SIMULTANEOUS SELFING AND PARTIAL DIALLEL TEST CROSSING

III.* OPTIMUM SELECTION PROCEDURES†

By C. Clark Cockerham‡ and D. F. Matzinger‡

[*Manuscript received March 7, 1966*]

Summary

Considered was the best method of combining information on half-sib, full-sib, and self progenies and check entries in a blocked experiment for selecting among parents. An index or least squares weighting approach was utilized. Index solutions were found for two situations, one general with respect to the genetic and environmental model, and the other with simplifying assumptions about the model. In each situation a restricted index such that no block effects appear in any comparisons and an index based on comparisons of only genetic entries within blocks were found. Selection gains for these six indexes plus an additional *ad hoc* one were compared in a numerical example for two characteristics in tobacco, assuming the estimated variances for the two situations to be parametric.

I. Introduction

A twofold purpose is accomplished if selection experiments are designed such that they are also amenable to the estimation of genetic and environmental parameters. Such a design was described previously (Matzinger and Cockerham 1963) which involves half-sib and full-sib progenies from biparental matings and also self progenies from the same parents. The previous paper was concerned with estimation procedures while the present one is concerned with selection procedures.

The best parents, or remnant self seed from the best parents, are to be selected and recombined for the next generation. The problem arises as to how information on biparental and self progenies should be combined to accomplish this objective. The problem is further augmented by the fact that several groups of biparental and self progenies from independent sets of parents are grown in different experimental blocks and with all blocks having the same check entries. A unified approach to the problem is presented which is very similar in concept to that by Lush (1947) of weighting individual and family information and to that of the selection index approach as given by Henderson (1963).

* Part II, *Biometrics*, 1966, in press.

† Paper No. 2132 of the Journal Series of the North Carolina Agricultural Experiment Station. This investigation was supported in part by Public Health Service Research Grant GM 11546 from the Division of General Medical Sciences. The support for computing was provided by the National Institutes of Health Grant No. FR 00011.

‡ Departments of Experimental Statistics and Genetics, North Carolina State University at Raleigh.

II. Description of Experiment

While the design of the experiment has been described previously, it will be reviewed briefly to avoid cross-referencing. An experimental block of material consists of m^2 full-sib families, $2m$ self families, and c plots of different check materials, all replicated r times. The m^2 full-sib families are from all possible matings of m parents used as males to another m parents used as females and the self families are from the same $2m$ parents. All together, d blocks of material are considered, each from entirely different sets of parents. The check entries, however, are the same in all blocks. In the experiment described previously, m was 4 and the number of checks was 4, consisting of two parents of an original cross and their F_1 and F_2 generation. The parents of the families were random members of the F_2 generation.

III. Procedure

(a) *General*

Let the true value of a parent of the families be designated as G_{ijk}. The ijk subscript notation uniquely designates a parent of the full-sib families in the kth block as the ith individual used as the jth sex. Various means of the experimental material with information about G_{ijk} and to be utilized in the selection procedure are set out below:

y_{ijk} = the mean of the m full-sib families averaged over replications from the ij parent in the kth block.

x_{ijk} = the mean of the single self progeny over replications from the ij parent in the kth block.

$y_{.jk} = \sum_i y_{ijk}/m = y_{..k}$ = the mean of all biparental progenies in the kth block.

$x_{.jk} = \sum_i x_{ijk}/m$ = the mean of the self progenies from the m parents used as the jth sex in the kth block.

$x_{..k} = \frac{1}{2}(x_{.1k}+x_{.2k})$ = the mean of all self progenies in the kth block.

$z_{.k}$ = the mean of all checks in the kth block.

$y_{...}$ = the mean of all biparental progenies in the experiment.

$x_{...}$ = the mean of all self progenies in the experiment.

$z_{..}$ = the mean of all checks in the experiment.

The progeny means y_{ijk}, x_{ijk}, and $x_{.jk}$ provide information about their respective parents. The block means $x_{..k}$ and $y_{..k}$ are pertinent to the comparison of parents in different blocks. They contain genetic effects for parents in the blocks and also environmental effects peculiar to blocks which need to be assessed and discounted. This assessment of block environmental effects is the only function of $z_{.k}$. The overall means $y_{...}$, $x_{...}$, and $z_{..}$ are constants in the comparing or ranking of parents but are useful in simplifying the formulations and certain solutions. There are six not wholly dependent linear functions of these nine types of means. The set of six which appears to be least correlated is utilized in the following prediction equation of the parental true value:

$$\bar{G}_{ijk} = B_1(y_{ijk}-y_{..k})+B_2(x_{ijk}-x_{.jk})+B_3(x_{.jk}-x_{..k}) \\ +B_4(y_{..k}-y_{...})+B_5(x_{..k}-x_{...})+B_6(z_{.k}-z_{..}), \tag{1}$$

or in matrix notation

$$g = \mathbf{Y}'\mathbf{B},$$

where

$$g = \bar{G}_{ijk}, \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} y_{ijk}-y_{..k} \\ x_{ijk}-x_{.jk} \\ x_{.jk}-x_{..k} \\ y_{..k}-y_{...} \\ x_{..k}-x_{...} \\ z_{.k}-z_{..} \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \\ B_5 \\ B_6 \end{bmatrix}. \tag{2}$$

Required to find the set of B's which maximizes the correlation between the $\bar{G}$'s and G's, or discriminates among parents for maximum gain from selection, are the variance–covariance matrix of the Y's,

$$V = E[\mathbf{YY}'],$$

and the covariances of G with the Y's,

$$\mathbf{C} = E[g\mathbf{Y}],$$

where E denotes expectation.

$$V = \begin{bmatrix} \frac{m-1}{m}N_{23} & \frac{m-1}{m}N_{33} & 0 & 0 & 0 & 0 \\ \frac{m-1}{m}N_{33} & \frac{m-1}{m}N_{13} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2m}N_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{d-1}{d}N_{21} & \frac{d-1}{d}N_{31} & \frac{d-1}{d}N_{61} \\ 0 & 0 & 0 & \frac{d-1}{d}N_{31} & \frac{d-1}{d}N_{11} & \frac{d-1}{d}N_{51} \\ 0 & 0 & 0 & \frac{d-1}{d}N_{61} & \frac{d-1}{d}N_{51} & \frac{d-1}{d}N_{41} \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} \frac{(m-1)}{m}(2\sigma^2_{2p}) \\ \frac{(m-1)}{m}(2\sigma_{sm}) \\ \frac{1}{2m}(2\sigma_{sm}) \\ \frac{(d-1)}{dm}(2\sigma^2_{2p}) \\ \frac{(d-1)}{d2m}(2\sigma_{sm}) \\ 0 \end{bmatrix}. \tag{3}$$

The N's in the V matrix are summarized in Table 1 in the form of the expectations of variances and covariances of means corresponding to the mean squares and products in the analyses of variance of the experimental data.

The true value, G, of the parent is defined as its breeding value (Falconer 1960) in biparental progenies, since the selected parents are to be intermated to produce the next generation. It is twice the effect of the parent in the biparental progenies which accounts for the factor of 2 in each of the components in the C's.

TABLE 1

SUMMARY OF ANALYSES PERTINENT TO SELECTION OF PARENTS OR SELF PROGENIES

Source	Degrees of Freedom	Expectations of Variance or Covariance of Means
		Self Progenies
Blocks	$d-1$	$\frac{M_{11}^{*}}{2rm} = N_{11} = \frac{\sigma_{1e}^2}{2rm} + \frac{\sigma_{1s}^2}{2m} + \frac{\sigma_{1r}^2}{r} + \sigma_{1d}^2$
Self/blocks	$d(2m-1)$	$\frac{M_{13}}{r} = N_{13} = \frac{\sigma_{1e}^2}{r} + \sigma_{1s}^2$
		Biparental Progenies
Blocks	$d-1$	$\frac{M_{21}}{rm^2} = N_{21} = \frac{\sigma_{2e}^2}{rm^2} + \frac{\sigma_{2mf}^2}{m^2} + \frac{2\sigma_{2p}^2}{m} + \frac{\sigma_{2r}^2}{r} + \sigma_{2d}^2$
Parents/blocks	$2d(m-1)$	$\frac{M_{23}}{rm} = N_{23} = \frac{\sigma_{2e}^2}{rm} + \frac{\sigma_{2mf}^2}{m} + \sigma_{2p}^2$
		Self Progenies $\times$ Biparental Progenies
Blocks	$d-1$	$N_{31} = \frac{\sigma_{sm}}{m} + \frac{\sigma_{12r}}{r} + \sigma_{12d}$
Parents/blocks	$2d(m-1)$	$\frac{M_{33}}{m} = N_{33} = \sigma_{sm}$
		Checks
Blocks	$d-1$	$N_{41} = \frac{\sigma_{4e}^2}{rc} + \frac{\sigma_{4cd}^2}{c} + \frac{\sigma_{4r}^2}{r} + \sigma_{4d}^2$
Checks	$c-1$	$N_{42} = \frac{\sigma_{4e}^2}{rd} + \frac{\sigma_{4cd}^2}{d} + \sigma_{4c}^2$
Blocks $\times$ checks	$(d-1)(c-1)$	$N_{43} = \frac{\sigma_{4e}^2}{r} + \sigma_{4cd}^2$
		Checks $\times$ Self Progenies
Blocks	$d-1$	$N_{51} = \frac{\sigma_{14r}}{r} + \sigma_{14d}$
		Checks $\times$ Biparental Progenies
Blocks	$d-1$	$N_{61} = \frac{\sigma_{24r}}{r} + \sigma_{24d}$

* The M's are expectations of mean squares or mean products given in a previous publication (Matzinger and Cockerham 1963) for $m = 4$ and discussed there in more detail.

The B's which maximize selection gain are solutions to the equations

$$V\mathbf{B} = \mathbf{C},$$

and

$$\mathbf{B} = V^{-1}\mathbf{C}.$$

The gain from selection for an arbitrary set of B's and based on normal theory is expected to be

$$\Delta = k\mathbf{B}'\mathbf{C}/(\mathbf{B}'V\mathbf{B})^{\frac{1}{2}}, \tag{4}$$

where k is the unit selection differential corresponding to a given intensity of selection. For the B's which maximize the gain the formulation simplifies to

$$\Delta = k(\mathbf{B}'\mathbf{C})^{\frac{1}{2}}. \tag{5}$$

It may be noted that V^{-1} involves only the inverses of the 2×2, 1×1, and 3×3 matrices.

Estimates of B's may be found by substituting estimates for the N's in V. From the work of Williams (1962) and Patel (1962) these may lead to poorer estimates than if some judgment is exercised and restrictions are placed on the solutions.

(b) Simplification

Several simplifications in the model are introduced if certain conditions can be met. The first of these to be considered is the interaction of genetic entries with blocks. The only direct information on this interaction in the experiment is from the check material in which a check by block interaction component of variance, σ^2_{4cd}, may be tested. If this is zero, and if genetic entries do not interact with environmental blocks in general,

$$\sigma^2_{4cd} = 0,$$

$$\frac{\sigma^2_{1r}}{r}+\sigma^2_{1d} = \frac{\sigma^2_{2r}}{r}+\sigma^2_{2d} = \frac{\sigma_{12r}}{r}+\sigma_{12d} = \frac{\sigma^2_{4r}}{r}+\sigma^2_{4d} = \frac{\sigma_{14r}}{r}+\sigma_{14d} = \frac{\sigma_{24r}}{r}+\sigma_{24d} = \frac{\sigma^2_{r}}{r}+\sigma^2_{d}. \tag{6}$$

This would mean that the environmental portions of the error variances are the same also. While the genetic portions of the error variance will differ with the kind of genetic entries, they will be largely averaged out in the variance of plot means for plots of 10 plants or more. In such case, with small discrepancies,

$$\sigma^2_{1e} = \sigma^2_{2e} = \sigma^2_{4e} = \sigma^2_{e}. \tag{7}$$

Further simplifications depend upon the genetic situation. If gene effects are entirely additive, as is often indicated for normally self-fertilizing species, then

$$\begin{aligned} \sigma^2_{1s} = 2\sigma_{sm} = 4\sigma^2_{2p} &= \sigma^2_{A}, \\ \sigma^2_{2mf} &= 0. \end{aligned} \tag{8}$$

With these simplifying assumptions, only the relative values of σ^2_e, σ^2_A, and $[(\sigma^2_r/r)+\sigma^2_d]$ are needed. Making use of the intra-block heritability,

$$h = \frac{\sigma^2_A}{[(\sigma^2_e/r)+\sigma^2_A]},$$

and the ratio of intra- to inter-block environmental components,

$$b = \frac{\sigma^2_e/r}{[(\sigma^2_r/r)+\sigma^2_d]},$$

the relative weights are

$$\mathbf{B}^s = \begin{bmatrix} B_1^s \\ B_2^s \\ B_3^s \\ B_4^s \\ B_5^s \\ B_6^s \end{bmatrix} = \begin{bmatrix} 2mh/(4+mh) \\ 2B_1^s/m \\ h \\ mh/[2+mh+2m(m+2)(1-h)/(c+b)] \\ 2B_4^s/m \\ -(B_4^s+B_5^s)c/(c+b) \end{bmatrix}. \tag{9}$$

The effect of increasing the number, c, of checks is to increase B_4 and B_5, although complicatedly, and to give a greater negative weight or correction to the check block means.

(c) *Inter-block Restriction*

Less information is available for the block component of variance than any of the others. Too, blocks are often set out to account for major soil differences and, as such, do not represent a sample of random effects with common variances. When blocks are considered fixed, or for other reasons it is thought best to remove block effects entirely from the comparisons, this is accomplished by setting

$$B_6 = -(B_4+B_5),$$

or of utilizing (Y_4-Y_6) and (Y_5-Y_6) for the inter-block comparisons.

The solutions which maximize the gain for this restricted situation are

$$\begin{bmatrix} B_4^r \\ \\ B_5^r \end{bmatrix} = \begin{bmatrix} \frac{d-1}{d}(N_{21}-2N_{61}+N_{41}) & \frac{d-1}{d}(N_{31}-N_{51}-N_{61}+N_{41}) \\ \\ \frac{d-1}{d}(N_{31}-N_{51}-N_{61}+N_{41}) & \frac{d-1}{d}(N_{11}-2N_{51}+N_{41}) \end{bmatrix}^{-1} \begin{bmatrix} \frac{(d-1)}{dm}(2\sigma_{2p}^2) \\ \\ \frac{(d-1)}{d2m}(2\sigma_{sm}) \end{bmatrix}, \tag{10}$$

and

$$B_6^r = -(B_4^r+B_5^r).$$

The intra-block weights, B_1, B_2, and B_3, are not affected by inter-block restrictions.

For the simplified model, i.e. when the conditions outlined in equations (6), (7), and (8) hold, explicit expressions for inter-block weights, restricted such that block effects are removed, are

$$\begin{bmatrix} B_4^{sr} \\ B_5^{sr} \\ B_6^{sr} \end{bmatrix} = \begin{bmatrix} mh/[2+mh+2m(m+2)(1-h)/c] \\ 2B_4^{sr}/m \\ -(B_4^{sr}+B_5^{sr}) \end{bmatrix}. \tag{11}$$

Without check entries in the design, and to eliminate block effects, requires $B_4 = B_5 = B_6 = 0$, and only intra-block deviations of the genetic entries are utilized for either the general or simplified model.

(d) *Selection*

Whatever weights are used, it is not necessary to perform all the computations indicated for the $\bar{G}_{ijk}$'s in equation (1) in order to make the selections. By rearrangement,

$$\begin{aligned} \bar{G}_{ijk} = {} & B_1 y_{ijk} + B_2 x_{ijk} + (B_3 - B_2) x_{.jk} \\ & + \{(B_4 - B_1) y_{..k} + (B_5 - B_3) x_{..k} + B_6 z_{.k}\} \\ & - [B_4 y_{...} + B_5 x_{...} + B_6 z_{..}]. \end{aligned} \tag{12}$$

The term in the braces is constant for all members of a block and the term in the square brackets can be omitted since it is constant for the experiment.

The actual selection differential, $\bar{g}_s - \bar{g}$, which equals the mean of the $\bar{G}$'s of the selected group minus the mean of all the $\bar{G}$'s, is sometimes used in the prediction of gain from selection in the form

$$\Delta = (\bar{g}_s - \bar{g})\mathbf{B}'\mathbf{C}/\mathbf{B}'V\mathbf{B}, \tag{13}$$

instead of equation (4). For the parametric V, $\mathbf{C}$, $\mathbf{B}$, the two forms give the same results on the average since

$$E(\bar{g}_s - \bar{g}) = k(\mathbf{B}'V\mathbf{B})^{\frac{1}{2}}.$$

In practice, where estimates must be used for $\mathbf{B}$ and V, it has not been clarified which form is the most accurate.

TABLE 2

ESTIMATES OF COMPONENTS OF VARIANCE AND COVARIANCE FOR PERCENTAGE ALKALOID AND YIELD IN A POPULATION OF TOBACCO

	Alkaloid	Yield		Alkaloid	Yield
$\hat{\sigma}^2_{1e}$	0·035001	18,371·35	$\hat{\sigma}^2_{1r}/2+\hat{\sigma}^2_{1d}$	0·063015	15,273·55
$\hat{\sigma}^2_{2e}$	0·037216	13,628·04	$\hat{\sigma}^2_{2r}/2+\hat{\sigma}^2_{2d}$	0·067605	5,827·08
$\hat{\sigma}^2_{4e}$	0·034313	9,896·36	$\hat{\sigma}_{12r}/2+\hat{\sigma}_{12d}$	0·064535	10,256·10
$\hat{\sigma}^2_{1s}$	0·028555	13,064·35	$\hat{\sigma}^2_{4r}/2+\hat{\sigma}^2_{4d}$	0·051654	15,178·35
$2\hat{\sigma}_{sm}$	0·025286	16,805·48	$\hat{\sigma}_{14r}/2+\hat{\sigma}_{14d}$	0·057560	16,764·86
$4\hat{\sigma}^2_{2p}$	0·023438	24,905·17	$\hat{\sigma}_{24r}/2+\hat{\sigma}_{24d}$	0·060840	11,923·80
$\hat{\sigma}^2_{2mf}$	0·000323	158·98	$\hat{\sigma}^2_{4cd}$	−0·001323	2,457·10
$\hat{\sigma}^2_e$	0·036248	14,547·82	$\hat{\sigma}^2_r/2+\hat{\sigma}^2_d$	0·060868	12,537·29
$\hat{\sigma}^2_A$	0·021487	28,124·92	$\hat{b}$	0·297759	0·580182
$\hat{h}$	0·542450	0·794516			

IV. NUMERICAL EXAMPLE

To be used for the purpose of illustration are the results of an experiment with *Nicotiana tabacum* for which $m = 4$, $r = 2$, $c = 4$, and $d = 13$ for percentage of total alkaloid, and $d = 14$ for yield of cured leaf. Estimates of all the components of variance and covariance appearing in Table 1 are given above the dashed line in Table 2. The components of variance below the dashed line are averages. The intra-block error variance, $\hat{\sigma}^2_e$, is a pooled average (by pooling sums of squares) of the three separate intra-block errors. The additive variance, $\hat{\sigma}^2_A$, is a least squares estimate obtained as outlined in Matzinger and Cockerham (1963). The block environmental variance, $\frac{1}{2}\hat{\sigma}^2_r + \hat{\sigma}^2_d$, is the arithmetic average of the six corresponding terms above the dashed line in Table 2.

For the general model the estimated variance–covariance matrix ($\hat{V}$) and $\hat{\mathbf{C}}$ are constructed by substituting the estimates of components of variance and covariance above the dashed line in Table 2. Ordinarily, the mean squares would be used directly to construct $\hat{V}$, but these have not been given. For the simplified model the conditions of equations (6), (7), and (8) are imposed and $\hat{V}^s$ and $\hat{\mathbf{C}}^s$ are constructed by substituting the components of variance below the dashed line in Table 2. The index weights which maximize the expected gain for the general model and the simplified model are, respectively,

$$\hat{\mathbf{B}} = \hat{V}^{-1}\hat{\mathbf{C}},$$

$$\hat{\mathbf{B}}^s = (\hat{V}^s)^{-1}\hat{\mathbf{C}}^s.$$

TABLE 3

INDEX WEIGHTS

	$\hat{\mathbf{B}}$	$\hat{\mathbf{B}}^r$	$\hat{\mathbf{B}}^i$	$\hat{\mathbf{B}}^s$	$\hat{\mathbf{B}}^{sr}$	$\hat{\mathbf{B}}^{si}$	$\mathbf{B}^a$
				Alkaloid			
B_1	0·670892	0·670892	0·670892	0·703362	0·703362	0·703362	$\frac{2}{3}$
B_2	0·364867	0·364867	0·364867	0·351681	0·351681	0·351681	$\frac{1}{3}$
B_3	0·549027	0·549027	0·549027	0·542450	0·542450	0·542450	$\frac{1}{3}$
B_4	0·478141	0·411619	0	0·233815	0·224608	0	$\frac{1}{3}$
B_5	0·098043	0·123080	0	0·116908	0·112304	0	$\frac{1}{6}$
B_6	−0·624568	−0·534699	0	−0·326424	−0·336912	0	$-\frac{1}{2}$
				Yield			
B_1	1·273076	1·273076	1·273076	0·885493	0·885493	0·885493	$\frac{2}{3}$
B_2	0·274523	0·274523	0·274523	0·442746	0·442746	0·442746	$\frac{1}{3}$
B_3	0·755302	0·755302	0·755302	0·794516	0·794516	0·794516	$\frac{1}{3}$
B_4	2·914769	0·966668	0	0·433479	0·415766	0	$\frac{1}{3}$
B_5	0·192315	0·917274	0	0·216740	0·207883	0	$\frac{1}{6}$
B_6	−2·230181	−1·883942	0	−0·567854	−0·623649	0	$-\frac{1}{2}$

Alternatively, $\hat{\mathbf{B}}^s$ can be found by direct substitution in equation (9). The restricted sets of weights, $\hat{\mathbf{B}}^r$ and $\hat{\mathbf{B}}^{sr}$, for the general and simplified model, respectively, are found as outlined in equations (10) and (11). For further restriction to intra-block selection only, the weights $\hat{\mathbf{B}}^i$ and $\hat{\mathbf{B}}^{si}$ are given by setting the last three in $\hat{\mathbf{B}}$ and $\hat{\mathbf{B}}^s$, respectively, to zero. These six sets of weights are given in Table 3 along with a seventh *ad hoc* set, $\mathbf{B}^a$, arrived at by reasoning as to the relative merits of averaging environmental effects for different sized progenies and of the various genetic relationships and such that block effects are eliminated among comparisons of entries In terms of m the *ad hoc* weights are

$$\mathbf{B}^{a'} = \left[\frac{m}{2+m} \quad \frac{2}{2+m} \quad \frac{2}{2+m} \quad \frac{m}{2(2+m)} \quad \frac{1}{(2+m)} \quad -\tfrac{1}{2}\right].$$

The large number of digits given in Tables 2 and 3 is required for computational checking.

In Table 4 are the expected gains for each of these indexes assuming that $\hat{V}$ is the true parametric variance–covariance matrix and alternatively assuming that $\hat{V}^s$ is the true matrix. In computing the expected gains the assumed true matrix and corresponding $\mathbf{C}$ are always used in equation (4) which will reduce to the simpler form (5) when the index weights are the maximizing solution, i.e. $\hat{\mathbf{B}}$ for $\hat{V}$, $\hat{\mathbf{C}}$, and $\hat{\mathbf{B}}^s$ for $\hat{V}^s$, $\hat{\mathbf{C}}^s$, and the gains are a maximum in these two cases. The simpler form (5) will also work for the restricted solutions, $\hat{\mathbf{B}}^r$ and $\hat{\mathbf{B}}^i$ for $\hat{V}$, $\hat{\mathbf{C}}$, and $\hat{\mathbf{B}}^{sr}$ and $\hat{\mathbf{B}}^{si}$ for $\hat{V}^s$, $\hat{\mathbf{C}}^s$. For all expected gains k is taken to be $1 \cdot 75$, corresponding to an intensity of selection of 10%.

TABLE 4

EXPECTED GAINS FOR THE INDEXES ASSUMING TWO TRUE PARAMETRIC V's, $\mathbf{C}$'s

Parameter	Index Weights						
	$\hat{\mathbf{B}}$	$\hat{\mathbf{B}}^r$	$\hat{\mathbf{B}}^i$	$\hat{\mathbf{B}}^s$	$\hat{\mathbf{B}}^{sr}$	$\hat{\mathbf{B}}^{si}$	$\mathbf{B}^a$
	Alkaloid (%)						
True V, $\mathbf{C}$ $\hat{V}$, $\hat{\mathbf{C}}$	0·222	0·222	0·211	0·220	0·220	0·211	0·220
True V, $\mathbf{C}$ $\hat{V}^s$, $\hat{\mathbf{C}}^s$	0·199	0·202	0·198	0·204	0·204	0·198	0·201
	Yield (*lb*)						
True V, $\mathbf{C}$ $\hat{V}$, $\hat{\mathbf{C}}$	281	257	228	245	243	224	243
True V, $\mathbf{C}$ $\hat{V}^s$, $\hat{\mathbf{C}}^s$	204	241	254	269	268	256	265

V. DISCUSSION

The optimal selection strategy depends on parametric variances and covariances. The use of estimates of these parameters always leads to less than optimal gain from selection (Williams 1962), and a restricted solution to the weights may be closer to optimal on the average (Patel 1962). Unless the block component of variance is well estimated, and in particular if there are major environmental differences among blocks such that it is best to consider them to be fixed, then one of the restricted solutions, $\hat{\mathbf{B}}^r$ or $\hat{\mathbf{B}}^{sr}$ if checks are included or $\hat{\mathbf{B}}^i$ or $\hat{\mathbf{B}}^{si}$ without checks, should be utilized. By comparing equations (9) and (11), as the plot component of variance becomes large or b small, it can be seen that $\mathbf{B}^s$ approaches $\mathbf{B}^{sr}$ for the simplified model and $\mathbf{B}^{sr}$ is near optimal for the simplified situation even when the parametric variances are known.

The type of model, general or simplified, that is appropriate will depend on the genetic material. For the restricted solution the main assumption in the simplified model is additive gene action [equation (8)]. The analysis of the material (Matzinger and Cockerham 1963) furnishes information on the tenability of this assumption. When appropriate, the procedures outlined for the simplified model should always be used because fewer estimates are required and they are estimated with less variance from all of the data. On the other hand, for genetic materials which show considerable heterosis and non-additive gene action by analysis, the procedure for the general model, which is also general for the genetic model, may need to be followed.

In the numerical example the data (Table 2) fit very closely the assumptions of the simplified model for percentage alkaloid, as they do for many characteristics in many populations of tobacco. The fit to the simplified model is much poorer for yield from the standpoint of both genetic and environmental assumptions. However, the example is not as extreme in the lack of fit as one might find in naturally cross-fertilizing species which exhibit a large amount of heterosis.

There is no basis for comparing the expected gains for the two assumed true V's, $\mathbf{C}$'s, or between the two rows for each characteristic in Table 4. Rather, the comparisons of the gains of the procedures must be confined within rows bearing in mind that the maximum gain is from $\hat{\mathbf{B}}$ for $\hat{V}$, $\hat{\mathbf{C}}$, and from $\hat{\mathbf{B}}^s$ for $\hat{V}^s$, $\hat{\mathbf{C}}^s$. Needed for comparisons among the gains is V, $\mathbf{C}$ which is never known.

The maximum differences among the gains within a row are of the order of 5% for alkaloids and 20% for yield. This difference for the two characteristics is primarily a reflection of how well the variance estimates fit the assumptions of the simplified model. For alkaloid they fit very well, thus the $\hat{\mathbf{B}}$'s, $\hat{V}$'s, and $\hat{\mathbf{C}}$'s do not vary much and give much the same gains. For yield the variance estimates fit less well the assumptions of the simplified model leading to fairly large discrepancies among the expected gains.

Based on the studies of Williams (1962) and Patel (1962) there is little question but that one of the simplified solutions would be superior on the average for percentage alkaloid. The answer is not so clear for yield. The dilemma arises from the fact [Williams (1962), translated in terms of the present experiment] that the situations for which the general index gives the greatest gain are the ones for which the estimates have the largest variances which reduces the advantage of the general index in practice.

The *ad hoc* index ($\mathbf{B}^a$, Table 3) was actually used before the results of this paper were developed, and it is included for comparative purposes. It is a restricted solution, based on the simplified model, and corresponds very closely to $\hat{\mathbf{B}}^{sr}$. It is near optimal when the simplified model holds, has the advantage of simple computation, and does not suffer from errors of estimation. The loss in gain from using it instead of the true $\mathbf{B}^s$ or $\mathbf{B}^{sr}$ is only slight, and in practice it is probably more often superior to $\hat{\mathbf{B}}^s$ and $\hat{\mathbf{B}}^{sr}$ than not. While this cannot be said for the yield example with certainty, one can be fairly sure that it will not be much inferior.

Intra-block selection, $\hat{\mathbf{B}}^i$ for true $\hat{V}$, $\hat{\mathbf{C}}$, and $\hat{\mathbf{B}}^{si}$ for true $\hat{V}^s$, $\hat{\mathbf{C}}^s$, accounts for most of the total possible gain in Table 4. An indication of the effectiveness of utilizing information available on the check materials is given by comparing gains for $\hat{\mathbf{B}}^r$ with $\hat{\mathbf{B}}^i$ and for $\hat{\mathbf{B}}^{sr}$ with $\hat{\mathbf{B}}^{si}$. These results are in line with those of Schutz and Cockerham (1966). Intra-block selection is near optimal for appropriate sizes of blocks. If information is available on checks, it should be utilized and will increase slightly the selection gain. However, selection gains will generally be reduced slightly if checks are included at the expense of genetic entries.

VI. References

FALCONER, D. S. (1960).—"Introduction to Quantitative Genetics." (The Ronald Press Co.: New York.)

HENDERSON, C. R. (1963).—Selection index and expected genetic advance. In "Statistical Genetics and Plant Breeding." National Academy of Sciences—National Research Council Publ. No. 982. pp. 141–63.

LUSH, J. L. (1947).—Family merit and individual merit as bases for selection. Pts. I and II. *Am. Nat.* **81**: 241–61; 362–79.

MATZINGER, D. F., and COCKERHAM, C. C. (1963).—Simultaneous selfing and partial diallel test crossing. I. Estimation of genetic and environmental parameters. *Crop Sci.* **3**: 309–14.

PATEL, R. M. (1962).—Selection among factorially classified variables. Ph.D. Thesis, North Carolina State University at Raleigh.

SCHUTZ, W. M., and COCKERHAM, C. C. (1966).—The effect of field blocking on gain from selection. *Biometrics* (In press.)

WILLIAMS, J. S. (1962).—Some statistical properties of a genetic selection index. *Biometrika* **49**: 325–37.