	...	...	65	26	
		Half sibs									
0.1	2	145	109	87	73	62	55	49	44	40	
0.2		...	...	...	...	...	...	305	153	102	
0.1	3	...	324	217	109	82	65	55	47	41	
0.2		...	...	...	...	...	...	...	229	115	
0.1	4	...	...	289	145	97	77	58	49	42	
0.2		...	...	...	...	...	...	...	305	122	

are 5, 7, 9 and 17 full sibs for heritabilities 0.1, 0.2, 0.3 and 0.4 respectively, and 37 and 77 half-sibs for heritabilities 0.1 and 0.2. If sib records are to be compared with performance testing using n records the "1" in the denominator of (12) is replaced by $[1+(n-1)\rho]/n$, and the sibs are more efficient if $h < 1/s$ as above, provided that if $h > \rho/s$, $n < (1-\rho)/(sh-\rho)$. Table 2 shows minimum values of N in these situations, for $n = 2, 3$ and 4. When $h < 1/s$ but $h = \rho/s$, then N must exceed $sn(s-h)/(1-\rho)$ for sib records to be more efficient than performance testing with n records; and if $h < \rho/s$, N exceeding $s(s-h)/(\rho-sh)$ implies that sib records are more

efficient even for infinitely large values of n . Table 3 shows values of this last expression.

TABLE 3

Minimum number of sibs, with single records, that are more efficient than performance testing with any number of records

Heritability	Repeatability							
	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
	Full sibs							
	39	20	13	10	8	7	6	5
	...	...	37	19	13	10	8	7
	...	...	...	...	35	18	12	9
	...	...	...	...	...	...	33	17
	Half sibs							
	...	...	157	79	53	40	32	37
	...	...	...	...	...	...	153	77

Finally we can compare sib records with progeny-testing using n progeny with single records by replacing r_L^2 in (11) with

$$nh/[4 + (n-1)h].$$

Thus N sibs are more efficient when

$$N > \frac{sn(s-h)}{4 - [n(s-1) + 1]h}.$$

i.e. only when there are less than $(4/h-1)/(s-1)$ progeny. Thus 8 progeny are more efficient than full sib records, for heritabilities of 0.4 and greater, and 38, 18 and 12 progeny for heritabilities of 0.1, 0.2 and 0.3 respectively. Compared to half-sibs, 2 progeny are more efficient for heritabilities 0.4 and greater, and 12, 6, and 4 progeny are more efficient for heritabilities 0.1, 0.2 and 0.3 respectively.

5. SUMMARY

An expression is derived for the efficiency of a selection index based on ancestor records, measured as the correlation between index and true additive genetic merit, and maximum values obtained for an infinite number of generations. It is shown that parent records alone have an efficiency that is at least 80 per cent. of maximum, and parent and grandparent records together at least 90 per cent.— in both instances for heritabilities not less than 0.1. But progeny-tests on 4 or more progeny and performance tests on even a single record

are always more efficient than ancestor records. The efficiency of using sib records is also considered.

Acknowledgment.—The problem of the limiting efficiency of ancestor records was originally posed by Dr C. R. Henderson, Animal Husbandry Department, Cornell University.

6. REFERENCES

- COMSTOCK, R. E. 1948. Statistics in animal breeding research, Proc. Auburn Conf. on Statistics Applied to Research Stat. Lab., Auburn.
- CZEKANOWSKI, JAN. 1933. Mendelistisches *Laws of Ancestral Heredity*. *Ž. indukt. Abstamm-u. Vererb.*, 64, 154-168.
- HAZEL, L. N. 1943. Genetic basis for selection indices. *Genetics*, 28, 476-490.
- HENDERSON, C. R. 1961. Selection index and genetic advance. Symposium on Statistical Genetics and Plant Breeding, Raleigh, N.C. (in press).
- JARDINE, R. 1958. Animal breeding and the estimation of genetic value. *Heredity*, 12, 499-511.
- KEMPTHORNE, OSCAR. 1960. Biometrical relations between relatives and selection theory. *Biometrical Genetics*, ed. Oscar Kempthorne, Pergamon Press, 12-23.
- LUSH, J. L. 1947. *Animal Breeding Plans*. Iowa State College Press, Ames, Iowa.
- OSBORNE, R. 1957. The use of sire and dam averages in increasing the efficiency of selective breeding under a hierarchical mating system. *Heredity*, 11, 93-116.
- SEARLE, S. R. 1962. Bull-sampling programmes in the artificial breeding of dairy cattle. *Proc. N.Z. Soc. Ann. Prod* 22, 54-66.
- YOUNG, S. S. Y. 1961. The use of sire's and dam's records in animal selection. *Heredity*, 16, 91-102.