THE EFFICACY OF EARLY OR INDIRECT TRUNCATED SELECTION

Fan H. Kung 1/

Abstract .--Early truncated selection causes a smaller selection differential for a given selection proportion and a smaller selection proportion for a given selection differential than the direct selection, the risk of misjudgement due to early or indirect selection depends on the magnitude of the juvenile-mature correlation and the selection intensity.

Additional keywords: correlation, selection differential.

INTRODUCTION

Tree breeding traditionally deals with long term projects. To shorten the delay and to improve efficiency, we often measure juvenile characters and correlate them with mature performance (Kung, 1973a). We also often measure easily obtainable characteristics which are indirect measures of the product that is sold and used. For example we may select for diameter outside bark at age 15, instead of volume inside bark at age 30. Obviously, the efficiency of indirect selection depends, in part, on the correlations and the selection intensities. These relationships are discussed in the paper.

CORRELATION AND SELECTION DIFFERENTIAL

The mathematical relationship between correlation and selection differential can be derived as follows by the regression model:

$$(\overline{Y}_i - \mu_y) = b_{y \cdot x}\,(\overline{X}_i - \mu_x) + e_i$$

Since

$$b_{y \cdot x} = \gamma\, \sigma_y/\sigma_x$$

we have

$$\overline{Y}_i - \mu_y = \gamma \frac{\sigma_y}{\sigma_x} (\overline{X}_i - \mu_x) + e_i \quad (1)$$

or

$$\frac{(\overline{Y}_i - \mu_y)}{\sigma_y} = \gamma \frac{(\overline{X}_i - \mu_x)}{\sigma_x} + e_i \quad (2)$$

1/ Assistant Professor, Department of Forestry, Southern Illinois University, Carbondale, Illinois 62901.

where $\overline{Y}_i$ = Mean of the selected group i of the dependent character (or character for direct selection).

$\overline{X}_i$ = Mean of the selected group i of the independent character (or character for indirect selection).

μ_y = The population average of the dependent character before selection.

μ_x = The population average of the independent character before selection.

σ_y = Standard deviation of the population of the dependent character.

σ_x = Standard deviation of the population of the independent character.

$b_{y \cdot x}$ = Regression coefficient (slope) of value Y due to X.

γ = Correlation coefficient between Y and X.

e_i = Error of estimation, with mean equal to 0.

Since $(Yi - \mu y)/\sigma y$ is the standardized selection differential of the desired character and $(Yi - \mu x)/\sigma x$ is the standardized selection differential of the selected character, one can consider the correlation as a reducing factor in standardized selection differential for indirect selection. For example, if the correlation between diameter outside bark and diameter inside bark is .95, and if the mean of the group selected by diameter outside bark is one standard deviation above the original population, the mean diameter inside bark of the selected group would be about .95 standard deviation higher than the original population.

If the actual rather than the standardized selection differential is desired, equation (1) may be used. Data from a study of the black walnut can be used to illustrate the use of the formula: given average height of black walnut trees at age 3, $\mu x = 7.1$ ft.; standard deviation of 3-year height, $\sigma x = 1.0$ ft.; average volume of 30-year old trees $\mu y = 16$ cu. ft.; standard deviation of 30-year volume $\sigma y = 7$ Cu. ft.; correlation coefficient between 3-year height and 30-year volume, $\gamma = .67$. If at age 3 the mean of the selected group is 9 ft. tall $(Xi = 9)$ this selected group at age 30 will be $\gamma(\sigma y/\sigma x)(Xi - \mu x)$ or .67 x 7 cu. ft. (9 ft. - 7.1 ft.)/1.0 ft. = 9 cu. ft. more than the population average. In other words, selecting trees with average height of 9 ft. at age 3 will give the same selection differential as selecting trees with average volume of 16 + 9 = 24 cubic feet at age 30.

It may be desirable to express selection differential as a percent of the mean. Substituting the product of the coefficient of variation (C.V.) and the mean (μ) for the standard deviation, equation (2) becomes:

$$\frac{Y_i - \mu_y}{\mu_y\ (C.V._y)} = \gamma\ \frac{(X_i - \mu_x)}{\mu_x (C.V._x)} + e_i$$

or

$$\frac{\text{Superiority of Y,\%}}{C.V._y,\%} = \gamma\left(\frac{\text{Superiority of X,\%}}{C.V._x,\%}\right) + e_i$$

Thus, $\text{Superiority of Y,\%} = \gamma\ \left(\frac{C.V._y}{C.V._x}\right)\ \text{Superiority of X,\%} + e_i$

Using the above example, we have C.V. for 3-year height and for 30-year volume as 14% and 44% respectively. The estimated superiority of 30-year volume for the selected group = (.67) (44%/14%) (Superiority of 3-year height) + ei = (.21) (Superiority of 3-year height) + ei. In other words, selecting a group of black walnut trees 10% taller than the average in 3-year height would have the same result as selecting a group of trees 21% larger than average in 30 year volume.

CORRELATION AND SELECTION PROPORTION

If direct selection is used, the corresponding values for selection proportion and selection differential can be obtained from tables (Nanson 1967, Namkoong and Snyder 1969) or by graphic method (Kung 1973b). For example, if the top half of the population is selected for diameter inside bark, the selection differential is .8 standard deviation above the mean. For indirect truncated selection, one may want to ask questions such as what is the selection proportion in diameter outside bark which will give a selection differential of .8 standardized score in diameter inside bark? Or: what is the proportion saved in diameter outside bark giving a selection differential equivalent to selecting 50% of the population based in diameter inside bark? To solve this problem, we must first compute the needed selection differential for diameter outside bark which is equal to the selection differential for diameter inside bark divided by the correlation coefficient (γ = .95). So we need .8/.95 = .84 standardized score for outside bark selection. The next step is to convert the selection differential to selection proportion which in this example is 0.47. Therefore, for the improvement of diameter inside bark, 50% selection by diameter inside bark is equivalent to 47% selection by diameter outside bark, when the correlation is .95. Other equivalent selection proportions between direct and indirect truncated selection at various correlation coefficients are shown in Table 1.

It can be seen that indirect selection is less effective when the correlation between the desired and the selected character is low. For example, correlation between 3-year and 30-year volume in black walnut trees is .3, so selecting only the top 1 percent in 3-year volume is the same as selecting half of the 30-year-old trees. In most finite populations

Table 1.--Equivalent selection proportion between direct selection and indirect selection for different levels of correlation

Percent Saved in Direct Selection	Percent of Population Saved for Indirect Selection when Correlation Coefficient is							
	.2	.3	.4	.5	.6	.7	.8	.9
1	---	---	---	---	---	.02	.11	.40
5	---	---	---	---	.07	.41	1.3	2.8
10	---	---	---	.10	.44	1.6	3.7	6.5
20	---	---	.06	.66	2.6	5.8	10	15
30	---	.01	.48	2.8	6.8	13	18	24
40	---	.16	2.1	6.8	13	21	27	34
50	---	1.0	5.8	14	23	30	38	44
60	.16	4.0	13	24	34	42	49	55
70	1.7	12	26	38	48	55	61	66
80	10	29	45	56	64	69	74	77
90	39	59	70	77	80	84	87	88

it would be almost impossible to practice indirect selection to achieve a 5 percent selection in the desired character when the correlation is below .5. Indirect selection may be used when the correlation is high and/or the selection proportion is low. For example, correlation between 30-year volume and 15-year volume in black walnut trees is .9; it may be economic to pay the price of thinning a few more trees so that the test period can be shortened. Thus, comparing the thinning of 56% (selection proportion 44%) of trees at age 15 to the, thinning of 50% of trees at age 30, the gain (15 years) may justify the cost (6% greater culling).

CORRELATION AND PROBABILITY OF IDENTICAL SELECTIONS

The effectiveness of indirect truncated selection may be judged by the probability of identical selections. Suppose that tree breeder A uses diameter inside bark to select the top 10% of the trees, and tree breeder B uses diameter outside bark to select the top 10% of the trees, what percent of the trees in two selection lists would be identical? The probability of identical selections, P(I.S.) can be expressed as follows:

$$
\begin{aligned}
P(\text{I.S.}) &= \frac{P(\text{individual chosen by both A and B})}{P(\text{individual chosen by A})} \\
&= \frac{P(\text{individual chosen by both A and B})}{P(\text{individual chosen by B})} \\
&= P(\text{individual chosen by B}) \text{ given that individual is chosen by A} \\
&= P(\text{individual chosen by A}) \text{ given that individual is chosen by B}
\end{aligned}
$$

If A and B are perfectly correlated, the probability of identical selections is 1 at any selection proportion. On the other hand, if all individuals are selected, the probability of identical selection is 1 for correlation. The probability can be 0 when a negative correlation exists and the selection proportion is less than half. By using computer simulation figure 1 was obtained. Curves were drawn using the average of ten runs for each correlation and level of selection proportions. Correlations coefficients used were from .1 to .9 with increment of .1. Selection proportions used were from .1 to .9 with increment of .05. The simulation program using OMNIIAB commands is available upon request.

The reader can use figure 1 to estimate whether an indirect selection is worth the risk of missing some superior trees. For example, correlation coefficient between 30-year and 15-year volume is .9. Therefore, of the top 20% of the trees by volume are selected at age 15, the probability of identical selection from selecting the top 20% by volume at age 30 is .75. In other words, one out of four (25%) of the top 20% of the trees at age 30 may be lost due to the early culling.

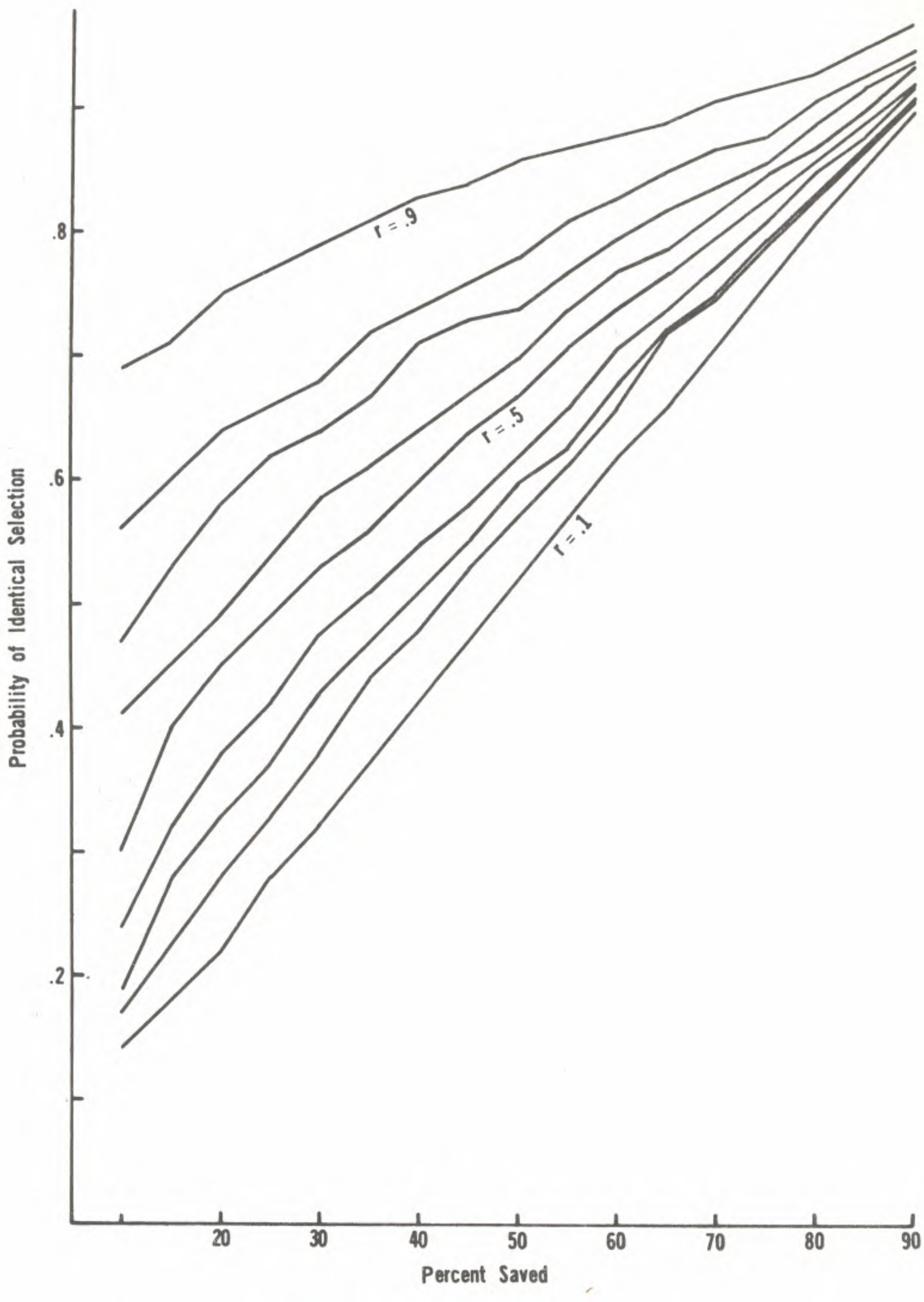

Figure 1.-- Probability of identical selection at various levels of selection proportion and correlation between the direct and the indirect character.

The probability of missing a true winner is greater when the correlation is low and when a small portion of the population is selected.

SUMMARY

Whether or not indirect truncated selection should be used depends on the magnitude of the correlation between the direct and the indirect character, as well as the intensity of selection. The desirability of indirect selection can be judged by (1) the ratio of selection differentials, (2) the ratio of selection proportion for a given standardized selection differential, and (3) the probability of identical selections. Indirect selection is usually inferior to direct selection in terms of selection differential, proportion saved, and the risk of misjudgement, because the correlation is seldom perfect. However, when the correlation is high and the culling is mild, indirect selection may be justified on economic grounds.

LITERATURE CITED

Kung, F. H. 1973a.
Development and use of juvenile - mature correlations in a black walnut tree improvement program.
South Conf. Forest Tree Impr. Proc. 12: 243-249.

Kung, F. H. 1973b.
A handbook of graphic solution to forest biometric problems.
Dept. of Forestry Publication No. 12. Dept. of Forestry Southern Illinois Univ. Carbondale, IL 89 pp.

Namkoong, G., and Snyder, E. B. 1969.
Accurate Values for selection intensities.
Silvae Genetica 18: 172-173.

Nanson, A. 1967.
Tables de la différentielle de sélection Bans la distribution normale (o, 1).
Biometrie Praximetrie. t VIII, No. 1 40-51.