The effect of errors in the economic weights on the accuracy of selection indexes [1]

W. M. VANDEPITTE and L. N. HAZEL

Centrum voor Landbouweconomisch Onderzoek, K.U. Leuven
Belgische Boerenbond, Minderbroederstraat 8, 3000 Leuven (Belgium)
Animal Science Dept., Iowa State University, Ames Iowa (U.S.A.)

Summary

The effects of errors in the economic weights on the efficiency of index selection were investigated. A selection index for the genetic improvement of pigs was used as test case.

Errors in single economic weights of $\pm$ 50 percent reduce the relative efficiency with less than 1 percent for all traits considered. Larger errors can result in considerable bias of the estimated genetic gain. The effect of errors in single economic weights are non linear and non symmetrical. Negative errors (underestimation) are in general more critical than positive errors (overestimation). This dissymmetry depends on the scale used to define the errors.

The effect of simultaneous sampling errors in the economic weight vector was studied by using both Monte-Carlo simulation and mathematical approximation. Result obtained by simulation indicate that the estimated genetic gain ($\Delta\widehat{H}$) and the realized genetic gain ($\Delta H!\widehat{I}$) are not normally distributed. For small sampling errors in the economic weight vector (coefficient of variation C.V $\leqslant$ 0.50) the loss in relative efficiency was less than 2.6 percent but increased to approximately 15 percent for C.V. = 1.0.

The loss in relative efficiency of a selection index, the biases and variances of $\Delta\widehat{H}$ and $\Delta H|\widehat{I}$ due to sampling errors in the economic weight vector are a function of the sampling variances and covariances of the economic weights and the genetic and phenotypic variance-covariance structure of the traits involved. $\Delta\widehat{H}$ is an overestimate of the maximum attainable genetic gain, while $\Delta H|\widehat{I}$ is an underestimate. The biases in $\Delta\widehat{H}$ and $\Delta H|\widehat{I}$ are small for small sampling errors, but increase rapidly when the sampling errors in the economic weight vector increase. The variance of $\Delta\widehat{H}$ is relatively much larger than the variance of $\Delta H|\widehat{I}$, and increasing sampling errors in the economic weight vector make variance of $\Delta\widehat{H}$ increase at a much faster rate than the variance of $\Delta H|\widehat{I}$.

Introduction

In order to construct a selection index, for multiple trait selection, estimates of genetic and economic parameters are required. These data may be obtained from different sources of information and following different computational

methods. However, the optimum response will only be obtained if precise para-
meter values are used. If the estimates are in error some efficiency will be lost.

This study deals with the effect of errors in the economic weights on the accu-
racy of index selection. The genetic parameters are treated as fixed constants. A
selection index for multiple trait selection in a pig population has been used as
exemple.

The results were obtained by a simulation approach and were verified by
mathematical approximations (using Taylor's series).

The construction of a selection index

The objective of selection will always be to affect population changes
in one or more traits which relate to the economic value of the members of the
population. Whatever these traits may be, one can always express the net
economic worth or net profit of an animal (W_j) as a function of these traits (x_1,
x_2, ..., x_n).

$$W_j = a_1x_{1j} + a_2x_{2j} + \ldots + a_nx_{nj} + e_j \qquad j = 1, \ldots, N \qquad (1)$$

Where $a_1, a_2, \ldots, a_n$ are the regression coefficients from the regression of net
profit (W_j) on the phenotypic values $x_1, x_2, \ldots, x_n$ and e_j is an error term. Notice
that $W_j = f(x_{ij})$ does not have to be linear in the $x_{ij}'s$. The estimated profi-
tability is (we will drop the j subscripts).

$$\widehat{W} = \widehat{a_1}x_1 + \widehat{a_2}x_2 + \ldots + \widehat{a_n}x_n. \qquad (2)$$

The $a_i's$ are the economic weights, defined by Hazel (1943) as "... the amount
by which net profit may be expected to increase for each unit of improvement in
that trait ". Under the additivity assumption $x = g + e$ (phenotypic value is
equal to an additive genetic component and an environmental component), we
can rewrite (2) as

$$\widehat{W} = \widehat{a_1}(g_1 + e_1) + \widehat{a_2}(g_2 + e_2) + \ldots + \widehat{a_n}(g_n + e_n). \qquad (3)$$

If we define

$$\widehat{H} = \widehat{a_1}g_1 + \widehat{a_2}g_2 + \ldots \widehat{a_n}g_n \qquad \text{and} \qquad (4)$$

$$\widehat{E} = \widehat{a_1}e_1 + \widehat{a_2}e_2 + \ldots \widehat{a_n}e_n$$

$\widehat{W}$ becomes

$$\widehat{W} = \widehat{H} + \widehat{E}.$$

Where $\widehat{H}$ is the complete average genetic net worth or complete aggregate
genotype and $\widehat{E}$ is the deviation from the complete average genetic net worth
associated with the particular environment in which the phenotype is observed.
$\widehat{H}$ is a linear function of n nonobservable variables and is the quantity to be im-
proved by selection. The most efficient selection criterion I (selection index)
is the one having the largest possible correlation with $\widehat{H}$ (Hazel, 1943). An obvious
index would be $\widehat{W}$ itself as illustrated by Manning (1956). However the x_i values

in $\widehat{W}$ are not always measurable on the individual. Or for practical reasons we may direct our attention to traits that have indirectly a unique significance with respect to yield or quality. For this reason an index of the form

$$I = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m \qquad (5)$$

can be considered, where x_1, x_2, ... x_m are representing objective measurements or numerical scores on the individual and or its relatives, and β_1, β_2, ..., β_m are constants obtained as solutions from a set of simultaneous equations.

The effect off errors in the economic weights

Relatively little research has been directed toward examining the problems and properties of economic weights. The estimation of economic weights is not very easy. Relative costs and prices may fluctuate periodically, or for some traits the necessary data are not available. For these reasons the estimates of economic weights are often rather approximate. It is therefore important to know to what extent they influence the index and the efficiency of it. Pease et al. (1967) have made some considerations on the effect of errors in the economic weights. The economic weights were not found to be very critical.

Extreme errors of as much as 50 percent resulted in the worst case in a loss of relative efficiency of 1.8 percent. Rønningen (1971) studied the effect of false economic ratio's between two traits on the change in aggregate genotype for a two-trait index. The criterion used was the difference between the correlation of the correct aggregate genotype and the correct index, and the correlation of the correct aggregate genotype with a biased index. Rønningen concluded that the loss in efficiency is not too serious when moderate deviations from the true economic ratio are used. The loss increased as the devaition from the true economic ratio increased. When the economically most important trait was given a negative weight, the loss was substantial, especially when the heritability was high.

The selection index theory assumes that the economic weights are known fixed constants. For various reasons this assumption is almost never fulfilled.

Economic weights are estimates

In the most favorable case, when we have complete information, multiple regression techniques can be used to determine the economic weights. In this case the economic weights are unbiased, but have usually fairly large sampling errors. For some traits, however, (i.e., loin-eye area and color score) the economic information is lacking or only partially available. In these circumstances economic weights are intelligent guesses rather than accurate estimates.

Fractionated structure of the industry

The industry is made up of many breeders, commercial producers, and packers. Every individual has his own ideas and goals, and each tends to work in the economic framework of his own enterprise. For this reason every breeder and

producer has his own set of economic weights. Because of the long-term nature of a breeding policy, the industry as a whole should have a well defined set of realistic goals.

Time dimension

Economic weights reflect production costs and consumer preference through the pricing mechanism. They are affected by price trends of feed grains, labor, construction costs, the qualitative and quantitative trends in the demand for pork meat and its substitutes and by technical innovations. Therefore the economic weights have to be adjusted periodically for changing economic situations and selection goals.

Improvement lag

The dissemination of additive genetic improvement through a multiple tier breeding structure (nucleus, multiplier and commercial herds) requires time. The extent to which each tier is genetically behind the previous one, has been termed the " improvement lag ". In order to use the proper economic weights to conform with consumer preference and state of technology at time t_0, we should use projected economic weights for time $(t_0 + \Delta t)$ where Δt is the total time lag between the genetic improvement in the nucleus herd and its appearance in the commercial herds. Since long term projections of price relationships and economic conditions in general are likely to have poor accuracy, frequent reevaluation of economic weights is indicated. Even then, sampling errors may influence the estimates actually chosen to establish breeding goals. This makes it evident that the " fixed constant " concept of economic weights is only relative. The important consequences of this are that by using a biased set of economic weights or by delaying the proper adjustments, the accuracy of selection will be reduced. In this section we will confine ourselves to the effect of variation and errors in the economic weights on index selection.

Concepts

A basic knowledge of how the genetic gain and the efficiency of a selection index is affected when the economic weights are biased and of how the relative efficiency or loss in relative efficiency will be measured is important for the understanding of what follows.

The gain in the aggregate genotype (ΔH) can be represented by:

$$\sum_{i=1}^{m} a_i \Delta G_i \quad \text{with} \quad \Delta G_i = R(IG_i)\sigma G_i i. \tag{6}$$

ΔG_i is the genetic gain in the ith trait, $R(IG_i)$ is the correlation between the index I and the genetic value of the ith trait, σG_i is the genetic standard deviation of the ith trait and i is the selection intensity,
or in matrix notation by

$$\Delta H = a' \Delta G$$

where a' is a $1 \times m$ vector of economic weights and ΔG is a $m \times 1$ vector of gains in the individual traits in metric units.

Assuming that all phenotypic and genetic parameters are known, we will denote ΔH, $\boldsymbol{a}$ and ΔG as the gain in the aggregate genotype, the economic weight vector and the vector of genetic gains in the individual traits when the economic weights are known without error. $\Delta\hat{H}$, $\hat{\boldsymbol{a}}$ and $\Delta\hat{G}$ are the corresponding parameters when the economic weights are biased.

We can express $\Delta\hat{G}$ and $\hat{\boldsymbol{a}}$ by

$$\Delta\hat{G} = \Delta G + (\Delta\hat{G} - \Delta G)$$
$$\hat{\boldsymbol{a}} = \boldsymbol{a} + (\hat{\boldsymbol{a}} - \boldsymbol{a})$$

and multiplication results in the following identities :

$$\hat{\boldsymbol{a}}'\Delta\hat{G} = \boldsymbol{a}'\Delta G + {}'(\hat{\boldsymbol{a}} - \boldsymbol{a})'\Delta G + \boldsymbol{a}'(\Delta\hat{G} - \Delta G) + (\hat{\boldsymbol{a}} - \boldsymbol{a})'(\Delta\hat{G} - \Delta G)$$
$$\hat{\boldsymbol{a}}'\Delta\hat{G} = \boldsymbol{a}'\Delta G + \boldsymbol{a}'(\Delta\hat{G} - \Delta G) + (\hat{\boldsymbol{a}} - \boldsymbol{a})'\Delta\hat{G} \tag{7}$$

Expression (7) clearly indicates that errors in economic weights affect the estimated genetic gain in two different ways.

(a) Indirectly, because the $\Delta\hat{G}$ vector is biased since ΔG_i is a linear function of $R(IG_i)$. This is the effect of biases in the economic weights on the genetic gain in the individual traits expressed in metric units.

(b) Directly, because in the third term of expression (7) $\Delta\hat{G}$ is multiplied by the vector of biases $(\hat{\boldsymbol{a}} - \boldsymbol{a})'$ of economic weights.

Three different estimates of genetic gain can be considered (selection intensity $= 1.0$).

$\Delta H = \sigma I$

$= (\boldsymbol{a}'G'P^{-1}Ga)^{1}/^{2}$ is the expected genetic gain in the real aggregate genotype H that results from selection on the corresponding real index I.

P is a $n \times n$ matric of phenotypic covariance between the n variables in index I and G is a $n \times m$ matrix of genotypic covariances between the n traits in I and the m traits in H.

$\Delta H \ \widehat{I} = \mathrm{Cov}\ (\widehat{HI})/\sigma\widehat{I}$

$= (\hat{\boldsymbol{a}}'G'P^{-1}Ga)/(\hat{\boldsymbol{a}}'G'P^{-1}G\hat{\boldsymbol{a}})^{1}/^{2}$ is the realized genetic gain in the real genotype when selection is practiced on a biased index (calculated with biased economic weights).

$\Delta\hat{H} = \sigma\hat{I}$

$= (\hat{\boldsymbol{a}}'G'P^{-1}G\hat{\boldsymbol{a}})^{1}/^{2}$ is the estimated genetic gain in the biased aggregate genotype resulting from selection on the corresponding biased index.

The relative efficiency (RE) of a selection index with respect to errors in the economic weights is

$$RE = 100\ \frac{R(\widehat{HI})}{R(HI)} = 100\ \frac{\mathrm{Cov}\ (\widehat{HI})}{\sigma I \sigma\widehat{I}} \quad \text{as } R(HI) = \frac{\sigma I}{\sigma H} \tag{8}$$

where $R(\widehat{HI})$ is the correlation between the real aggregate genotype with the biased index, and $R(HI)$ is the correlation between the real aggregate genotype and the unbiased index, while σI and σH are the standard deviations of index I and aggre-

gated breeding value H. The loss in relative efficiency due to errors in the economic weights is

$$LRE = \left(1 - \frac{R(\widehat{HI})}{R(HI)}\right) 100.$$

Data

To study the effect of errors in the economic weights a selection index was designed for a " Combined Testing " scheme for pigs. Four pigs from the same litter were to be tested, two boars in a performance-test (individual pens), and two full-sibs to obtain carcass information (in the same pen). The parameters used to construct the selection index are given in tables 1, 2 and 3.

TABLE 1

Summary of the economic weights in dollars

Pondérations économiques en dollars

Trait	Unit	Ec. weight	Stand dev.	Ec. stand dev.
Daily gain	lb.	4.50	0.16	0.720
Feed efficiency		— 5.80	0.25	— 1.450
Carcass backfat	in.	— 2.95	0.16	— 0.472
Dressing percentage	%	0.61	1.65	1.006
Percent ham and loin	%	0.46	1.77	0.814
Loin-eye area	sq. in.	0.64	0.62	0.397
Number of pigs weaned	#	1.86	2.18	4.055

TABLE 2

Genetic correlations (top triangle) and phenotypic correlations (bottom triangle) and heritabilities

Corrélations génétiques (au-dessus de la diagonale) et phénotypiques (au-dessous de la diagonale) et héritabilités

	1	2	3	4	5	6	7	8	h^2
1. Daily Gain		— 0.68	0,20	0.15	— 0.13	— 0.20	— 0.15	0.00	0.30
2. Feed eff.	— 0.62		0.38	0.38	— 0.15	— 0.35	— 0.13	0.00	0.35
3. B. F. Probe	0.15	0.35		1.00	0.15	— 0.45	— 0.13	0.00	0.45
4. B. F. Carcass	0.10	0.32	0.70		0.15	— 0.40	— 0.30	0.00	0.45
5. Dressing %	— 0.10	— 0.17	0.10	0.15		0.17	0.20	0.00	0.30
6. %ham + loin	— 0.15	— 0.35	— 0.40	— 0.45	0.10		0.60	0.00	0.45
7. Loin-eye	— 0.10	— 0.10	—0.20	— 0.30	0.20	0.54		0.00	0.55
8. # weaned	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.10

TABLE 3

List of traits included in index and aggregate genotype

Liste des caractères inclus dans l'index et valeur génétique globale

Trait	Number	Description of trait	$\widehat{\beta}$ Value
Index	Unit		
1	1b.	Daily gain of boar 1	— 0.140
2		Feed efficiency of boar 1	— 0.817
3	in.	Probe of boar 1	— 0.164
4	1b.	Daily gain of boar 2	— 0.024
5		Feed efficiency of boar 2	— 0.159
6	in.	Probe of boar 2.	— 0.014
7	in.	Average carcass backfat of barrow and gilt. . . .	— 0.036
8	%	Average dressing percentage of barrow and gilt .	0.193
9	%	Average percent ham and loin of barrow and gilt .	0.236
10	sq. in.	Average loin-eye of barrow and gilt	0.239
11	1b.	Average daily gain of barrow and gilt	0.167
12		Average feed efficiency of barrow and gilt. . . .	— 0.066
13	#	Number of pigs weaned	0.203
Aggregate genotype			
1′		Daily gain.	
2′		Feed efficiency.	
3′		Carcass backfat.	
4′		Dressing percentage.	
5′		Percentage ham and loin.	
6′		Loin-eye area.	
7′		Number of pigs weaned.	

The genetic and phenotypic parameters used to calculate the appropriate variances and covariances were compiled from the litterature. The sources were JONSSON (1965), FREDEEN (1953), PEASE *et al.* (1967), CHRISTIAN (1970), VINT (1971), SIERS and THOMSON (1972).

Errors in single economic weights

A. — *Effect on estimated genetic gain* $(\Delta\widehat{H})$

The change in $\Delta\overline{H}$ for a marginal change in the economic weights can be evaluated by differentiation with respect to the economic weight vector.

Second differentials of $\Delta\widehat{H}$ with respect to the economic weights indicate that the effect of marginal changes in the economic weights are neither linear nor additive.

The effect of errors in individual economic weights, errors ranging from minus 200 percent to plus 200 percent, were introduced in each economic weight separately and $\Delta\widehat{H}$ was calculated for each error-trait combination.

The effect on estimated genetic gain is plotted in figure 1. These figures indicate that errors in the economic weights can result in relatively important

under (negative errors) and over estimation (positive errors) of the genetic gain. For errors of minus 50 percent, the bias in estimated genetic gain ranges from minus 1.33 percent for number weaned to minus 22.91 percent for feed efficiency.
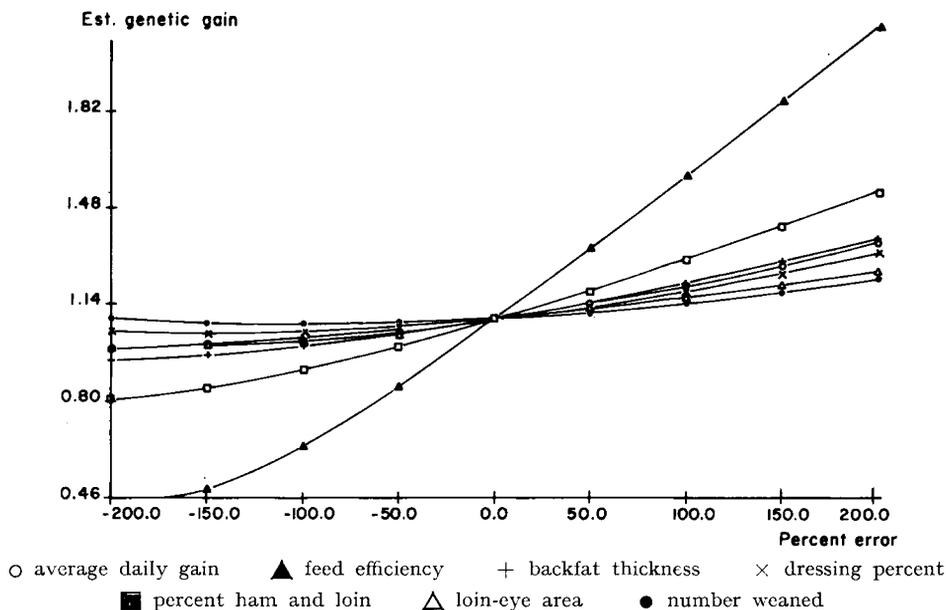

FIG. 1. — *Estimated genetic gain* ($\widehat{\Delta H}$), *in dollars per generation, when single economic weights are affected by errors ranging from minus 200 percent to plus 200 percent.*

Gain génétique estimé ($\widehat{\Delta H}$ *), en dollars par génération, en fonction d'erreurs allant de — 200 à + 200 p. 100 dans les pondérations économiques individuelles.*

For positive errors of 50 percent, the bias ranges from plus 2.15 percent for number weaned to plus 27.82 percent for feed efficiency. The magnitude of the bias is a function of the value of the affected economic weight, the heritability of the corresponding trait and the covariance structure of that trait with the other traits.

B. — *Effect on real genetic gain* ($\Delta H | \widehat{I}$)

A similar approach was used as for $\widehat{\Delta H}$. $\Delta H | \widehat{I}$ was calculated for each error-trait combination.

Errors ranging from minus 50 percent to plus 50 percent have relatively little effect on the real genetic gain. Over this range, the reduction in real genetic gain is in all cases less than 1 percent.

This is also reflected by the loss in relative efficiency (LRE). The loss in relative efficiency due to errors in the economic weights ranging from minus 200 percent to plus 200 percent is plotted in figure 2.
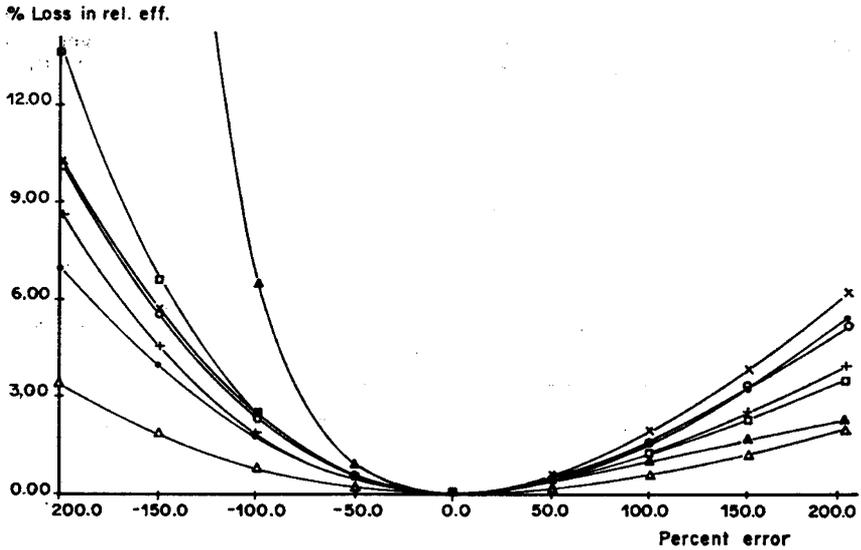
FIG. 2. — *Loss in relative efficiency (LRE) when single economic weights are affected by errors ranging from minus 200 percent to plus 200 percent.*

Perte d'efficacité relative en fonction des erreurs sur les pondérations économiques individuelles.

The results indicate that:

— the loss in relative efficiency of a selection index due to errors in single economic weights is not symmetrical,

— negàtive errors (underestimation of economic weights) are more critical than positive errors (overestimation of economic weights),

— the dissymmetry depends on the scale used to define the errors,

— for errors between minus and plus 50 percent, the loss in relative efficiency varies between 0.16 percent and 0.90 percent,

— larger errors (beyond the plus and minus 50 percent interval) can result in losses up to 10.14 percent and 76.44 percent for errors of minus 200 percent in the economic weights of daily gain and feed efficiency respectively.

Random errors in the economic weight vector

A. — *Simulation of random errors and index samples*

In cases where complete information is available, economic weights can be estimated by multiple regression technique and the variances and covariances of the regression coefficients (economic weights) can be calculated. Since this vital information could not be produced, Monte-Carlo simulation was used to study the effect of random errors in the economic weight vector on real and estimated genetic gain and the loss in relative efficiency.

Errors were introduced in the economic weight vector in the following way: $\hat{a}_i = a_i(1 + \varepsilon_i)$, where a_i is the ith element of the economic weight vector used previously and ε_i is a random error drawn from a distribution that is NID(0, σ^2). It is easily demonstrated that the new economic weights, $a_i(1 + \varepsilon_i)$, are unbiased and have variance $a_i^2\sigma\varepsilon^2$.

TABLE 4

Major distribution statistics of $\widehat{\Delta H}, \Delta H\widehat{|I}$ and LRE for four generated index samples
($N = 100$)

Paramètres des distributions de $\Delta H, \Delta H\widehat{|I}$ et de LRE pour 4 échantillons de 100 index
obtenus par simulation

Sampling error (σ_ε)	Exp. val. (a)	Variance (a)	Std. dev.	Min. val.	Max. val.	β_1 (b)	β_2 (c)
Estimated genetic gain ($\widehat{\Delta H}$) in $ per generation							
0.10	1.09085	0.00371	0.06093	0.97509	1.30509	0.21468	3.52099
	1.08800	0.00312					
0.20	1.09779	0.01483	0.12178	0.86549	1.52519	0.21245	3.51258
	1.09197	0.01237					
0.50	1.13562	0.09016	0.30027	0.56059	2.19149	0.22192	3.54848
	1.12168	0.07280					
1.00	1.26717	0.30964	0.55645	0.25929	3.30909	0.40114	3.87259
	1.24036	0.24727					
Real genetic gain ($\Delta H|\widehat{I}$) in $ per generation							
0.10	1,08529	0.00000	0.00132	1.07999	1.08669	4.07848	8.17399
	1.08538	0.00000					
0.20	1.08115	0.00002	0.00453	1.06199	1.08669	3.91379	7.74180
	1.08135	0.00009					
0.50	1.05008	0.00110	0.03323	0.88799	1.08659	4.82347	9.19628
	1.04990	0.00343					
1.00	0.90005	0.05137	0.22665	— 0.10189	1.08449	4.73826	7.72679
	0.92472	0.04269					
Loss in relative efficiency (LRE) in %							
0.10	0.12738	0.01087	0.10427	0.00059	0.61979	4.19367	8.32184
	0.12059	0.04972					
0.20	0.50805	0.17409	0.41724	0.00089	2.27009	3.92706	7.75134
	0.49146	0.78966					
0.50	2.61190	10.82238	3.28974	0.00105	18.28739	4.21573	8.54555
	3.37830	29.09834					
1.00	17.17443	435.05249	20.85791	0.20559	109.37959	4.73878	7.72721
	14.90486	361.55920					

(a) The upper value of each pair is the simulated estimate, while the lower value is derived from the approximate equation.

Chaque couple de valeur est obtenu par simulation (au-dessus) ou par approximation mathématique (au-dessous).

(b) β_1 measures skewness, for the normal distribution $\beta_1 = 0$; β_1 = coefficient d'asymétrie.

(c) β_2 measures kurtosis, for the normal distribution $\beta_2 = 3$; β_2 = coefficient d'aplatissement.

Four random samples (with sample size N = 100) of economic weight vectors were generated. The random errors were drawn from normal distributions with standard deviations ($\sigma\varepsilon$) of 0.1; 0.2; 0.5 and 1.0; respectively. Each sample of economic weight vectors resulted in 100 selection indexes. The major distribution statistics of each sample were calculated for $\Delta\hat{H}$, $\Delta H|\hat{I}$, LRE, and are listed in Table 4. The value of ΔH was equal to $ 1.086 per generation.
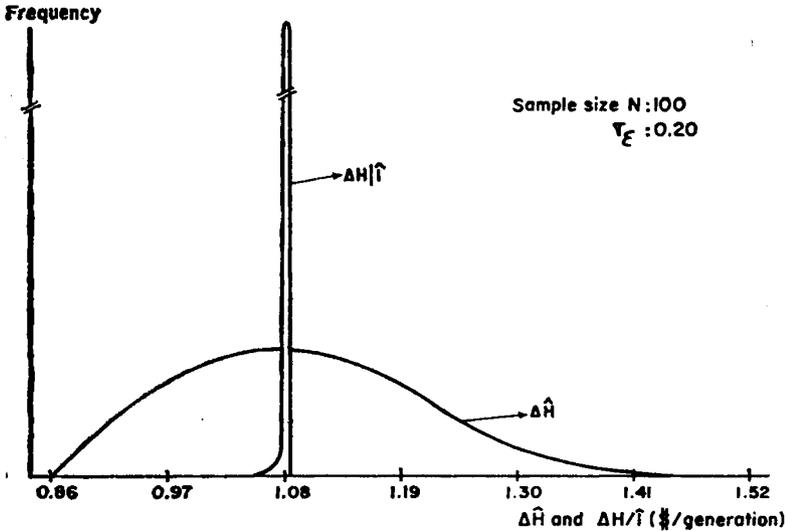

FIG. 3. — *Distribution of $\Delta\hat{H}$ and $\Delta H|\hat{I}$.*

Distribution de $\Delta\hat{H}$ et $\Delta H|\hat{I}$.

The β_1 and β_2 statistics calculated from the simulated samples indicate that ΔH, $\Delta H|\hat{I}$ and LRE are not normally distributed. To illustrate the shape, the distribution of $\Delta\hat{H}$, $\Delta H|I$ and LRE are plotted in figures 3 and 4.

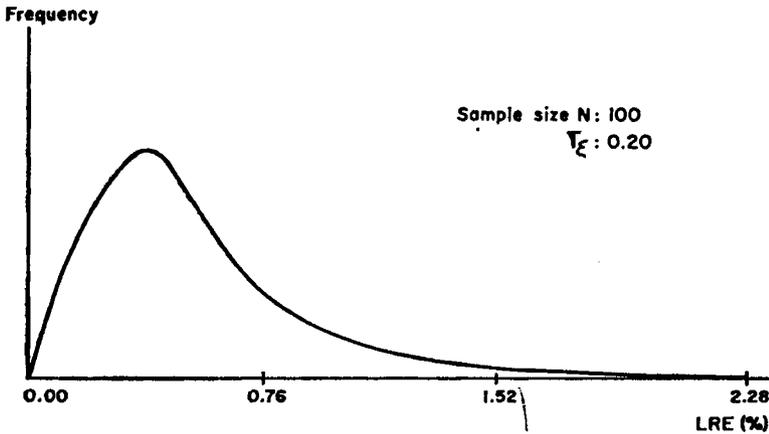FIG. 4. — *Distribution of LRE.*

Distribution de LRE.

B. — *Derivations of expectations and variances*

To relate the errors in the estimated economic weights to the variances and expectations of $\Delta\widehat{H}$, $\Delta H|\widehat{I}$ and LRE and to get a better insight in the mathematics involved, an attempt was made to derive expectations and variances for the generated distributions.

Expectations and variances of complex functions of the general form $(x'Ax)^{1}/^{2}$, $\dfrac{x'Ay}{(x'Ax)^{1}/^{2}}$ (where x and y are vectors of random variables and A is a matrix of constants) were derived by approximation as suggested by KENDALL and STUART (1958) by taking the first terms of a Taylor's expansion. A similar approach was used by HARRIS (1963) and SALES and HILL (1976). The accuracy of these approximations depends on the smallness of the higher order terms of the expansion which are truncated. This method is only useful in cases where the Taylor series is converging at a relatively fast rate.

Approximations for the expected values and variances of $\Delta\widehat{H}$, $\Delta H|\widehat{I}$ and LRE, have been derived. We further assumed that the genetic parameters and the economic weights in the correct aggregate genotype are known constants, and that the economic weights have the same error structure $(a_i = a_i (1 + \varepsilon_i))$ as the one used for the simulation.

Discussion

The agreement between the numerical values obtained from the approximate equations and those from the Monte-Carlo simulation is not perfect. However, both series of estimates follow the same directional trends. The simulated and approximated numerical values of the expectations of $\Delta\widehat{H}$, $\Delta H|\widehat{I}$ and LRE, and of the variances of $\Delta H|\widehat{I}$ agree reasonably well.

The discrepancy between the simulated variance estimates of $\Delta\widehat{H}$ and LRE and those obtained by the approximate procedure are somewhat larger. The simulated variance estimates are consistently higher, and the differences tend to increase with increasing sampling variance (σ_{ε}^2) of the economic weight vector.

The discrepancies between the Monte-Carlo estimates and the results derived from the approximate equations may be caused by (1) the inaccuracies of the approximation technique (Taylor series) used in the development of expectations and variances, and (2) the sample size used in the Monte-Carlo process (N = 100).

The effects of random errors in the economic weight vector can be evaluated by looking at the biases of $\Delta\widehat{H}$ and $\Delta H|\widehat{I}$, at the expected relative efficiency and at the sampling variances of $\Delta\widehat{H}$ and $\Delta H|\widehat{I}$.

Three different relative biases can be calculated :

(a) $B_1 = \dfrac{E(\Delta H) - \Delta\widehat{H}}{\Delta H}$. 100 measures with how much the estimated genetic

gain ($\Delta\widehat{H}$) overestimates the maximum attainable genetic gain (ΔH).

(b) $B2 = \dfrac{E(\Delta H \mid \widehat{I}) - \Delta H}{\Delta H}$. 100 measures the reduction in realized genetic gain $(\Delta H \mid \widehat{I})$ relative to the maximum attainable genetic gain (ΔH). It can be verified that the $E(LRE) = - B2$.

(c) $B3 = \dfrac{E(\widehat{\Delta H}) - E(\Delta H \mid \widehat{I})}{E(\Delta H \mid \widehat{I})}$. 100 measures with how much the estimated genetic gain $(\widehat{\Delta H})$ overestimates the realized genetic gain $E(\Delta H \mid \widehat{I})$.

Estimates of B1, B2 and B3 for different sampling errors in the economic weight vector are listed in Table 5 and plotted in figure 5.

TABLE 5

Relative bias of $\widehat{\Delta H}$ and $\Delta H \mid \widehat{I}$ estimates, and relative over-estimation
of realized genetic gain for different sampling errors in the economic weight vector

Biais relatif de $\widehat{\Delta H}$ et $\Delta H \mid \widehat{I}$ et surestimation relative du gain génétique
réalisé pour différentes erreurs d'échantillonnage du vecteur des pondérations économiques

Sampling error σ_ε)	$B1 = \dfrac{E(\widehat{\Delta H}) - \Delta H}{\Delta H} \cdot 100$	$B2 = \dfrac{E(\Delta H / \widehat{I}) - \Delta H}{\Delta H} \cdot 100$	$B3 = \dfrac{E(\widehat{\Delta H}) - E(\Delta H / \widehat{I})}{E(\Delta H / \widehat{I})} \cdot 100$
0.1	0.1197 (a)	— 0.1215	0.2413
	0.3819 (b)	— 0.1298	0.5123
0.2	0.4850	— 0.4924	0.9821
	1.0206	— 0.5108	1.5391
0.5	3.2190	— 3.3864	6.8368
	4.5017	— 3.3699	8.1461
1.0	14.1401	— 14.9057	34.1335
	16.6072	— 17.1759	40 7889

(a) Mathematically derived (approximation). Obtenue par approximation mathématique.
(b) Obtained by simulation. Obtenue par simulation.

The values of B1 and B2 indicate that $\widehat{\Delta H}$ is an overestimate of ΔH by about the same amount as $\Delta H \mid \widehat{I}$ is an underestimate of ΔH. This is illustrated in figure 5, where B1 and B2 are almost symmetrical. For sampling errors (σ_ε) smaller than 0.5, the values of B1 and B2 are relatively small and approximately less than 1 percent. For larger sampling errors $(\sigma_\varepsilon = 1.0)$ the biases are considerable and the values of B1 and B2 are respectively 16.61 and — 17,18 percent for the simulated estimates, and 14.14 and — 14.90 percent for the derived estimates. The expected loss in relative efficiency (plotted in figure 5) is equal to B2 but with positive sign.

The effects of sampling errors in the economic weights on B3 (the relative overestimation of $\Delta H \mid \widehat{I}$ by $\widehat{\Delta H}$) are relatively small for small sampling errors, but

fairly large for large sampling errors.   For sampling errors of 0.2 the estimates for
B3 are 0.98 percent (derived) and 1.54 percent (simulated), while for sampling
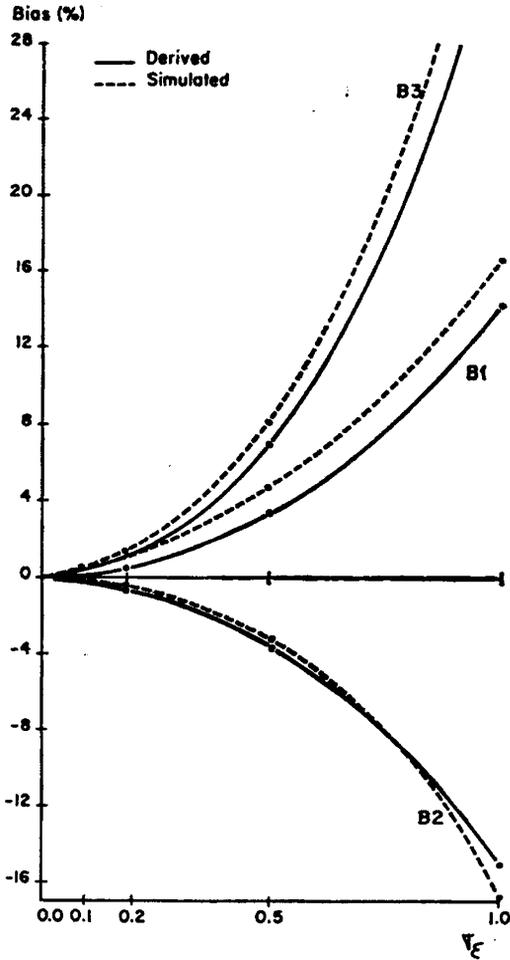errors of 1.0, B3 increases to 34.13 percent (derived) and 40.79 percent (simulated).

FIG. 5. — *Relative biases B1, B2 and B3 for different sampling errors in the economic weight vector.*
Biais relatifs B1, B2 et B3 en fonction des erreurs d'échantillonnage sur le vecteur
des pondérations économiques

 The variance of $\Delta \widehat{H}$ and $\Delta H |\widehat{I}$ for different sampling errors in the economic
weight vector are given in Table 4 and plotted in Figure 6.   The variance of
$\Delta \widehat{H}$ is relatively large and much larger than the variance of $\Delta H |\widehat{I}$.   Furthermore,
increasing errors in the economic weight vector make the variance of $\Delta \widehat{H}$ increase
at a much faster rate than the variance of $\Delta H |\widehat{I}$.   Since $\Delta \widehat{H}$ is likely to have
relatively large sampling errors we should be cautious not to draw far-reaching
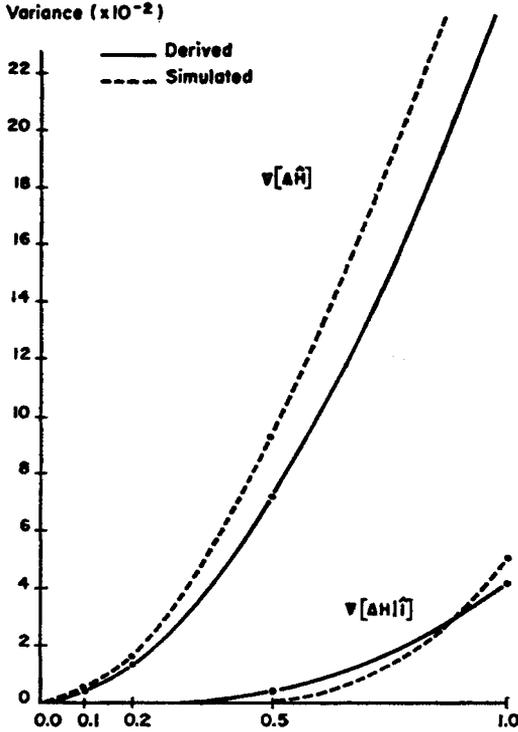conclusions or to make derivations based on the value of $\Delta \widehat{H}$.

FIG. 6. — *Variance of* $\Delta \widehat{H}$ *and* $\Delta H / \widehat{I}$ *for different sampling errors in the economic weight vector.*
Variance de ΔH *et* $\Delta H / \widehat{I}$ *en fonction des erreurs d'échantillonnage sur le vecteur des pondérations économiques*

Based on the symmetry between B1 and B2 and the relationship between B2 and the bias of $\Delta \widehat{H}$, $\Delta H / \widehat{I}$ and E(LRE) can be approximated by taking the first terms of a Taylor's expansion :

$$\widehat{B} = E(\Delta \widehat{H}) - \Delta H$$

$$\doteq [V(I) + \sigma_\varepsilon^2 tr(\mathbf{AU})]^{1/2} \left[1 - \frac{2\sigma_\varepsilon^4 tr(\mathbf{AU})^2 + 4\sigma_\varepsilon^2(\mathbf{a'AUAa})}{8[V(I) + \sigma_\varepsilon^2 tr(\mathbf{AU})]^2} \right] - [V(I)]^{1/2} \quad (10)$$

where $\widehat{B}$ = estimated bias
    V = variance
    tr = trace
    σ_ε^2 = the residual mean square from the regression procedure
    $\mathbf{U} = (\mathbf{X'X})^{-1}$ the inverse of the variance-covariance matrix from the regression procedure
    $\mathbf{A} = \mathbf{G'P^{-1}G}$ from the selection index
    $\mathbf{a}$ = the economic weight vector.

Then the estimates corrected for bias are :

$$\Delta H \doteq \widehat{\Delta H} - \widehat{B}$$

$$\Delta H | \widehat{I} \doteq \widehat{\Delta H} - 2\widehat{B}$$

$$LRE \doteq 100 . \frac{\widehat{B}}{\widehat{\Delta H}}$$

In general the loss in relative efficiency of a selection index and the biases and variances of $\widehat{\Delta H}$ and $\Delta H | \widehat{I}$ are function of :

(a)   the sampling variances and covariances of the economic weights $\sigma_{\mathfrak{z}}^{\mathfrak{e}}(\mathbf{X'X})^{-1}$ in which $\mathbf{X'X}$ is the variance covariance matrix and $\sigma_{\mathfrak{e}}^2$ the sampling variance.

This was studied by using different values for the sampling variance $(\sigma_{\mathfrak{e}}^2)$. When multiple regression procedures result in large residual mean squares $(\sigma_{\mathfrak{e}}^2)$ or in a small determination coefficient, the loss in relative efficiency of the resulting selection index can be large.   The variances and biases of $\widehat{\Delta H}$ and $\Delta H | \widehat{I}$ can be considerable and make the estimates rather doubtful.   However, when the sampling errors are small the resulting discrepancies are relatively unimportant.

(b)   The $\mathbf{G'P^{-1}G}$ matrix.   This not only involves the genetic and phenotypic variances and covariances but also the number of traits in the aggregate genotype and index.   Since this was not the real subject of our work, the genetic and phenotypic variances and covariances were considered as known constants, so that the effect of the $\mathbf{G'P^{-1}G}$ matrix has not been studied.

The results of this study being relative to the $\mathbf{P}$ and $\mathbf{G}$ matrices and the set of economic weights from our example, we must be cautious with generalisation.

Reçu pour publication en février 1977.

Acknowledgements

Résumé

L'influence d'erreurs dans les pondérations économiques
sur l'efficacité des index de sélection

Les effets des erreurs dans les pondérations économiques sur l'efficacité d'un index de sélection ont été étudiés. Comme exemple on a utilisé un index de sélection pour l'amélioration génétique des porcs.

Des erreurs de ± 50 pour cent dans la pondération économique d'un caractère font diminuer l'efficacité relative de l'index de sélection de moins d'1 pour cent pour tous les caractères considérés. Des erreurs plus grandes peuvent résulter en un biais assez important du gain génétique estimé. Les effets des erreurs dans les pondérations économiques individuelles ne sont ni linéaires ni symétriques. Dans le cas étudié, les erreurs négatives (sous-estimation) sont plus graves que les erreurs positives (surestimation). Cette dissymétrie est fonction cependant de l'échelle utilisée.

Les effets des erreurs d'échantillonnage simultanées des pondérations économiques ont été étudiés au moyen d'une simulation du type Monte-Carlo et par approximation mathématique (développement de Taylor). Les résultats obtenus par simulation indiquent que le gain génétique estimé ($\widehat{\Delta H}$) et le gain génétique réalisé ($\Delta H | \widehat{I}$) ne suivent pas une distribution normale.

Pour des erreurs dans le vecteur des poids économiques relativement petites (coefficient de variation C.V. $\leqslant$ 0.50) la diminution de l'efficacité relative est inférieure à 2.6 pour cent, mais s'élève approximativement jusqu'à 15 p. 100 pour C.V. = 1.0.

La perte d'efficacité relative d'un index de sélection, les biais et les variances de $\Delta\widehat{H}$ et $\Delta H|\widehat{I}$ dus aux erreurs d'échantillonnage dans le vecteur des poids économiques sont fonction des variances et des covariances des poids économiques et des variances-covariances génétiques et phénotypiques (matrices **P** et **G**) des caractères incorporés dans l'index de sélection. $\Delta\widehat{H}$ est une surestimation du gain génétique maximal et $\Delta H|\widehat{I}$ est une sous-estimation. Les biais de $\Delta\widehat{H}$ et $\Delta H|\widehat{I}$ sont petits pour des erreurs relativement petites, mais augmentent rapidement avec l'accroissement des erreurs d'échantillonnage dans le vecteur des poids économiques. La variance de $\Delta\widehat{H}$ est plus grande que celle de $\Delta H|\widehat{I}$ et augmente plus vite que la variance de $\Delta H|\widehat{I}$ quand les erreurs sur les pondérations économiques augmentent.

A cause des risques d'erreurs d'estimation dans $\Delta\widehat{H}$, il faut être prudent quant aux conclusions basées sur $\Delta\widehat{H}$. Pour des erreurs d'échantillonnage modérées dans les pondérations économiques, ces résultats constituent un argument assez fort pour la robustesse de la théorie des indices de sélection.

Les résultats précédents étant relatifs aux matrices **P** et **G** et aux pondérations économiques de l'exemple porcin utilisé, il faut être prudent dans la généralisation à d'autres situations.

References

CHRISTIAN L. L., 1970. Quantitative inheritance of swine. *Hog Farm Management*, **7**, 25-29.

FREDEEN Th. H., 1953. *Genetic aspects of Canadian bacon production*. Canadian Dept. of Agric. Publication 889. Ottawa, Canada.

HARRIS D. L., 1964. Expected and predicted progress from index selection involving estimates of population parameters. *Biometrics*, **20**, 46-72.

HAZEL L. N., 1943. The genetic basis for constructing selection indexes. *Genetics*, **28**, 476-490.

JONSSON P., 1965. *Analyze uf egenskaber hos svin af Dansk Landrace med en historisk indledning*. 350. Beretn. Forsøgslab., København. 490 pp.

KENDALL M. G. and STUART A., 1958. *The advanced theory of statistics*. Vol. I. *Distribution theory*. New York, N. Y., Hafner Publishing Co.

MANNING H. L., 1956. Yield Improvement with cotton. *Heredity*. **10**, 303-322.

PEASE A. H. R., COOK G. L., GREIG M., CUTHBERTSON A., 1967. *Combined testing*. Report DA 188 of the P.I.D.A. Hitchin, Herts, England.

RØNNINGEN K., 1971. Tables for estimating the loss in efficiency when selecting according to an index based on a false economic ratio between two traits. *Acta Agric. Scand.*, **21**, 33-49.

SALES J. and HILL W. G., 1976. Effect of sampling errors on efficiency of selection indices. I. Use of information from relatives for single trait improvement. *Anim. Prod.*, **22**, 1-17.

SIERS D. G. and THOMSON G. M., 1972. Heritabilities and genetic correlations of carcass and growth traits in swine. *J. Anim. Sci.*, **35**, 311-320.

VINT L. F., 1971. *Evaluation of selection practices used in a swine herd*. Unpublished Ph. D. Thesis. Ames, Iowa, Library, Iowa State University.