

# Prediction of rates of inbreeding in selected populations

NAOMI R. WRAY<sup>1</sup>† AND ROBIN THOMPSON<sup>2</sup>

<sup>1</sup> Institute of Animal Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN

<sup>2</sup> AFRC Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, Roslin, Midlothian, EH25 9PS

(Received 15 December 1988 and in revised form 18 July 1989)

## Summary

A method is presented for the prediction of rate of inbreeding for populations with discrete generations. The matrix of Wright's numerator relationships is partitioned into 'contribution' matrices which describe the contribution of the Mendelian sampling of genes of ancestors in a given generation to the relationship between individuals in later generations. These contributions stabilize with time and the value to which they stabilize is shown to be related to the asymptotic rate of inbreeding and therefore also the effective population size,  $N_e \approx 2N/(\mu_r^2 + \sigma_r^2)$ , where  $N$  is the number of individuals per generation and  $\mu_r$  and  $\sigma_r^2$  are the mean and variance of long-term relationships or long-term contributions. These stabilized values are then predicted using a recursive equation via the concept of selective advantage for populations with hierarchical mating structures undergoing mass selection. Account is taken of the change in genetic parameters as a consequence of selection and also the increasing 'competitiveness' of contemporaries as selection proceeds. Examples are given and predicted rates of inbreeding are compared to those calculated in simulations. For populations of 20 males and 20, 40, 100 or 200 females the rate of inbreeding was found to increase by as much as 75% over the rate of inbreeding in an unselected population depending on mating ratio, selection intensity and heritability of the selected trait. The prediction presented here estimated the rate of inbreeding usually within 5% of that calculated from simulation.

## 1. Introduction

In a random mating population and in the absence of differences in viability and fecundity, all families have equal probabilities of contributing offspring to be parents of the next generation. In a population undergoing selection, families superior for the selected trait will contribute more offspring to the next generation than inferior families. Therefore, the rate of inbreeding is higher in selected populations than in randomly mated populations. In a given generation,  $t$ , the mean level of inbreeding,  $\mathcal{F}_t$ , and rates of inbreeding,  $\Delta\mathcal{F}$ , (defined as  $\Delta\mathcal{F} = (\mathcal{F}_t - \mathcal{F}_{t-1})/(1 - \mathcal{F}_{t-1})$ ) can easily be calculated from pedigree information after selection has occurred, but prediction of inbreeding rate in the planning stage of a breeding programme has proved to be difficult. Frequently, advantages of new breeding schemes are discussed solely in terms of responses to selection with little regard to the effect of selection on inbreeding, with

inbreeding rate per generation assumed to be simply that of Wright (1931) ( $\Delta\mathcal{F}_w$ ),

$$\Delta\mathcal{F}_w = \frac{1}{8M} + \frac{1}{8F},$$

where  $M(F)$  are the number of males (females) entering the population each generation. This rate of inbreeding is only appropriate for randomly mated populations with Poisson distributed variances of family size.

Inbreeding is often considered in terms of the effective population size, which has an inverse relationship with rate of inbreeding,  $\Delta\mathcal{F} = 1/2N_e$ . Effective population size is a subject which has received much attention in the study of natural populations (see Crow & Denniston, 1988 for a review) but selection is assumed to be absent. Where selection has been considered in natural populations, the selection criterion is usually assumed to be non-genetically determined, that is, the selective advantage of an individual is not inherited. In this way natural selection is included via differential variance in family size (Latter, 1959; Hill, 1972; Hill, 1979). Nei & Murata

† Present address: Edinburgh School of Agriculture, West Mains Road, Edinburgh, EH9 3JG.

(1966) considered effective population size in natural populations when fertility is inherited, but their derivation is similar to that of Robertson (1961) discussed below.

Robertson (1961) first considered prediction of inbreeding in selected populations. He considered a population consisting of  $N$  full-sib families, that is a total of  $2N$  parents. Robertson's argument was based on the variance of the change in gene frequency and suggests a relationship between  $N/N_e$ ,  $Q$ ,  $i$  and  $\rho$  (where  $Q$  is a measure of relative selective advantage,  $i$  is the selection intensity and  $\rho$  is the intra-class correlation of selection criteria between sibs) of the form,

$$\frac{N}{N_e} = 1 + Q^2 i^2 \rho. \quad (1)$$

We are unable to follow his derivation, since  $N$  must now be re-interpreted because under no selection, when  $\rho$  is zero,  $N_e = N$  and from Wright's formula,  $N_e$  is the total number of parents. In equation (1),  $i^2 \rho$  is the increment in the ratio of  $N/N_e$  following one generation of selection and accounts for the effect of the differential selective advantage between parents of different genetic merit for the selected trait on the variance of family size contributing to the next generation.  $Q$  accounts for the way in which the selective advantage accumulates with subsequent generations of selection expressed relative to the selective advantage in the initial generation,  $i^2 \rho$ . Following a single round of selection  $Q = 1$  by definition, and Robertson argued that  $Q = \frac{3}{2}$  following two generations of selection and then increases as the sum of a geometric series tending to a limiting value of 2 when the number of generations of selection is large. The increase in relative selective advantage affects the variance of the long-term family size and so the square of  $Q$  is used, and equation (1) with  $Q^2$  as 4 is Robertson's expression for the steady state rate of inbreeding in a selected population.

The intra-class correlation of full-sibs is assumed to be  $\rho = h^2/2$  where  $h^2$  is the heritability of the trait under selection. The parents, themselves, are likely to be a selected group and so to account for the effect of reduction in variance Robertson used  $\rho = (h^2/2)(1 - kh^2)$  where  $k = i(i - v)$  and  $v$  is the truncation deviate. More correctly it should be  $\rho = (h^2/2)(1 - kh^2)/(1 - kh^4/2)$  (Bulmer, 1971).

In practice, equal numbers of offspring from each family ( $n$  of each sex) are likely to be measured prior to selection, which is likely to induce a larger effective population size than if the family sizes are Poisson distributed (as assumed in Robertson's derivation). Robertson (1961) corrected for this using a formula of Crow & Morton (1955),

$$\frac{N}{N_e} = \left(1 - \frac{1}{2n}\right) + Q^2 i^2 \rho \left(1 - \frac{1}{n}\right). \quad (2)$$

Experimental checks of the predictions presented in equations (1) and (2) have been few, probably because the prediction only allows for a limited family structure – only full-sib or only half-sib families. Long-term experiments with *Drosophila* (Jones, 1969; Yoo, 1980) have shown the prediction to be reasonable, although it is noted that the traits under selection were of low heritability (0.16 and 0.05 respectively). Simulation studies (e.g. Hill, 1985) have found the prediction to severely overestimate inbreeding at moderate heritabilities and high selection intensities.

Burrows (1984a) estimated probabilities of pairwise coancestry (i.e. probabilities of two selected individuals being sibs) following a single generation of selection using the same full-sib family structure as Robertson (1961). He derived a prediction of inbreeding rate for which equation (1) with  $Q = 1$  is only a first order approximation, the higher order terms becoming important at high selection intensities and high correlations between selection criteria of sibs. The prediction was not extended to subsequent generations and so the prediction is not for asymptotic rates of inbreeding. Different family structures were considered (Burrows, 1984b), for example, the hierarchical structure commonly used in animal breeding programmes. However, the candidates available for selection are envisaged as a mixed group of males and females, a scenario which is useful in plant breeding since a selected plant can take on the role of either male or female parent but which is not applicable to animal populations. By not separating the sexes there is no allowance for the differential selection intensities which are an essential feature of an animal hierarchical breeding programme. Differential selection criteria for males and females also cannot be accommodated. In addition, Woolliams (1989) found the estimates of frequencies of co-selection of sibs derived from the formulae of Burrows (1984b) to be gross over-estimates, in some cases generating frequencies substantially greater than 1.

This paper presents a method of predicting asymptotic rates of inbreeding in closed populations undergoing selection, which improves on equation (1) for populations of full-sib families and which can be extended to populations with a hierarchical mating structure with differing numbers of males and females. A list of notation is given in Appendix I.

## 2. Model, assumptions and simulations

The trait under selection is assumed to be under the control of many genes each of small additive genetic effect, the infinitesimal model (Bulmer, 1980). Breeding values of individuals in the base population, prior to selection, are assumed to be distributed normally with mean zero and variance  $\sigma_A^2$ . In subsequent generations the breeding value ( $A'$ ) of an individual  $i$  is assumed to be,

$$A'_i = \frac{1}{2}A'_s + \frac{1}{2}A'_d + \phi_i, \quad (3)$$

where  $A'_s$  and  $A'_d$  are the breeding values of the sire and dam respectively and  $\phi_i$  is the Mendelian sampling effect unique to individual  $i$ .  $\phi_i$  is distributed normally with mean zero and variance

$$\frac{1}{2} \left( 1 - \frac{f_s + f_d}{2} \right) \sigma_A^2,$$

where  $f_s$  and  $f_d$  are the inbreeding coefficients of the sire and dam. Generations are assumed to be discrete with the selected population consisting of  $M$  male and  $F$  female parents each generation. Populations are assumed to be large enough such that  $\Delta\mathcal{F}$  is approximately constant each generation. Truncation selection is assumed, whereby the individuals with the highest scores under the selection criteria are selected, and selected individuals are mated at random.

Rates of inbreeding are assumed to be predicted for loci whose effect is neutral with respect to the selected trait. Under an infinitesimal model, however, rates of inbreeding are unlikely to be substantially different for loci affecting and not affecting the selected trait.

In section 3 the relationship between  $\Delta\mathcal{F}$  and long-term contributions or relationships is demonstrated. In section 4 the prediction of long-term relationships is given for a population with a hierarchical mating structure ( $F \geq M$ ) with  $F/M$  being an integer. Each female and male parent is assumed to have  $n_f$  and  $n_m$  offspring respectively of each sex available for selection, where  $n_m = (F/M)n_f$ . The selection criterion is assumed to be individual phenotype (mass selection).

Checks are made on the derived methodology by comparing predicted results with those calculated from simulation. Simulated populations have the same population, mating and selection structure as described in the assumptions above. Individual breeding values are simulated by sampling from a normal distribution as described in equation (3). Phenotypes are simulated as the sum of breeding values and environmental or error effects. The latter are sampled from a normal distribution with mean zero and variance  $\sigma_E^2$ . The total phenotypic variance,  $\sigma_P^2$  is defined as  $\sigma_P^2 = \sigma_A^2 + \sigma_E^2 = 100$  units. Traits with different heritabilities are considered,  $h^2 = \sigma_A^2/\sigma_P^2 = 10^{-6}, 0.1, 0.2, 0.4, 0.6$ . Different population structures are considered,  $M = 20, F = 20, 40, 100, 200$ . Simulation results are the average of 100 replicates and selection is continued for up to 20 generations. Asymptotic rates of inbreeding are calculated from simulation as,

$$\Delta\mathcal{F}_{stm} = \frac{1}{10} \sum_{t=5}^{14} \frac{\mathcal{F}_t - \mathcal{F}_{t-1}}{1 - \mathcal{F}_{t-1}},$$

where  $\mathcal{F}_t$  is the level of inbreeding in generation  $t$  averaged over 100 replicates.

### 3. Relating the rate of inbreeding to long-term contributions

The matrix of additive genetic relationships between individuals and how its components relate to  $\Delta\mathcal{F}$  is considered initially.

#### (i) Partitioning the relationship matrix

Let  $A$  be the symmetric matrix of the numerator of Wright's coefficient of additive genetic relationship (Wright, 1922) such that the  $(p, q)$  element  $a_{pq}$  represents the numerator relationship between individuals  $p$  and  $q$ . Consider individuals  $p$  (with parents  $w$  and  $x$ ) and  $q$  (with parents  $y$  and  $z$ ) then when  $p$  is older than  $q$ ,

$$a_{pq} = (a_{wq} + a_{xq})/2.$$

If  $p$  and  $q$  are in the same generation then,

$$a_{pq} = (a_{wy} + a_{wz} + a_{xy} + a_{xz})/4 \tag{4}$$

and the numerator relationship of individual  $p$  with itself is,

$$a_{pp} = 1 + f_p = 1 + a_{wx}/2 = (a_{ww} + a_{wx} + a_{xw} + a_{xx})/4 + (1 - a_{ww}/4 - a_{xx}/4), \tag{5}$$

(Thompson, 1977). In (5)  $a_{pp}$  has been partitioned into two components, the first is analogous to the relationship between two different individuals as in (4) and the second is related to the Mendelian sampling of genes with a correction for the way in which the mean level of inbreeding of the parents reduces it.

Consider individuals from many generations, with  $M+F$  individuals per generation. Let  $A_{ij}$  be the relationship matrix between individuals of generations  $i$  and  $j$  of order  $(M+F) \times (M+F)$  so that  $A$  is composed of sub-matrices,

$$A = \begin{bmatrix} A_{00} & A_{01} & A_{02} \dots A_{0t} \\ A_{10} & A_{11} & A_{12} \dots A_{1t} \\ A_{20} & A_{21} & A_{22} \dots A_{2t} \\ \vdots & \vdots & \vdots & \vdots \\ A_{t0} & A_{t1} & A_{t2} & A_{tt} \end{bmatrix}.$$

Therefore, if the relationships between individuals of the  $i$ th and  $k$ th generations are known, relationships between the offspring of the  $i$ th generation (i.e. the  $(i+1)$ th generation) and the  $k$ th generation can be derived. Let,

$$A_{(i+1)k} = Z_i A_{ik} \text{ for } k \leq i,$$

where  $Z_i$  is a matrix of order  $(M+F) \times (M+F)$  which relates individuals of generation  $(i+1)$  to generation  $i$  and whose elements are either 0 or  $\frac{1}{2}$ . Element  $(p, q)$  of  $Z_i$  is  $\frac{1}{2}$  if the  $q$ th individual of generation  $i$  is a parent of the  $p$ th individual in generation  $(i+1)$ .

The relationship matrix between individuals in the  $(i+1)$ th generation,  $A_{(i+1)(i+1)}$ , can be written in a similar way,

$$A_{(i+1)(i+1)} = Z_i A_{ii} Z_i' + D_{(i+1)},$$

(Thompson, 1977) where  $D_{(i+1)}$  is a diagonal matrix

containing terms analogous to the term  $(1 - a_{ww}/4 - a_{xx}/4)$  from (5), the terms of which are  $\frac{1}{2}$  if the parents are non-inbred and which tend to 0 as the parents become more inbred. It follows that the matrix **A** can be written as ( $t = 2$ ),

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_0 & \mathbf{I}_0 \mathbf{Z}'_0 & \mathbf{I}_0 \mathbf{Z}'_0 \mathbf{Z}'_1 \\ \mathbf{Z}_0 \mathbf{I}_0 & \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 + \mathbf{D}_1 & \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 \mathbf{Z}'_1 + \mathbf{D}_1 \mathbf{Z}'_1 \\ \mathbf{Z}_1 \mathbf{Z}_0 \mathbf{I}_0 & \mathbf{Z}_1 \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 + \mathbf{Z}_1 \mathbf{D}_1 & \mathbf{Z}_1 \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 \mathbf{Z}'_1 + \mathbf{Z}_1 \mathbf{D}_1 \mathbf{Z}'_1 + \mathbf{D}_2 \end{bmatrix},$$

where, for example,  $\mathbf{Z}'_1 \mathbf{Z}'_2 \dots \mathbf{Z}'_t$  is a square matrix of order  $(M + F)$  which relates ancestors of generation 1 to descendants of generation  $t$ . The  $p$ th column of  $\mathbf{Z}'_1 \mathbf{Z}'_2 \dots \mathbf{Z}'_t$  gives the proportion of genes of the  $p$ th individual in generation  $t$  derived from the ancestors in generation 1. The sum of the  $M + F$  columns,  $\mathbf{Z}'_1 \mathbf{Z}'_2 \dots \mathbf{Z}'_t \mathbf{1} = \mathbf{r}_t$  (where  $\mathbf{1}$  is the unit vector) gives the number of descendants arising from the  $M + F$  individuals in generation 1 weighted according to their genetic contribution to the individuals in generation  $t$ . Each column of  $\mathbf{Z}'_1 \mathbf{Z}'_2 \dots \mathbf{Z}'_t$  is the average of two columns of  $\mathbf{Z}'_1 \mathbf{Z}'_2 \dots \mathbf{Z}'_{t-1}$ , so as  $t$  increases the variance of the column vectors is reduced and each column becomes closer to the mean vector  $\mathbf{r}_t / (M + F)$ , and

$$\prod_{j=1}^t \mathbf{Z}'_j = \prod_{j=1}^{t+1} \mathbf{Z}'_j \quad \text{as } t \rightarrow \infty.$$

In addition, as  $t$  increases, the sum of the columns will tend to the same vector,  $\mathbf{r}$ , and  $\mathbf{Z}'_1 \mathbf{Z}'_2 \dots \mathbf{Z}'_t$  will tend to  $\mathbf{r}_t \mathbf{1} / (M + F)$  to a limit of  $\mathbf{r} \mathbf{1} / (M + F)$ . An alternative argument for a limiting value of  $\mathbf{Z}'_1 \mathbf{Z}'_2 \dots \mathbf{Z}'_t$  is found from standard matrix theory by noting that the  $\mathbf{Z}_t$  are stochastic matrices with a single largest eigenvalue of 1.

In generation  $t$ , **A** can be considered as a sum of  $(t + 1)$  'contribution' matrices (**C**) such that  $\mathbf{A} = \sum_{k=0}^t \mathbf{C}_k$  is the matrix initiated in generation  $k$  and contains the contributions to the numerator relationship matrix from the Mendelian sampling of genes which took place in generation  $k$ .  $\mathbf{C}_k$  is composed of submatrices  $\{\mathbf{C}_{k(i,j)}\}$  relating generations  $i$  and  $j$  by the contributions from  $k$ . For example when  $t = 2$ ,  $\mathbf{A} = \mathbf{C}_0 + \mathbf{C}_1 + \mathbf{C}_2$ , where,

$$\mathbf{C}_0 = \begin{bmatrix} \mathbf{I}_0 & \mathbf{I}_0 \mathbf{Z}'_0 & \mathbf{I}_0 \mathbf{Z}'_0 \mathbf{Z}'_1 \\ \mathbf{Z}_0 \mathbf{I}_0 & \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 & \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 \mathbf{Z}'_1 \\ \mathbf{Z}_1 \mathbf{Z}_0 \mathbf{I}_0 & \mathbf{Z}_1 \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 & \mathbf{Z}_1 \mathbf{Z}_0 \mathbf{I}_0 \mathbf{Z}'_0 \mathbf{Z}'_1 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{0(0,0)} & \mathbf{C}_{0(0,1)} & \mathbf{C}_{0(0,2)} \\ \mathbf{C}_{0(1,0)} & \mathbf{C}_{0(1,1)} & \mathbf{C}_{0(1,2)} \\ \mathbf{C}_{0(2,0)} & \mathbf{C}_{0(2,1)} & \mathbf{C}_{0(2,2)} \end{bmatrix}$$

$$\mathbf{C}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{D}_1 & \mathbf{D}_1 \mathbf{Z}'_1 \\ 0 & \mathbf{Z}_1 \mathbf{D}_1 & \mathbf{Z}_1 \mathbf{D}_1 \mathbf{Z}'_1 \end{bmatrix} \quad \mathbf{C}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbf{D}_2 \end{bmatrix}.$$

These contribution matrices can be calculated in a similar way to the numerator relationship matrix.

$\mathbf{C}_{k(i,j)}$  are null sub-matrices when  $i$  or  $j < k$ .  $\mathbf{C}_k$  has its first non-null sub-matrix at  $\mathbf{C}_{k(k,k)}$  and this is simply a diagonal matrix of Mendelian sampling contributions to individuals in generation  $k$ .

(ii) *Relating the partitioned relationship matrix to  $\Delta \mathcal{F}$*

The mean of the diagonals of  $\mathbf{A}_t$  (denoted  $\overline{\text{diag}}\{\mathbf{A}_t\}$ ) is 1 plus the mean inbreeding coefficient of the selected individuals in generation  $t$  ( $\mathcal{F}_t$ ), that is,  $\overline{\text{diag}}\{\mathbf{A}_t\} = 1 + \mathcal{F}_t$ , and therefore  $\overline{\text{diag}}\{\mathbf{A}_t\} = \sum_{k=0}^t \overline{\text{diag}}\{\mathbf{C}_{k(t,t)}\} = 1 + \mathcal{F}_t$ . With discrete generations  $\overline{\text{diag}}\{\mathbf{C}_{0(0,0)}\} = 1$ ,  $\overline{\text{diag}}\{\mathbf{C}_{0(1,1)}\} = \overline{\text{diag}}\{\mathbf{C}_{1(1,1)}\} = \frac{1}{2}$  and  $\overline{\text{diag}}\{\mathbf{C}_{t(t,t)}\} \approx (1 - \mathcal{F}_{t-1})/2$ . The latter is only approximate because it depends on the exact contribution from each parent after selection has taken place. Following from this,

$$\frac{\overline{\text{diag}}\{\mathbf{C}_{t+1(t+1,t+1)}\}}{\overline{\text{diag}}\{\mathbf{C}_{t(t,t)}\}} \approx \frac{(1 - \mathcal{F}_t)}{(1 - \mathcal{F}_{t-1})}, \quad \text{and analogously,}$$

$$\frac{\overline{\text{diag}}\{\mathbf{C}_{k+1(t+1,t+1)}\}}{\overline{\text{diag}}\{\mathbf{C}_{k(t,t)}\}} \approx \frac{(1 - \mathcal{F}_t)}{(1 - \mathcal{F}_{t-1})} \quad \text{for } 0 < k \leq t. \quad (6)$$

In addition, as  $t \rightarrow \infty$  then  $\overline{\text{diag}}\{\mathbf{C}_{k(t,t)}\} \rightarrow \overline{\text{diag}}\{\mathbf{C}_{k(t+1,t+1)}\}$ . This can be derived from the stochastic nature of the **Z** matrices so that when this asymptote is reached, all diagonal elements, and indeed all off-diagonal elements of  $\mathbf{C}_{k(t,t)}$  are equal. The interpretation is that the Mendelian sampling of selected genes in the  $k$ th generation contributes equally to all animals by the  $t$ th generation, where  $t$  is sufficiently large to have allowed stabilization of the dispersal of genes throughout the population. It is this dispersal of genes that James and McBride (1958) examined in retrospect in data collected from selected poultry populations.

Using the identities presented above, a relationship between the stabilised contributions of individuals of the first generation and  $\Delta \mathcal{F}$  can be found, the details of which are presented in Appendix II,

$$\Delta \mathcal{F} \approx \overline{\text{diag}}\{\mathbf{C}_{1(t,t)}\} / 2. \quad (7)$$

Table 1 presents  $\overline{\text{diag}}\{\mathbf{C}_{k(t,t)}\}$  for  $k = 0, 4$  and  $t = 0, 20$  observed from simulation (mean of 100 replicates) of a population of  $M = 20$  males and  $F = 40$  females with  $n_f = 3$  undergoing mass selection on a trait with heritability 0.4. The calculated mean  $\Delta \mathcal{F}$  from simulation is 0.0118 and  $\Delta \mathcal{F}$  using (7) is 0.0110.

As  $t \rightarrow \infty$ , all elements of  $\mathbf{C}_{1(t,t)}$  are equal so that the mean of all elements is the same as the mean of the diagonals,

$$\Delta \mathcal{F} \approx \frac{1}{2(M + F)^2} \mathbf{1}' \mathbf{C}_{1(t,t)} \mathbf{1} = \frac{1}{2(M + F)^2} \mathbf{1}' \mathbf{C}_{1(t,1)} \mathbf{D}_1^{-1} \mathbf{C}_{1(1,t)} \mathbf{1}.$$

Table 1. The mean of the diagonal elements of the first five contribution matrices for a population of  $M = 20$  and  $F = 40$  with  $n_f = 3$  undergoing mass selection on a trait with  $h^2 = 0.4$

$t$	diag $\{C_{0(t,t)}\}$	diag $\{C_{1(t,t)}\}$	diag $\{C_{2(t,t)}\}$	diag $\{C_{3(t,t)}\}$	diag $\{C_{4(t,t)}\}$	$1 + F_t$
0	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000
1	0.5000	0.5000	0.0000	0.0000	0.0000	1.0000
2	0.2653	0.2500	0.5000	0.0000	0.0000	1.0153
3	0.1546	0.1310	0.2500	0.4919	0.0000	1.0275
4	0.1009	0.0748	0.1311	0.2459	0.4863	1.0390
5	0.0734	0.0470	0.0737	0.1282	0.2432	1.0469
⋮						
10	0.0490	0.0226	0.0230	0.0244	0.0277	1.1023
15	0.0485	0.0221	0.0218	0.0219	0.0226	1.1536
20	0.0485	0.0221	0.0218	0.0219	0.0225	1.2032

Since, firstly,  $D_1$  is a diagonal matrix with each diagonal element being  $\frac{1}{2}$ , secondly,  $C_{1(t,t)} \mathbf{1} = D_1 Z'_1 Z'_2 \dots Z'_t \mathbf{1} = \frac{1}{2} \mathbf{r}_t$ , and thirdly, when  $t$  is sufficiently large that  $\mathbf{r}_t = \mathbf{r}$ , it follows that,

$$\Delta \mathcal{F} \approx \frac{1}{4(M+F)^2} \mathbf{r}' \mathbf{r} = \frac{1}{4(M+F)^2} \sum_{i=1}^{M+F} r_i^2, \tag{8}$$

where  $r_i$  is the  $i$ th element of  $\mathbf{r}$  and represents the total long-term contribution from the  $i$ th ancestor to descendants, or equivalently the total long-term additive relationship between ancestor and descendants.

**4. Prediction of long-term contributions**

In the previous section we demonstrated how rate of inbreeding may be related to the sum of squares of long-term contributions or relationships ( $\sum r_i^2$ ) of first generation ancestors to descendants. The objective of this section is to predict this quantity. The assumption of a hierarchical mating scheme is now imposed. In this case, males and females must be considered separately due to their differential selection intensities. Long-term contributions are then considered separately from male and female ancestors to male and female descendants. Initially, the prediction of the mean of the long-term contribution is considered. Subsequently, the accumulation of the total sums of squares over several generations of selection is examined, in which the total sums of squares in a given generation depends on the selection in all previous generations.

(i) Notation

Subscripts  $m$  or  $f$  relate to the sex, male or female, of individuals in a given generation for which there are  $M$  males and  $F$  females per generation as before. For generality,  $w, x$  and  $y$  may be used to allow for any combination of male and female subscripts; correspondingly  $W, X$  and  $Y$  denotes the numbers of individuals of that sex in each generation. Subscripts

of the form  $xy, t$  refer to descendants of sex  $y$  in the  $t$ th generation from ancestors (in the first generation) of sex  $x$ , and those in the form  $xwy, t$  refer to descendants of sex  $y$  in the  $t$ th generation from families of descendants of sex  $w$  in generation  $t-1$  from ancestors of sex  $x$ . The subscript  $xwy^*, t$  is as for  $xwy, t$  but denotes a conditioning on selection of individuals in generation 1 up to  $t-1$ , accounting for the sexes represented by the first two subscripts. A parent of sex  $x$  has  $n_x$  offspring of each sex, with  $n_m = (F/M)n_f$ .

Truncation selection is assumed and so the notation  $p_x, i_x, v_x, z_x$  and  $k_x$  is used to denote the proportion selected, the selection intensity, the normal deviate, the normal ordinate and the variance reduction constant  $k_x = i_x(i_x - v_x)$  associated with the selection of individuals of sex  $x$ . In addition,  $k = (k_m + k_f)/2$ .

The phenotypic and additive genetic variances and the heritability in the base generation before selection are  $\sigma_p^2, \sigma_A^2$  and  $h^2$  respectively. The offspring of the base generation are denoted generation 1. The term ancestors will be used to refer to the selected individuals of generation 1 and the genetic variance of the ancestors of sex  $x$  is,

$$V_x = (1 - k_x h^2) \sigma_A^2. \tag{9}$$

$h_t^2$  is the heritability appropriate for individuals in generation  $t$  (prior to selection) which accounts for the reduction in genetic variance due to previous generations of selection (Bulmer, 1971), e.g.  $h_1^2 = h^2$  and  $h_2^2 = h^2(1 - \frac{1}{2}kh^2)/(1 - \frac{1}{2}kh^4)$ .

$r_{t(xy),t}$  is the sum of the contributions from the  $i$ th ancestor of sex  $x$  in the first generation to descendants of sex  $y$  in generation  $t$ .

$b_{xy,t}$  is used as a regression of the selection score of descendants of sex  $y$  in generation  $t$  on the breeding value deviation ( $A_i$ ) of its ancestor of sex  $x$ . The breeding value deviation of the ancestor is defined as  $A_i = A'_i - A$ , where  $A'_i$  is the actual breeding value of the ancestor and  $A$  is the mean breeding value of all the ancestors, such that,  $E(A_i) = 0$  and  $\text{Var}(A_i) = V_x$ . Breeding value deviation will hereafter be called breeding value.

Where  $i$  appears as a subscript in terms such as  $\mu_{i(xy),t}$ ,  $\sigma_{i(xy),t}^2$ ,  $SS_{i(xy),t}$ ,  $CP_{i(x),t}$  etc. which are introduced in the text, a conditioning on the breeding value of the ancestor  $i$  of sex  $x$  is implied. When the subscript is omitted then an expectation has been taken over all ancestors e.g.  $SS_{xy,t} = E[SS_{i(xy),t}]$ .

$\beta_{P,Q}$  is the regression of  $P$  on  $Q$ , where  $P$  and  $Q$  can be any random variables.

(ii) *Prediction of the mean of long-term relationships*

In this section, the mean (over all ancestors) of the sum of long-term genetic contributions from ancestors (with known breeding value) to descendants is predicted considering males and females separately, i.e.  $E[r_{i(xy),t}]$  is predicted. Under no selection each ancestor of sex  $x$  is expected to contribute  $2^{t-2}(Y/X)$  descendants of sex  $y$  in the  $t$ th generation. For example, a male ancestor is expected to have one son, two grandsons (one via his son and an expected  $M/F$  grandsons from each of his  $F/M$  daughters), four great grandsons, etc. The relationship between ancestor and descendant along a single pathway is  $1/2^{t-1}$ , the total contribution is, therefore,  $(1/2)(Y/X)$ . Under selection some ancestors are expected to contribute more descendants than others; that is they have a selective advantage which is a function of the superiority of their breeding value over the breeding value of their contemporaries,  $A_i$ . It follows that,

$$E[r_{i(xy),t}] = 2^{t-2} \left[ \frac{Y}{X} + b_{xy,t} A_i \right] \left( \frac{1}{2} \right)^{t-1} = \frac{1}{2} \left[ \frac{Y}{X} + b_{xy,t} A_i \right] = \mu_{i(xy),t}$$

where  $2^{t-2}b_{xy,t}$  can be interpreted as the regression coefficient of number of descendants of sex  $y$  in generation  $t$  on the breeding value of their ancestors of sex  $x$  and where  $E[r_{i(xy),t}]$  is strictly an expectation conditional on  $A_i$ . (The assumption of a linear regression was checked by simulation and quadratic effects were found to be non-significant for the example populations.) In generation  $t$ , descendants can be chosen either from male or female descendants in generation  $t-1$ , that is,

$$\left[ \frac{Y}{X} + b_{xy,t} A_i \right] = \frac{1}{2} \left[ \frac{Y}{X} + b_{xmy,t} A_i \right] + \frac{1}{2} \left[ \frac{Y}{X} + b_{xfy,t} A_i \right]. \tag{10}$$

The coefficients of  $\frac{1}{2}$  are part of the accumulation of  $(\frac{1}{2})^{t-1}$ . The number of offspring selected from male parent families is the product of the number of male families  $[M/X + b_{xm,(t-1)} A_i]$  with the number of descendants per male family  $[Y/M + b_{xmy*,t} A_i]$ .  $Y/M$  is the expected number of descendants of sex  $y$  from each male parent family and the  $b_{xmy*,t}$  is the regression coefficient which accounts for the additional selective advantage of the ancestor to the descendant given the selective advantage of the ancestor to the male parent. Superior ancestors are expected to contribute more

male descendants than an average ancestor. In addition, the male descendants of superior ancestors are themselves expected to contribute more descendants due to the superior genes they have inherited. It follows that,

$$\left[ \frac{Y}{X} + b_{xy,t} A_i \right] = \frac{1}{2} \left[ \frac{M}{X} + b_{xm,(t-1)} A_i \right] \left[ \frac{Y}{M} + b_{xmy*,t} A_i \right] + \frac{1}{2} \left[ \frac{F}{X} + b_{xf,(t-1)} A_i \right] \left[ \frac{Y}{F} + b_{xfy*,t} A_i \right] \tag{11}$$

which can be written equivalently as,

$$\mu_{i(xy),t} = \mu_{i(xm),t-1} \mu_{i(xmy*),t} + \mu_{i(xf),t-1} \mu_{i(xfy*),t} \tag{12}$$

where

$$\mu_{i(xwy*),t} = \frac{1}{2} \left[ \frac{Y}{F} + b_{xfy*,t} A_i \right].$$

In addition,

$$b_{xy,t} = \frac{1}{2} \left[ \frac{Y}{M} b_{xm,(t-1)} + \frac{Y}{F} b_{xf,(t-1)} + \frac{M}{X} b_{xmy*,t} + \frac{F}{X} b_{xfy*,t} \right]. \tag{13}$$

Therefore, the prediction of  $b_{xy,t}$  depends on the regression coefficients of previous generations.

(a) *Prediction of regression coefficients of selection score of offspring (generation 2) on breeding value of their ancestors in generation 1*

The selection score of an offspring (generation 2),  $S_2$ , can take the value  $S_2 = 1$  if the offspring is selected, and  $S_2 = 0$  if it is not, so that  $S_2$  is an all-or-none variable.  $b_{xy,2}$  is the regression coefficient of total selection score of offspring of sex  $y$  on the breeding value of its parent (the ancestor in generation 1) of sex  $x$ . Let  $\beta_{S_2,A^*}$  be the same regression coefficient but for a single offspring and where the \* represents that the ancestors were selected on the basis of their phenotype, so that the regression is conditioned on this selection. The regression of selection score on  $A^*$  can be completely explained through the relationship of  $S_2$  with the phenotype of the offspring ( $P_2$ ) on which the selection decisions are based. Therefore it follows that,

$$b_{xy,2} = n_x \beta_{S_2,A^*} = n_x \beta_{S_2,P_2^*} \beta_{P_2,A^*}, \tag{14}$$

where  $\beta_{S_2,P_2^*}$  is the regression coefficient of offspring selection score on offspring phenotype given that the ancestors are a selected group. The regression coefficient  $\beta_{P_2,A^*} = \frac{1}{2}$ , regardless of the conditioning, and,

$$\beta_{S_2,P_2^*} = \frac{i_y P_y \sigma_{P_2^*}}{\sigma_{P_2^*}^2} = \frac{z_y}{\sigma_{P_2^*}} \tag{15}$$

(Robertson, appendix in Dempster and Lerner, 1950), where  $\sigma_{P_2^*}^2$  is the phenotypic variance of the offspring

given that their parents of sex  $x$  were selected on their phenotype,

$$\sigma_{P_2^*}^2 = \sigma_P^2(1 - \frac{1}{4}h^2k_x),$$

and finally,

$$b_{xy,2} = \frac{1}{2} n_x \frac{z_y}{\sigma_{P_2^*}}.$$

(b) Prediction of regression of selection score of descendants (generation  $t$ ) on the breeding value of their ancestors in generation 1

In order to predict  $b_{xy,3}$ , from (11),  $b_{xmy^*,3}$  and  $b_{xfy^*,3}$  are needed, where  $b_{xmy^*,3}$  is then the regression of selection score of grandoffspring ( $S_3$ ) of sex  $y$  on the breeding value of their paternal grandparent (ancestor) of sex  $x$  given that both the grandparent and parent were selected on the basis of their phenotype. Using similar notation as in (14), but replacing subscript 2 with 3 and the \* now representing conditioning on two generations of selection, it follows that,

$$b_{xmy^*,3} = n_m \beta_{S_3, A^*} = n_m \beta_{S_t, P_t^*} \beta_{P_t, A^*},$$

and analogous to (15),

$$\beta_{S_3, P_3^*} = \frac{z_y}{\sigma_{P_3^*}}.$$

In general, the regression of number of descendants of sex  $y$  in generation  $t$  on breeding value of ancestors of sex  $x$  in generation 1, via descendants of sex  $w$  in generation  $t-1$ , given that all generations from 1 to  $t-1$  have been selected on the basis of their phenotype (represented by \*),  $b_{xwy^*,t}$  can be written as,

$$b_{xwy^*,t} = n_w \beta_{S_t, A^*} = n_w \beta_{S_t, P_t^*} \beta_{P_t, A^*},$$

where the notation is as defined in (14), with subscript 2 replaced by  $t$ . Finally, analogous to (15) is,

$$\beta_{S_t, P_t^*} = \frac{z_y}{\sigma_{P_t^*}}.$$

In order to estimate the general  $\beta_{P_t, A^*}$  and  $\sigma_{P_t^*}^2$  consider the variance-covariance matrix of the breeding value of the ancestor in generation 1,  $A$ , phenotype of the ancestor in generation 1,  $P_1$ , and the phenotype of the descendants in subsequent generations, for illustration  $P_2$  and  $P_3$ , in a situation in which selection has not taken place,

$$V \begin{bmatrix} A_1 \\ P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} \sigma_A^2 & & & \\ & \sigma_P^2 & & \\ & & \frac{1}{2}\sigma_A^2 & \frac{1}{4}\sigma_A^2 \\ & & \frac{1}{2}\sigma_A^2 & \frac{1}{4}\sigma_A^2 \\ \text{symmetric} & & & \sigma_P^2 \end{bmatrix} \quad (16)$$

If selection is imposed on the phenotypes of the ancestors in generation 1 then the variance-covariance matrix of (16) conditioned on this selection, using

conditional variance and covariance theory (Tallis, 1987) is,

$$\begin{bmatrix} \sigma_A^2(1-h^2k_x) & \sigma_A^2(1-k_x) & \frac{1}{2}\sigma_A^2(1-h^2k_x) \\ & \sigma_P^2(1-k_x) & \frac{1}{2}\sigma_A^2(1-k_x) \\ & & \sigma_P^2(1-\frac{1}{4}h^4k_x) \end{bmatrix} \begin{bmatrix} \frac{1}{4}\sigma_A^2(1-h^2k_x) \\ \frac{1}{4}\sigma_A^2(1-k_x) \\ \frac{1}{2}\sigma_A^2(1-\frac{1}{4}h^2k_x) \\ \sigma_P^2(1-\frac{1}{16}h^4k_x) \end{bmatrix} \quad (17)$$

symmetric

The regression  $\beta_{P_2, A^*}$  for an ancestor of sex  $x$  can be found from (17) to be,

$$\beta_{P_2, A^*} = \frac{\frac{1}{2}\sigma_A^2(1-h^2k_x)}{\sigma_A^2(1-h^2k_x)} = \frac{1}{2}$$

as expected, and also as expected,

$$\sigma_{P_2^*}^2 = \sigma_P^2(1 - \frac{1}{4}h^4k_x).$$

To obtain matrix (18), selection has been imposed on the phenotypes of the ancestors. Now to obtain  $b_{xwy^*,3}$  selection must also be imposed on the phenotypes of individuals in generation 2. The sex of the ancestor is now irrelevant, but the sex ( $w$ ) of the descendant in generation 2 is important. Therefore before imposing further generations of selection on the variance-covariance matrix replace the  $k_x$  in (17) by the average  $k$ , and for simplicity, represent the resulting matrix by,

$$\begin{bmatrix} V_{00} & V_{01} & V_{02} & V_{03} \\ & V_{11} & V_{12} & V_{13} \\ & & V_{22} & V_{23} \\ & & & V_{33} \end{bmatrix} \quad (18)$$

By next imposing selection onto the index values of individuals of sex  $w$  in generation 2 and performing as analogous step as from (16) to (17), matrix (18) becomes,

$$\begin{bmatrix} V_{00} - \frac{V_{02}^2}{V_{22}}k_w & V_{01} - \frac{V_{02}V_{12}}{V_{22}}k_w & V_{02}(1-k_w) \\ & V_{11} - \frac{V_{12}^2}{V_{22}}k_w & V_{12}(1-k_w) \\ & & V_{22}(1-k_w) \\ & & & V_{03} - \frac{V_{02}V_{23}}{V_{22}}k_w \\ & & & V_{13} - \frac{V_{12}V_{23}}{V_{22}}k_w \\ & & & V_{23}(1-k_w) \\ & & & & V_{33} - \frac{V_{23}^2}{V_{22}}k_w \end{bmatrix} \quad (19)$$

and  $\beta_{P_3, A^*}$  and  $\sigma_{P_3^*}^2$  are simply found from (19) to be,

$$\beta_{P_3, A^*} = \frac{V_{03} - \frac{V_{02} V_{23}}{V_{22}} k_w}{V_{00} - \frac{V_{02}^2}{V_{22} k_w}} \quad \text{and} \quad \sigma_{P_3^*}^2 = V_{33} - \frac{V_{23}^2}{V_{22}} k_w.$$

By this method, all the conditional regression coefficients can be found. Their exact expressions are cumbersome to write explicitly, but the recursion can be easily incorporated into a computer algorithm in which a variance-covariance matrix, as defined in (16) is considered, but which has dimensions of one plus the total number of generations. In each generation, selection must be imposed on both male and female parents of the descendants in generation  $t$  using  $k_m$  and  $k_f$ . However, the variance-covariance matrix which is the basis for the conditioning imposed in subsequent generations, must have the selection imposed using the mean  $k$ -value, because only the sex of the descendant (in generation  $t$ ) and parent of the descendant are relevant at each stage.

Conditioning on all generations of previous selection accounts for the reduction in variance as consequence of selection (Bulmer, 1971). The method accounts for the increasing 'competitiveness' of the contemporaries, since all individuals born in a given generation  $t$  are offspring born as a result of  $t-1$  generations of selection. An alternative derivation for  $\beta_{P_3, A^*}$  is presented in Appendix III which gives greater insight into the concept of the increasing competitiveness of the contemporaries.

(c) *Properties of the regression coefficients of selective advantage*

The regression coefficients of selection of offspring (generation 2) are related from (14) by,

$$b_{mm,2} = \frac{z_m}{z_f} b_{mf,2} = \frac{F}{M} b_{fm,2} = \frac{F}{M} \frac{z_m}{z_f} b_{ff,2},$$

and the asymptotic regression coefficients are related, due firstly to the distributive process implicit in (13), and secondly because the terms  $b_{xwy^*,t}$  tend to zero

because they are a function of  $\frac{1}{2^{t-1}}$ ,

$$b_{mm,\infty} = \frac{M}{F} b_{mf,\infty} = \frac{F}{M} b_{fm,\infty} = b_{ff,\infty}.$$

Clearly, when  $M = F$ , the regression coefficients in any generation are equal.

In addition, the ratios

$$\frac{b_{mm,\infty}}{b_{mm,2}} \quad \text{and} \quad \frac{b_{fm,\infty}}{b_{fm,2}} \rightarrow 1 + \frac{z_f M}{z_m F} \quad \text{as} \quad h^2 \rightarrow 0$$

and

$$\frac{b_{mf,\infty}}{b_{mf,2}} \quad \text{and} \quad \frac{b_{ff,\infty}}{b_{ff,2}} \rightarrow 1 + \frac{z_m F}{z_f M} \quad \text{as} \quad h^2 \rightarrow 0.$$

This ratio is the relative selective advantage of Robertson (1961) ( $Q$  in equations (1) and (2)) who expected it to asymptote to 2 when  $M = F$ . His expectation is therefore a limiting case. From simulation (see Tables 2 and 5) this ratio has been found to asymptote to considerably less than 2, even for low heritabilities. Robertson's argument did not account for the increasing 'competitiveness' of contemporaries as selection proceeds as discussed above.

Example

For an example of  $M = 20$ ,  $F = 40$ ,  $n_f = 3$ ,  $h^2 = 0.4$  and  $\sigma_p = 10$  units, predicted regression coefficients ( $b_{xy,t}$ ) and those calculated from simulation for this example are presented in Table 2. The  $b_{xwy^*,t}$  terms decrease by more than fifty percent each generation and soon become zero.

(ii) *Accumulation of the total sum of squares of long-term relationships*

Let the expectation of the total sum of squares of long-term relationships of contributions ( $\sum r_{i,t}^2$ ) from ancestors of the first generation to descendants of generation  $t$  be notated  $TSS_t$ . Considering male and female ancestors separately,  $TSS_t$  can be partitioned into the sum of squares due to long-term contributions of male ancestors  $TSS_{m,t}$  and that due to female ancestors  $TSS_{f,t}$ ,

$$TSS_t = TSS_{m,t} + TSS_{f,t}. \tag{20}$$

The Mendelian sampling of the male and female ancestors is independent so there is no cross-product term.  $TSS_{m,t}$  and  $TSS_{f,t}$  can be partitioned further by considering the sum of squares ( $SS$ ) of long term contributions to male and female descendants separately. If a given ancestor has more than the average number of male descendants because of its selective advantage, then it is also likely to have more than the average number of female descendants, therefore a cross-product is involved ( $CP$ ),

$$TSS_{m,t} = SS_{mm,t} + 2CP_{m,t} + SS_{mf,t} \quad \text{and} \quad TSS_{f,t} = SS_{fm,t} + 2CP_{f,t} + SS_{ff,t}. \tag{21}$$

(a) *Total sum of squares of contributions following one generation of selection*

Following one generation of selection, that is selection of the first generation ancestors when  $t = 2$ , the sum of squares from parents of sex  $x$  to offspring of sex  $y$  is,

$$SS_{xy,2} = E[SS_{i(xy),2}] = E\left[\sum_{i=1}^x r_{i(xy),2}^2\right] = E\left[\sum_{i=1}^x [\mu_{i(xy),2}^2 + \sigma_{i(xy),2}^2]\right],$$



Table 2. Predicted and simulated regression coefficients for a population of  $M = 20$ ,  $F = 40$ ,  $n_f = 3$ , undergoing mass selection for a trait with  $h^2 = 0.4$

Generation $t$	Predicted (simulated) regression coefficients							
	$b_{mm,t}$	$b_{mf,t}$	$b_{fm,t}$	$b_{ff,t}$	$b_{xmm^*,t}$	$b_{xfm^*,t}$	$b_{xmf^*,t}$	$b_{xjf^*,t}$
2	0.076 (0.076)	0.111 (0.112)	0.038 (0.036)	0.055 (0.060)	0.000	0.000	0.000	0.000
3	0.094 (0.088)	0.173 (0.168)	0.047 (0.048)	0.087 (0.088)	0.028	0.014	0.040	0.021
4	0.101 (0.096)	0.196 (0.184)	0.050 (0.052)	0.098 (0.100)	0.010	0.005	0.015	0.008
5	0.103 (0.096)	0.204 (0.188)	0.052 (0.052)	0.102 (0.100)	0.004	0.002	0.005	0.003
6	0.104 (0.100)	0.207 (0.196)	0.052 (0.052)	0.104 (0.100)	0.001	0.001	0.002	0.001
⋮								
10	0.104 (0.100)	0.209 (0.196)	0.052 (0.052)	0.104 (0.100)	0.000	0.000	0.000	0.000
$\frac{b_{xy,10}}{b_{xy,2}}$	1.37 (1.32)	1.88 (1.75)	1.37 (1.44)	1.88 (1.67)				

since  $E[r_{i(xy),2}] = \mu_{i(xy),2}$  and  $\text{Var}[r_{i(xy),2}] = \sigma_{i(xy),2}^2$  by definition.  $\mu_{i(xy),2} = (\frac{1}{2})^t [Y/X + b_{xy,2} A_i]$ , where the  $(\frac{1}{2})^t$  comes from number of descendants in generation  $t$  ( $2^{t-2}$ ) divided by the value of each contribution in generation  $t$  ( $2^{t-1}$ ).  $\mu_{i(xy),2}$  can be interpreted as a  $(\frac{1}{2})^t$  multiplied by a binomial mean,  $n_x [p_y + (b_{xy,2}/n_x) A_i]$  where  $[p_y + (b_{xy,2}/n_x) A_i]$  is the probability that an offspring is selected from the  $i$ th ancestor. Similarly it follows that,

$$\sigma_{i(xy),2}^2 = \frac{1}{4} \left\{ n_x \left[ p_y + \frac{b_{xy,2}}{n_x} A_i \right] \left[ 1 - p_y - \frac{b_{xy,2}}{n_x} A_i \right] + f_{S_{xy,2}} \right\}, \quad (22)$$

where the  $\frac{1}{4}$  is from (total number of descendants)/(the value of each contribution) $^2 = \frac{1}{2}^t$ . The first part of  $\sigma_{i(xy),2}^2$  is the binomial variance of offspring of sex  $y$  from parents of sex  $x$ , and  $f_{S_{xy,2}}$  is the expected additional covariance due to the  $n_f$  members of a full-sib family having selective advantage due to the mate of the parent of sex  $x$ . Under the assumption of hierarchical mating,

$$f_{S_{xy,2}} = \frac{F}{X} n_f (n_f - 1) \frac{1}{4} h_2^2 z_y^2 \quad (23)$$

where  $\frac{1}{4} h_2^2$  is the intra-class correlation of full-sibs given that the correlation via the parent  $x$  has already been accounted for. Equation (23) depends on bivariate normal distribution theory and is only a first order approximation; Mendell & Elston (1974) provide a second order approximation which is necessary at high selection intensities and heritabilities,

$$f_{S_{xy,2}} = \frac{F}{X} n_f (n_f - 1) [\Phi(\frac{1}{4} h_2^2 i_x - v_x) p_y - p_y^2],$$

where  $\Phi$  is the cumulative normal distribution probability.

It follows that  $SS_{xy,2}$  is,

$$SS_{xy,2} = E[SS_{i(xy),2}] = \frac{X}{4} [n_x^2 p_y^2 + n_x p_y (1 - p_y) + b_{xy,2}^2 (1 - 1/n_x) V_x + f_{S_{xy,2}}], \quad (24)$$

where  $V_x$  was defined in equation (9). The cross-product terms are similarly derived,

$$CP_{i(x),2} = \sum_{i=1}^x [\mu_{i(xm),2} \mu_{i(xf),2} + \sigma_{i(xm,xf),2}], \quad (25)$$

where  $\sigma_{i(xm,xf),2}$  is the covariance of selective advantage between male and female full-sibs due to the mate of the parent of sex  $x$ , so that  $\sigma_{i(xm,xf),2} = \frac{1}{4} f_{S_{xm,xf,2}}$  and analogous to (23),

$$f_{S_{xm,xf,2}} = \frac{F}{X} n_f^2 \frac{1}{4} h_2^2 z_m z_f. \quad (26)$$

(b) Total sums of squares of contributions in subsequent generations

In subsequent generations the sums of squares will be dependent on the consequences of the latest generation of selection conditional on the previous selection. For a simple analogy, consider the variance of a variable  $v$ , which occurs subsequently to, and is dependent on a variable  $u$ , then  $\text{Var}(v) = E_u[\text{Var}(v|u)] + \text{Var}_u(E[v|u])$ . If  $v$  is the sum of  $u$  random variables with mean  $\mu_u$  and variance  $\sigma_u^2$  then  $E(v) = \mu_u \mu_u$  and  $\text{Var}(v) = \mu_u \sigma_u^2 + \sigma_u^2 \mu_u^2$  and then for  $N$  observations  $\text{TSS}_v = N[\mu_u \sigma_u^2 + \mu_u^2 (\sigma_u^2 + \mu_u^2)]$ . Extending this analogy to the two-sex case relevant here, a recursive relationship then exists between the total sum of squares in generation  $t$  with those in generation  $t - 1$ ,

$$S_{x,t} = E[S_{i(x),t}] = E[\mathbf{M}_{i(x^*),t} \mathbf{S}_{i(x),t-1} \mathbf{M}'_{i(x^*),t}] + X E[\mu_{i(xm),t-1} \mathbf{V}_{i(xmy^*),t}] + X E[\mu_{i(xf),t-1} \mathbf{V}_{i(xfy^*),t}], \quad (27)$$

where

$$\mathbf{S}_{i(x),t} = \begin{bmatrix} SS_{i(xm),t} & CP_{i(x),t} \\ CP_{i(x),t} & SS_{i(xf),t} \end{bmatrix},$$

$$\mathbf{M}_{i(x^*),t} = \begin{bmatrix} \mu_{i(xmm^*),t} & \mu_{i(xfm^*),t} \\ \mu_{i(xmf^*),t} & \mu_{i(xjf^*),t} \end{bmatrix}$$

and

$$\mathbf{V}_{i(xwy^*),t} = \begin{bmatrix} \sigma_{i(xwm^*),t}^2 & \sigma_{i(xwm,xfm^*),t} \\ \sigma_{i(xwm,xfm^*),t} & \sigma_{i(xwf^*),t}^2 \end{bmatrix}.$$

Where  $\mu_{i(xwy^*),t}$  (from equation (12)) and  $\sigma_{i(xwy^*),t}^2$  are the mean and variance of the additional contributions of an ancestor of sex  $x$  to a descendant of sex  $y$  in generation  $t$  via a descendant of sex  $w$  in generation  $t-1$ , over the contribution from  $x$  to  $w$ . It follows that,

$$\mu_{i(xwy^*),t} = \frac{1}{2} [n_w p_y + b_{xwy^*,t} A_i], \tag{28}$$

$$\sigma_{i(xwy^*),t}^2 = \frac{1}{2^t} \{n_w [p_y + b_{xwy^*,t} A_i] [1 - p_y - b_{xwy^*,t} A_i] + f_{s_{wy,t}}\}, \tag{29}$$

$$\sigma_{i(xwm, xw^*),t} = \frac{1}{2^t} f_{s_{wm, wf,t}}. \tag{30}$$

The terms  $f_{s_{wy,t}}$  and  $f_{s_{wm, wf,t}}$  are as defined in equations (23) and (26) but replacing the subscript 2 with  $t$ , however, using  $h_2^2$  throughout has little consequence.

The prediction of the total sums of squares requires only the correct accumulation of terms via (27) of the regression coefficients of selective advantage predicted in the previous section.

Examples

To demonstrate the accumulation of terms consider a simple example when  $M = F = 20$ ,  $n_f = 3$  and  $h^2 = 0$ . In this case there is no selective advantage for ancestor breeding value so all regression coefficients are null. Since all  $A_i$  are zero the  $i$  subscripts are omitted. With  $M = F$  the  $SS_{xy,t}$  are all equal. Equation numbers relevant to the calculations are presented in square brackets following the answer in this section.

$$\mu_{xy,2} = \frac{1}{2} 3 \frac{1}{3} = \frac{1}{2} \quad \text{and} \quad \sigma_{xy,2}^2 = \frac{1}{4} 3 \frac{1}{3} \frac{2}{3} = \frac{1}{6} [22].$$

$$\text{Then } SS_{xy,2} = 20 \left[ \frac{1}{2} \frac{1}{2} + \frac{1}{6} \right] = 8 \cdot 3 [24]$$

$$\text{and } CP_x = 20 \left[ \frac{1}{2} \frac{1}{2} + 0 \right] = 5 [25].$$

$$TSS_{m,2} = TSS_{f,2} = 8 \cdot 3 + 8 \cdot 3 + 2 \times 5 = 26 \cdot 6 [21]$$

$$\text{and } TSS_2 = 53 \cdot 3 [20].$$

$$\mu_{xwy^*,3} = \frac{1}{2} [28] \quad \sigma_{xwy^*,3}^2 = \frac{1}{2^3} \frac{2}{3} = 0 \cdot 083 [29]$$

$$\sigma_{xwm, xwf^*,3} = 0 [30].$$

$$\begin{aligned} & \begin{bmatrix} SS_{xm,3} & CP_{x,3} \\ CP_{x,3} & SS_{xf,3} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 8 \cdot 3 & 5 \\ 5 & 8 \cdot 3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ &+ 20/2 \begin{bmatrix} 0 \cdot 083 & 0 \\ 0 & 0 \cdot 083 \end{bmatrix} + 20/2 \begin{bmatrix} 0 \cdot 083 & 0 \\ 0 & 0 \cdot 083 \end{bmatrix} \\ &= \begin{bmatrix} 6 \cdot 6 & 6 \cdot 6 \\ 6 \cdot 6 & 6 \cdot 6 \end{bmatrix} + \begin{bmatrix} 1 \cdot 6 & 0 \\ 0 & 1 \cdot 6 \end{bmatrix} \\ &= \begin{bmatrix} 8 \cdot 3 & 6 \cdot 6 \\ 6 \cdot 6 & 8 \cdot 3 \end{bmatrix} [27] \end{aligned}$$

$$\text{and } TSS_3 = 60 [20].$$

Similarly,

$$\begin{aligned} & \begin{bmatrix} SS_{xm,4} & CP_{x,4} \\ CP_{x,4} & SS_{xf,4} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 8 \cdot 3 & 6 \cdot 6 \\ 6 \cdot 6 & 8 \cdot 3 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ &+ 20/2 \begin{bmatrix} 0 \cdot 042 & 0 \\ 0 & 0 \cdot 042 \end{bmatrix} + 20/2 \begin{bmatrix} 0 \cdot 042 & 0 \\ 0 & 0 \cdot 042 \end{bmatrix} \\ &= \begin{bmatrix} 8 \cdot 3 & 7 \cdot 5 \\ 7 \cdot 5 & 8 \cdot 3 \end{bmatrix} [27] \quad TSS_4 = 63 \cdot 3 [20]. \end{aligned}$$

Under no selection or zero heritability the  $SS_{xy,t}$  do not change as  $t$  increases, however, the  $CP_x$  increase geometrically to a limiting value of  $CP_{x,t} = SS_{xy,2}$  when  $TSS = 66 \cdot 6$ , and from (8),  $\Delta \mathcal{F} = 0 \cdot 0104$ .

For a population under selection,  $M = 20$ ,  $F = 40$ ,  $n_f = 3$ ,  $h^2 = 0 \cdot 4$ , the predicted and total sums of squares are presented in Table 3. The recursion was continued until  $t = 15$ . The limiting value of  $TSS = E[\sum r_i^2]$  was predicted to be 172 \cdot 1. From equation (8),  $\Delta \mathcal{F}$  is predicted to be 0 \cdot 0120 which compares to that of 0 \cdot 0118 calculated directly from simulation. Predicted ( $\Delta \mathcal{F}_c$ ) and simulated ( $\Delta \mathcal{F}_{sim}$ ) rates of inbreeding for the range of mating ratios, heritabilities and offspring/dam (selection intensities) are presented in Table 4.  $\Delta \mathcal{F}_c$  is in very good agreement with (usually within 5% and always within 11% of)  $\Delta \mathcal{F}_{sim}$  for these example populations.

5. Discussion

A method has been presented which relates rate of inbreeding to long-term additive genetic contributions between ancestors and descendants. An expression for effective population size can be derived from equation (8) by expressing  $\sum r_i^2$  as  $(M+F)(\mu_r^2 + \sigma_r^2)$ , where  $\mu_r$  and  $\sigma_r^2$  are the mean and variance of long-term contributions, and if  $N = M+F$ , it follows that,

$$N_e \approx \frac{2N}{\mu_r^2 + \sigma_r^2}. \tag{31}$$

This equation can be related to the more usual expression,

$$N_e \approx \frac{4N}{\mu_s + \sigma_s^2}, \tag{32}$$

(Wright, 1931) where  $\mu_s$  and  $\sigma_s^2$  are the mean and variance of one generation family size. Under no selection (31) and (32) should be equal. Let  $M = F$  and let the distribution of family size be Poisson, then each individual is expected to contribute two offspring to the next generation, one of each sex, so that  $\mu_s = \sigma_s^2 = 2$ , so that  $N_e$  from (32) is  $N$  as expected. Considering the contributions from parents to offspring, each parent contributes half the genes to both their expected offspring, so that the mean and variance of the contributions to the first generation are  $\mu_r = \mu_s/2$  and  $\sigma_r^2 = \sigma_s^2/4$ . From the example presented in section 4, it is observed that under no selection  $\mu_r$  remains con-

Table 3. Predicted and simulated sums of squares for a population of  $M = 20$ ,  $F = 40$  and  $n_f = 3$  undergoing mass selection for a trait with  $h^2 = 0.4$

Generation $t$	Predicted (simulated) SS of contributions from first generation ancestors						
	$SS_{mm}$	$SS_{mf}$	$TSS_m$	$SS_{fm}$	$SS_{ff}$	$TSS_f$	$TSS$
2	10.2 (9.9)	28.9 (28.3)	63.1 (61.4)	7.4 (7.2)	18.0 (17.8)	38.3 (37.4)	101.4 (98.8)
3	11.2 (10.2)	37.0 (34.3)	82.8 (76.7)	7.8 (7.3)	24.3 (23.2)	52.4 (51.4)	136.6 (128.2)
4	11.3 (10.3)	41.7 (37.9)	93.6 (85.3)	7.8 (7.5)	27.9 (27.2)	66.4 (60.7)	155.9 (145.9)
5	11.4 (10.5)	43.8 (37.9)	98.4 (90.2)	7.8 (7.8)	29.6 (28.9)	68.2 (65.4)	164.8 (155.7)
6	11.4 (10.5)	44.8 (41.5)	100.6 (93.2)	7.8 (7.8)	30.4 (31.0)	68.2 (68.3)	168.8 (161.5)
⋮							
10	11.4 (10.7)	46.8 (42.7)	102.3 (96.0)	7.8 (7.9)	31.0 (31.4)	68.2 (70.7)	172.1 (166.7)

Table 4. Rates of inbreeding ( $\times 100$ ) predicted from long-term contributions ( $\Delta\mathcal{F}_c$ ) and observed from simulation ( $\Delta\mathcal{F}_{sim}$ ) for populations with  $M = 20$

$h^2$	Predicted (simulated)			
	$F = 20$	$F = 40$	$F = 100$	$F = 200$
$n_f = 3$				
0.0	1.04 (1.07)	0.83 (0.83)	0.71 (0.71)	0.67 (0.66)
0.1	1.19 (1.23)	0.99 (0.98)	0.89 (0.82)	0.87 (0.82)
0.2	1.30 (1.33)	1.10 (1.10)	0.99 (0.94)	0.99 (0.86)
0.4	1.42 (1.42)	1.20 (1.18)	1.07 (1.05)	1.06 (1.00)
0.6	1.47 (1.50)	1.21 (1.23)	1.05 (1.02)	1.02 (0.99)
$n_f = 6$				
0.0	1.15 (1.13)	0.89 (0.88)	0.73 (0.73)	0.68 (0.68)
0.1	1.46 (1.44)	1.17 (1.17)	0.99 (0.95)	0.97 (0.87)
0.2	1.68 (1.61)	1.35 (1.30)	1.13 (1.10)	1.14 (1.03)
0.4	1.91 (1.92)	1.52 (1.50)	1.24 (1.24)	1.24 (1.17)
0.6	2.02 (1.98)	1.52 (1.50)	1.21 (1.23)	1.18 (1.19)

Table 5. Rates of inbreeding ( $\times 100$ ) observed from simulation ( $\Delta\mathcal{F}_{sim}$ ), predicted from long term contributions ( $\Delta\mathcal{F}_c$ ) and by the methods of Burrows (1984a) and Robertson (1961) for populations with full-sib families

$h^2$	$\Delta\mathcal{F}_{sim}$	$\Delta\mathcal{F}_c$	$\Delta\mathcal{F}_B$	$\Delta\mathcal{F}_{1,R}$	$\Delta\mathcal{F}_R$	$\Delta\mathcal{F}_Q$	$Q$
$M = 20, F = 20, n_f = 3$							
0.1	1.23	1.19	1.13	1.09	1.23	1.20	1.87
0.2	1.33	1.30	1.17	1.13	1.38	1.31	1.77
0.4	1.42	1.42	1.27	1.19	1.61	1.43	1.60
0.6	1.50	1.47	1.30	1.24	1.72	1.46	1.46
$M = 20, F = 20, n_f = 6$							
0.1	1.44	1.46	1.27	1.25	1.58	1.52	1.86
0.2	1.61	1.68	1.38	1.35	1.93	1.76	1.75
0.4	1.92	1.91	1.57	1.49	2.42	1.99	1.57
0.6	1.98	2.02	1.68	1.57	2.61	2.01	1.42

stant, and the mean of the long-term contributions is therefore  $\mu_r = 1$ . However, the variance increases to twice its initial value and  $\sigma_r^2 = \sigma_s^2/2 = 1$ , and from (31),  $N_e = N$ . The increase in the contribution variance is attributed to the covariance induced between males and females when more than one generation is considered.

In table 5  $\Delta\mathcal{F}_c$  is compared to predictions from Burrows (1984a) ( $\Delta\mathcal{F}_B$ ) and from Robertson (1961) for populations with a full-sib family structure.  $\Delta\mathcal{F}_{1,R}$  is from equation (2) using  $Q^2 = 1$  and the corrected  $\rho = (h^2/2)(1 - kh^2)/(1 - kh^4/2)$ , and is directly comparable to  $\Delta\mathcal{F}_B$  (which used the same  $\rho$ ).  $\Delta\mathcal{F}_R$  is from (2) using  $Q^2 = 4$  and the incorrect  $\rho = (h^2/2)(1 - kh^2)$  and  $\Delta\mathcal{F}_Q$  is from (2) using the correct  $Q$  and  $\rho$ ;  $Q$  is also tabulated and was calculated as the increase in regression coefficients as discussed in section 3.  $\Delta\mathcal{F}_B$  gives better and higher predictions than  $\Delta\mathcal{F}_{1,R}$  because it accounts for higher order terms, but both are underpredictors of  $\Delta\mathcal{F}_{sim}$ .  $\Delta\mathcal{F}_R$  is a great overestimator of  $\Delta\mathcal{F}_{sim}$  but  $\Delta\mathcal{F}_Q$  is a very good predictor and is similar to the  $\Delta\mathcal{F}_c$  predictions. The values of  $Q$  (Table 5) illustrate how the relative selective advantage

$\Delta\mathcal{F}_{1,R}$  from equation (2), with  $Q = 1$  and  $\rho = \frac{1}{2}h^2(1 - kh^2)/(1 - kh^4/2)$ .  
 $\Delta\mathcal{F}_R$  from equation (2), with  $Q = 2$  and  $\rho = \frac{1}{2}h^2(1 - kh^2)$ .  
 $\Delta\mathcal{F}_Q$  from equation (2), with  $Q$  as listed and  $\rho = \frac{1}{2}h^2(1 - kh^2)/(1 - kh^4/2)$ .

asymptotes to a limiting value of much less than the 2 which Robertson (1961) expected. Two important facets of the recursion prediction presented here are that account is taken of reduction in genetic variance due to selection and also of the increasing competitiveness of contemporaries as selection proceeds.

An approximate relationship between the recursion equation (27) and Robertson's equation (2) can be shown. Consider the recursion (27) to be,

$$\begin{aligned}
 S_{x,t} &= E[S_{t(x),t}] \approx E[M_{t(x^*),t} S_{t(x),t} M'_{t(x^*),t}] \\
 &= E[M_{t(x^*),t} M_{t(x^*),t-1} \dots M_{t(x^*),3} \\
 &\quad S_{t(x),2} M'_{t(x^*),3} \dots M'_{t(x^*),t-1} M'_{t(x^*),t}] \quad (33)
 \end{aligned}$$

where the elements of  $\prod_{j=3}^t M'_{t(x^*),j}$  are the accumulation of the regression coefficients. In populations with  $M = F$  equation (33) reduces to a scalar equation in which  $\prod_{j=3}^t M'_{t(x^*),j}$  is equivalent to  $Q$ . Even in

populations with  $M < F$  an approximation to (27) can be made,

$$\Delta\mathcal{F} = \Delta\mathcal{F}_0 + Q^2(\Delta\mathcal{F}_1 - \Delta\mathcal{F}_0) \quad (34)$$

where  $Q$  is the mean of the four ratios  $b_{xy,\infty}/b_{xy,2}$ ,  $\Delta\mathcal{F}_0$  is the rate of inbreeding expected under random mating and  $\Delta\mathcal{F}_1$  is the rate of inbreeding following a single generation of selection predicted, for example via the equation of Latter (1959) and Hill (1979) using estimated variances of family size under selection (Wray, 1989). The term  $(\Delta\mathcal{F}_1 - \Delta\mathcal{F}_0)$  is the difference in rate of inbreeding due to a single round of selection which is scaled by the  $Q^2$  to predict the total inbreeding accumulated;  $\Delta\mathcal{F}_1$  is of course a larger term than the corresponding quantity implied in  $S_{x,2}$ . Predicted  $\Delta\mathcal{F}$  using the approximation for the same examples as before are found in Table 6. In general, the predictions are not as good as those from the recursion (Table 4). However, these predictions may be more easily attainable and may be sufficient to ensure correct ranking of alternative breeding schemes with respect to  $\Delta\mathcal{F}$ .

In the examples presented in Table 3,  $\Delta\mathcal{F}_c$  tends to overpredict slightly for populations which have both high mating ratios and are selected on traits with high heritabilities. This overprediction is found to be reduced when total population size is increased (Wray, 1989) and is attributed to the normal distribution, large population assumptions ( $i, v$  and  $k$  values) in the derivation. For this reason, the prediction is not good for very small populations ( $M < 15$ ). In addition, the effect of inbreeding itself on heritability is ignored which in very small populations may be important. Finally, the derivation assumed  $\Delta\mathcal{F}$  to be approximately constant each generation, an assumption which is more likely to be violated in very small populations.

Extension of the recursion to consider index selection should be straightforward, affecting only  $b_{xy,v}fs_{xy,t}$  and  $fs_{xm,xf,v}$  although the recursion is likely to overpredict if there is a very high intra-class correlation of sibs. The assumption of fixed family size

Table 6. Predicted  $\Delta\mathcal{F}$  ( $\times 100$ ) using the approximation of equation (34),  $M = 20$

$h^2$	Predicted (simulated)			
	$F = 20$	$F = 40$	$F = 100$	$F = 200$
$n_f = 3$				
0.1	1.23 (1.23)	1.02 (0.98)	0.90 (0.82)	0.88 (0.82)
0.2	1.35 (1.33)	1.13 (1.10)	1.01 (0.94)	1.00 (0.86)
0.4	1.46 (1.42)	1.23 (1.18)	1.12 (1.05)	1.12 (1.00)
0.6	1.48 (1.50)	1.24 (1.23)	1.12 (1.02)	1.12 (0.99)
$n_f = 6$				
0.1	1.54 (1.44)	1.21 (1.17)	1.01 (0.95)	0.95 (0.87)
0.2	1.78 (1.61)	1.40 (1.30)	1.17 (1.10)	1.12 (1.03)
0.4	2.00 (1.92)	1.57 (1.50)	1.32 (1.24)	1.27 (1.17)
0.6	2.01 (1.98)	1.57 (1.50)	1.32 (1.23)	1.27 (1.19)

available for selection is considered reasonable for the types of livestock populations in which prediction of inbreeding might be important; extension to variable family size requires further consideration as do populations with non-hierarchical mating schemes. In general, other mating schemes are likely to introduce a situation of overlapping generations. Hill (1979) demonstrated that a discrete generation derivation of effective population size at equilibrium holds for overlapping generations when averaged over generation interval. A similar approximation is likely to be applicable to the prediction method presented here.

We would like to thank Professor W. G. Hill and Mr. John Woolliams for many useful discussions and for constructive comments on manuscripts of this paper.

## References

- Bulmer, M. G. (1971). The effect of selection on genetic variability. *The American Naturalist*, **105**, 201–211.
- Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- Burrows, P. M. (1984a). Inbreeding under selection from unrelated families. *Biometrics* **40**, 357–366.
- Burrows, P. M. (1984b). Inbreeding under selection from related families. *Biometrics* **40**, 895–906.
- Crow, J. F. & Denniston, C. (1988). Inbreeding and variance effective number. *Evolution* **42**, 482–495.
- Crow, J. F. & Morton, N. E. (1955). The measurement of gene frequency in small populations. *Evolution* **9**, 482–495.
- Dempster, E. R. & Lerner, I. M. (1950). Heritability of threshold characters. *Genetics* **35**, 212–236.
- Hill, W. G. (1972). Effective size of populations with overlapping generations. *Theoretical Population Biology* **3**, 278–289.
- Hill, W. G. (1979). A note on effective population size with overlapping generations. *Genetics* **92**, 317–322.
- Hill, W. G. (1985). Fixation probabilities of mutant genes with artificial selection. *Génétiques Sélection et Evolution* **17**, 351–358.
- James, J. W. & McBride, G. (1958). The spread of genes by natural and artificial selection in a closed poultry flock. *Journal of Genetics* **56**, 55–62.
- Jones, L. P. (1969). Effects of artificial selection on rates of inbreeding in populations of *Drosophila Melanogaster*. I. Effect in early generations. *Australian Journal of Biological Science* **22**, 143–155.
- Mendell, N. C. & Elston, R. C. (1974). Multifactorial traits: Genetic analysis and prediction of recurrence risks. *Biometrics* **30**, 41–57.
- Nei, M. & Murata, M. (1966). Effective population size when fertility is inherited. *Genetical Research* **8**, 257–260.
- Robertson, A. (1961). Inbreeding in artificial selection programmes. *Genetical Research* **2**, 189–194.
- Thompson, R. (1977). The estimation of heritability with unbalanced data. II. Data available on more than two generations. *Biometrics* **33**, 497–504.
- Tallis, G. M. (1987). Ancestral covariance and the Bulmer effect. *Theoretical and Applied Genetics* **73**, 815–820.
- Woolliams, J. A. (1989). Modifications to MOET nucleus breeding schemes to improve rates of genetic progress and decrease rates of inbreeding in dairy cattle. *Animal Production* **49**, 1–14.
- Wray, N. R. (1989). *Consequences of selection in closed populations with particular reference to closed nucleus herds of pigs*. PhD thesis. University of Edinburgh.

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist* **56**, 330–338.  
 Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.  
 Yoo, B. H. (1980). Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster*. V. The inbreeding effect of selection. *Australian Journal of Biological Science* **33**, 713–723.

**Appendix I**

*Summary of symbols used in the text.*

This summary should be used in conjunction with the notation defined in section 4.

The matrices defined in equation (27) are not included.

<b>1</b>	Vector with all elements unity.
<b>A</b>	relationship matrix, $A_{ij}$ is a sub-matrix of <b>A</b> representing relationships between individuals in generations $i$ and $j$ .
$A_i$	deviation of the true breeding value of individual $i$ , $A'_i$ from the mean breeding value of its contemporaries; for convenience $A_i$ is referred to simply as the breeding value of $i$ .
$a_{pq}$	additive genetic relationship between individuals $p$ and $q$ ; the $(p, q)$ element of <b>A</b> .
$b_{xy,t}$	regression of total number of descendants of sex $y$ in generation $t$ selected from ancestors (in generation 1) of sex $x$ .
$b_{xwy,t}$	regression as $b_{xy,t}$ above but via descendants of sex $w$ in generation $t-1$ (the parents of the descendants of sex $y$ in generation $t$ ).
$b_{xwy^*,t}$	regression as $b_{xwy,t}$ above but conditioned on the selection in generations 1 to $t-1$ .
<b>C<sub>k</sub></b>	matrix of contributions to the relationship matrix matrix from the Mendelian sampling of genes which took place in generation $k$ .
<b>C<sub>k(t,t)</sub></b>	a submatrix of <b>C<sub>k</sub></b> which contains the partial relationships between individuals in generations $i$ and $j$ which are attributable to the Mendelian sampling of genes which took place in their ancestors of generation $k$ .
$CP_{i(x),t}$	cross products of relationships of male and female descendants in generation $t$ from ancestor $i$ of sex $x$ , $CP_{x,t} = E[CP_{i(x),t}]$ of all ancestors of sex $x$ .
<b>D<sub>t</sub></b>	diagonal matrix of Mendelian sampling effects of generation $t$ defined in equation (5).
$\overline{\text{diag}}\{ \}$	the mean of the diagonal elements of the matrix within $\{ \}$ .
$\Delta\mathcal{F}$	rate of inbreeding, may have subscripts, <b>B</b> = Burrows (1984a), $c$ = contribution method presented in this paper, $Q$ = equation (2) using correct $Q$ , $R$ = Robertson (1961) (equation (2), $Q = 2$ ), $1, R$ = Robertson (1961) (equation (2) $Q = 1$ ), $sim$ = observed from simulation, $0$ = random mating population, $1$ = single generation selection approach e.g. Hill (1979).
$\mathcal{F}_t$	mean inbreeding coefficient of individuals in generation $t$ .
<b>F</b>	total number of females each generation (may be represented by $X$ or $Y$ ).
$f_s, f_d$	inbreeding coefficient of sire or dam.

$f$	subscript, representing sex of ancestor or descendant is female (may be represented by $w, x$ or $y$ ).
$fs_{xy,t}$	relationship between full-sibs of sex $y$ in generation $t$ given that their relationship due to their parent of sex $x$ in generation $t-1$ has already been accounted for.
$fs_{wm, wf,t}$	relationship between male and female full-sibs in generation $t$ given that the relationship due to their parent of sex $w$ in generation $t-1$ has already been accounted for.
$\phi_i$	Mendelian sampling additive genetic effect unique to individual $i$ .
$h^2$	heritability, with subscript $t$ represents heritability in generation $t$ , $h_1^2 = h^2$ .
<b>I<sub>0</sub></b>	relationship matrix between individuals in the base population, identity matrix.
$i$	selection intensity (may have subscript $x$ for selection of sex $x$ ).
$k$	$k = i(i-v)$ and $k_x = i_x(i_x - v_x)$ .
<b>M</b>	total number of males in population each generation (may be represented by $X$ or $Y$ ).
$m$	subscript, representing sex of ancestor or descendant is male (may be represented by $w, x$ or $y$ ).
$\mu_r$	mean of stabilised long-term contributions from ancestors to descendants.
$\mu_{i(xy),t}$	mean of long-term contributions of ancestor $i$ (with known breeding value $A_i$ ) of sex $x$ to descendants of sex $y$ in generation $t$ .
$\mu_{i(xwy^*),t}$	as $\mu_{i(xy),t}$ but via descendants of sex $w$ in generation $t-1$ and conditioned on the selection in generations 1 to $t-1$ .
<b>N</b>	total number of males and females in population each generation, $N = M + F$ .
<b>N<sub>e</sub></b>	effective population size.
$n$	number of offspring of each sex available for selection in a full-sib family
$n_x$	number of offspring of each sex available for selection from parents of sex $x$ , $n_f = n$ .
$P_t$	phenotype (selection criterion in mass selection) of descendants in generation $t$ .
$p$	proportion of individuals selected, probability of selection (may have subscript $x$ for selection of sex $x$ ).
<b>Q</b>	measure of relative selective advantage used in Robertson (1961), can be interpreted as the mean $b_{xy,\infty}/b_{xy,2}$ over all combinations of $x = m, f$ and $y = m, f$ .
$r_t$	vector of long-term contributions of ancestors to descendants in generation $t$ , of length the number of ancestors. Without the subscript, stabilisation of the contributions is implied.
$r_{i,t}$	total long-term relationship or contribution of ancestor $i$ to descendants in generation $t$ .
$r_{i(xy),t}$	total long-term relationship or contribution of ancestor $i$ of sex $x$ to descendants of sex $y$ in generation $t$ .
$\rho$	intra-class correlation of selection criteria of sibs.
$S_t$	selection score of descendants in generation $t$ , $S_t = 1$ if selected and $0$ otherwise.
$SS_{i(xy),t}$	sums of squares of contributions from ancestor $i$ of sex $x$ to descendants of sex $y$ in generation $t$ . $SS_{xy,t} = E[SS_{i(xy),t}]$ over all ancestors of sex $x$ .
$\sigma_a^2$	additive genetic variance.
$\sigma_E^2$	environmental and error variance.

$\sigma_p^2$	phenotypic variance, $\sigma_{P_t}^2$ phenotypic variance of descendants born in generation $t$ , $\sigma_{P_t^*}^2$ as $\sigma_{P_t}^2$ but conditioned on the selection of ancestors from generation 1 to $t-1$ .
$\sigma_r^2$	variance of stabilized long-term contributions from ancestors to descendants.
$\sigma_{i(xy),t}^2$	variance of long-term contributions of ancestor $i$ (with known breeding value $A_i$ ) of sex $x$ to descendants of sex $y$ in generation $t$ .
$\sigma_{i(xwy^*),t}^2$	as $\sigma_{i(xy),t}^2$ but via descendants of sex $w$ in generation $t-1$ and conditioned on the selection in generations 1 to $t-1$ .
$\sigma_{i(xm,xf),t}$	covariance of long-term contributions of ancestor $i$ (with known breeding value $A_i$ ) of sex $x$ between male and female descendants generation $t$ .
$\sigma_{i(xwm,xwf^*),t}$	as $\sigma_{i(xm,xf),t}$ but via descendants of sex $w$ in generation $t-1$ and conditioned on the selection in generations 1 to $t-1$ .
$TSS_t$	total sum of squares of long-term contributions from ancestors of generation 1 to descendants in generation $t$ , $TSS_t = E[\sum r_{i,t}^2]$ ; $TSS_{x,t}$ as $TSS_t$ but refers to ancestors of sex $x$ only, $TSS_{x,t} = E[\sum (r_{i(xm),t} + r_{i(xf),t})^2]$ .
$V_x$	variance of breeding values of ancestors of sex $x$ in generation 1.
$v$	truncation deviate of normal distribution corresponding to a proportion $p$ selected, $v_x$ truncation deviates for sex $x$ .
$z$	truncation ordinate of normal distribution, corresponding to a proportion $p$ selected, $z_x$ truncation ordinate for sex $x$ .
$Z_t$	matrix with elements 0 or $\frac{1}{2}$ which relates individuals in generation $t+1$ to individuals in generation $t$ .

**Appendix II**

Relationship between the stabilized contribution matrices and  $\Delta\mathcal{F}$ .

Firstly,

$$(1 + \mathcal{F}_{t+1}) - (1 + \mathcal{F}_t) = \sum_{k=0}^{t+1} \text{diag}\{C_{k(t+1,t+1)}\} - \sum_{k=0}^t \text{diag}\{C_{k(t,t)}\},$$

and as  $t \rightarrow \infty$ ,  $\text{diag}\{C_{0(t,t)}\} = \text{diag}\{C_{k(t,t)}\}$  so it follows that,

$$\mathcal{F}_{t+1} - \mathcal{F}_t = \sum_{k=1}^{t+1} \text{diag}\{C_{k(t+1,t+1)}\} - \sum_{k=1}^t \text{diag}\{C_{k(t,t)}\}.$$

The base generation is eliminated since it is the only generation derived from unselected parents,

$$\mathcal{F}_{t+1} - \mathcal{F}_t = \text{diag}\{C_{1(t+1,t+1)}\} + \sum_{k=1}^t \text{diag}\{C_{k+1(t+1,t+1)}\} - \sum_{k=1}^t \text{diag}\{C_{k(t,t)}\},$$

and using (6),

$$\mathcal{F}_{t+1} - \mathcal{F}_t \approx \text{diag}\{C_{1(t+1,t+1)}\} - \Delta\mathcal{F} \sum_{k=1}^t \text{diag}\{C_{k(t,t)}\},$$

$$\Delta\mathcal{F}(1 - \mathcal{F}_t) \approx \text{diag}\{C_{1(t+1,t+1)}\} - \Delta\mathcal{F}(1 + \mathcal{F}_t - \text{diag}\{C_{0(t+1,t+1)}\}),$$

and so,

$$\Delta\mathcal{F} \approx \frac{\text{diag}\{C_{1(t+1,t+1)}\}}{2 - \text{diag}\{C_{0(t+1,t+1)}\}} \approx \text{diag}\{C_{1(t+1,t+1)}\}/2.$$

**Appendix III**

Alternative derivation for the regression coefficient  $\beta_{P_3, A^*}$ .

Definitions:

$\beta_{P_3, A^*}$  is the regression coefficient of phenotype of grandoffspring (generation 3) on the breeding value of its grandparent (ancestor from generation 1) given that its grandparent and parent (ancestor in generation 2) were both selected.

$\beta_{P_3, A}$  is the regression coefficient of phenotype of grandoffspring on the breeding value of its grandparent given that the grandparent was selected on its phenotype, and  $\beta_{P_3, A} = \frac{1}{4}$ .

$P_2^* - A_t/2$  is the deviation of the phenotype of the selected parent from the breeding value of the grandparent.

The expression  $\beta_{P_3, A^*} A_t$  can expressed as the sum of two independent regressions,

$$\beta_{P_3, A^*} A_t = \beta_{P_3, A} A_t + \beta_{P_3, P_2^* - A_t/2} (P_2^* - A_t/2) \tag{A 1}$$

If the grandparent is of sex  $x$ , the parent of sex  $w$  and the grandoffspring of sex  $w$ , then

$$\beta_{P_3, P_2^* - A_t/2} = \frac{[\frac{1}{2}(1 - h_1^2 k/2)(1 - h_2^2 k_w) - \frac{1}{8}(1 - h_1^2 k_x)] h_1^2}{(1 - h_1^4 k)(1 - k_w) - \frac{1}{4}(1 - h_1^2 k_x) h_1^2} \approx \frac{1}{2} h_2^2,$$

From equation (A 1) the only unknown is the value of  $P_2^*$  which is predicted as,

$$P_2^* = \beta_{P_2^*, A_t} A_t = \left[ \frac{1}{2} + \frac{1}{2} \frac{\partial(i_w) \sigma_{P_2}}{\partial(A)} \right] A_t = \left[ \frac{1}{2} + \frac{\partial(i_w)}{\partial(x_w)} \frac{\partial(x_w) \sigma_{P_2}}{\partial(A)} \right] A_t = \left[ \frac{1}{2} - \frac{p_w \partial(z_w)/\partial(x_w) - z_w \partial(p_w)/\partial(x_w)}{p_w^2} \right] A_t = \frac{\sigma_{P_2}}{2} A_t = \frac{1}{2} [1 - k_w] A_t$$

It follows that,

$$\beta_{P_3, A^*} A_t \approx \frac{1}{4} [1 - k_w h_2^2] A_t.$$

If no account had been taken of the parents being a selected group, then  $\beta_{P_3, A^*} A_t$  would simply be  $\frac{1}{4} A_t$ , this is the assumption made by Robertson (1961). In addition,  $b_{xwy^*, 3}$  can be approximated as,

$$b_{xwy^*, 3} \approx n_w \frac{1}{4} [1 - k_w h_2^2] \frac{z_y}{\sigma_{P_3}}$$

By using similar arguments to those presented above, the general regression coefficient  $b_{xwy^*, t}$  for  $t > 3$  can be predicted as,

$$b_{xwy^*, t} \approx n_w (\frac{1}{2})^{t-1} (1 - k_w h_2^2) \left[ \prod_{i=2}^{t-2} (1 - kh_i^2) \right] \frac{z_y}{\sigma_{P_t}}$$