

# The Relative Efficiency of Two Multistage Linear Phenotypic Selection Indices to Predict the Net Genetic Merit

J. Jesus Cerón-Rojas, Fernando H. Toledo, and Jose Crossa<sup>★</sup>

## ABSTRACT

Multistage linear phenotypic selection indices predict the individual net genetic merit at different individual ages or stages and are cost-saving strategies for improving several traits. In a two-stage context, we compared the relative efficiency of the optimum multistage linear phenotypic selection index (OMLPSI) and the decorrelated multistage linear phenotypic selection index (DMLPSI) theory to predict the individual net genetic merit and selection response using a real and a simulated dataset. In addition, we described a method for obtaining the OMLPSI selection intensity in a two-stage context. The criteria used to compare the relative efficiency of both indices were that the total selection response of each index must be lower than or equal to the single-stage linear phenotypic selection index (LPSI) selection response, similar to the accuracy of each index to predict the net genetic merit. Using four different total proportions ( $p = 0.05, 0.10, 0.20, \text{ and } 0.30$ ) for the real dataset, the total DMLPSI selection response was 22.80% higher than the estimated single-stage LPSI selection response, whereas the total OMLPSI selection response was only 2.21% higher than the estimated single-stage LPSI selection response. In addition, at Stage 2, OMLPSI accuracy was 62.24% higher than the DMLPSI accuracy for predicting the net genetic merit. We found similar results for the simulated data. Thus, we recommend using OMLPSI when performing the multistage phenotypic selection.

Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600, Mexico City, Mexico. Received 10 Nov. 2018. Accepted 28 Feb. 2019. <sup>★</sup>Corresponding author (j.crossa@cgiar.org). Assigned to Associate Editor Timothy Beissinger.

**Abbreviations:** DMLPSI, decorrelated multistage linear phenotypic selection index; LPSI, single-stage linear phenotypic selection index; OMLPSI, optimum multistage linear phenotypic selection index; QTL, quantitative trait locus.

**T**HE linear phenotypic selection index (LPSI) (Smith, 1936; Hazel, 1943) is a linear combination of several observable and optimally weighted phenotypic trait values useful for predicting the net genetic merit, which, in turn, is a linear combination of the true unobservable breeding values of the traits weighted by their respective economic values. The LPSI incorporates the genetic correlations between traits in the prediction of the net genetic merit and allows extra merit in some traits to offset slight defects in another, and by its use, individuals with very high merit in some traits are saved for breeding even when they are inferior in other traits (Hazel and Lush, 1942). The main parameters of the LPSI are selection response and accuracy. The selection response is the mean of the net genetic merit (Cochran, 1951) in the selected population, whereas the accuracy is the maximized correlation between the LPSI and the net genetic merit. These two parameters give breeders a clearer base on which to validate the success of the adopted selection method that is technically most effective on an objective basis; they are also useful for comparing the efficiency of different types of selection indices.

Cochran (1951) and Young (1964) combined the LPSI theory with the independent culling selection method and developed the optimum multistage (one stage is the animal or plant age at which breeders can measure and select the trait of economic interest) linear

Published in Crop Sci. 59:1037–1051 (2019).  
doi: 10.2135/cropsci2018.11.0678

© 2019 The Author(s). This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

phenotypic selection index (OMLPSI) to select several traits in the multistage selection context. Suppose two vectors of individual traits,  $\mathbf{x}$  and  $\mathbf{y}$ , become evident at different animal or plant stages. We can make a selection at one stage according to the LPSI using both vectors of information jointly, or we can perform a two-stage selection, in which case, we select  $\mathbf{x}$  in the first stage, and  $\mathbf{x}$  and  $\mathbf{y}$  in the second stage. When we use the OMLPSI in a two-stage context, at Stage 1, we have a partial index, but at Stage 2, we have a complete index. The OMLPSI is more efficient than the independent culling method because it uses all available information at each stage and incorporates the genetic correlations between traits in the prediction. The OMLPSI can be applied to any number of stages.

Breeders apply OMLPSI mainly in animal and tree breeding where, due to early culling, OMLPSI is a cost-saving strategy for improving multiple traits because it is not necessary to measure all traits at each stage. Thus, when traits have a developmental sequence in ontogeny or when there are large differences in the costs of measuring several traits, the efficiency of OMLPSI over LPSI, in terms of cost saving, can be substantial. The OMLPSI increases selection intensity on traits measured at an earlier age, and, with fixed facilities, OMLPSI selects a greater number of individuals at an earlier age (Xu et al., 1995; Xie et al., 1997; Hicks et al., 1998).

Some problems associated with the OMLPSI are the following. After the first selection stage, the distribution of OMLPSI values could be non-normal. For more than two stages, the OMLPSI requires numerical multiple integration techniques to derive selection intensities for each stage, and there are problems of convergence when the trait and index values at successive stages are highly correlated. Also, the computational time may be unacceptable if the number of selection stages becomes too high (Xu and Muir, 1992; Börner and Reinsch, 2012).

The decorrelated multistage linear phenotypic selection index (DMLPSI) is another multistage selection index (Börner and Reinsch, 2012), also called *selection index updated* (Xu and Muir, 1992). It also combines the LPSI theory with the independent culling selection method to select several traits in the multistage selection context. The DMLPSI minimizes the mean squared difference between the DMLPSI and the net genetic merit at each stage under the restriction that the covariance between the DMLPSI values at different stages be zero, thus preventing the correlation between DMLPSI values at different stages; hence the name *decorrelated*. Under such restriction, the selected individual DMLPSI values after the first selection stage could be normally distributed, and in addition, it is not necessary to use computationally sophisticated multiple integration techniques to derive the selection intensities, and exact truncation points can be determined for a fixed selection proportion before selection is performed (Xu and

Muir, 1991, 1992). Problems associated with DMLPSI are that its selection responses and accuracy could be lower than the OMLPSI selection response and accuracy after the first selection stage. The OMLPSI and DMLPSI are extensions of the LPSI theory to the multistage selection context.

We compared the relative efficiency of OMLPSI and DMLPSI using a real and a simulated dataset in a two-stage context. We obtained the theoretical results of both indices under the assumption that the indices, and the net genetic merit values have multivariate normal distribution at each stage. Under this assumption, the regression of the net genetic merit on any linear function of the phenotypic values is linear (Kempthorne and Nordskog, 1959), and the total selection response for two or more stages is the sum of each response obtained at each stage (Cochran, 1951; Young, 1964).

The criteria used to compare the relative efficiency of both indices were that the total selection response of each index must be lower than or equal to the LPSI selection response (Young, 1964; Saxton, 1983) and the accuracy of each index to predict the net genetic merit. Using four different total proportions for the real dataset, we found that the OMLPSI efficiency was higher than the DMLPSI efficiency when predicting the net genetic merit. We found similar results for the simulated data. Börner and Reinsch (2012) reported similar results in the genomic selection context when they used multistage selection indices in a dairy cattle breeding program.

## MATERIALS AND METHODS

### Objectives of the Multistage Linear Phenotypic Selection Indices

The OMLPSI and DMLPSI have four main objectives. The first objective is to predict the individual net genetic merit for  $n$  traits,  $H = \mathbf{w}'\mathbf{g}$ , where  $\mathbf{w}' = [w_1 \ w_2 \ \dots \ w_n]$  and  $\mathbf{g}' = [g_1 \ g_2 \ \dots \ g_n]$  are  $1 \times n$  vectors of economic weights and true breeding values, respectively. The second objective is to select individuals with the highest  $H$  values as parents of the next generation, and the third objective is to maximize the OMLPSI (DMLPSI) selection response. Finally, OMLPSI and DMLPSI should provide the breeder with an objective rule for evaluating and selecting several traits simultaneously.

When selection is based on all individual traits of interest jointly, the LPSI vector of coefficients that maximizes the selection response is  $\mathbf{b} = \mathbf{P}^{-1}\mathbf{C}\mathbf{w}$ , where  $\mathbf{C}$  (see Appendix A, subsection “Phenotypic and Genotypic Matrices for Two Stages” for details) is the covariance matrix of the true breeding values ( $\mathbf{g}$ ), and  $\mathbf{P}^{-1}$  is the inverse matrix of the covariance matrix ( $\mathbf{P}$ ) of trait phenotypic values ( $\mathbf{y}$ ). In addition,  $k$  is the selection intensity of the LPSI.

### The Multistage Linear Phenotypic Selection Index at Stage $i$

Let  $\mathbf{y}' = [y_1 \ y_2 \ \dots \ y_n]$  be a vector with  $n$  traits of interest and assume that we can select only  $n_i$  of them ( $n_i < N$ ) at Stage  $i$

( $i = 1, 2, \dots, N$ ), such that after  $N$  stages ( $N < n$ ),  $\sum_{i=1}^N n_i = n$ . Thus, for each stage, we should have a selection index with a different number of traits; for example, at Stage  $i$ , the index would be  $I_i = \sum_{j=1}^{n_i} b_{ij} y_{ij}$ , and at Stage  $N$ , the index would be  $I_N = \sum_{j=1}^{n_1} b_{1j} y_{1j} + \sum_{j=n_1+1}^{n_2} b_{2j} y_{2j} + \dots + \sum_{j=n_{N-1}+1}^{n_N} b_{Nj} y_{Nj} = \sum_{i=1}^N I_i$  (Young, 1964), where the double subscript of  $y_{ij}$  indicates that the  $j$ th trait is measured at Stage  $i$ , so that at each subindex  $I_i$ , all the  $n_i$  traits are measured at the same age.

The vector of scores  $\mathbf{y}' = [y_1 \ y_2 \ \dots \ y_n]$  can be partitioned into  $N$  subvectors as  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2 \ \dots \ \mathbf{x}'_N]$ , where  $\mathbf{x}'_1 = [y_1 \ y_2 \ \dots \ y_{n_1}]$ ,  $\mathbf{x}'_2 = [y_{(n_1+1)} \ y_{(n_1+2)} \ \dots \ y_{(n_1+n_2)}]$ , etc., are subvectors of  $\mathbf{y}'$  obtained at Stages 1, 2, etc. Thus, another way of writing the index at Stage  $i$  is  $I_i = \sum_{j=1}^{n_i} b_{ij} y_{ij} = \mathbf{b}'_i \mathbf{x}'_i$ , where  $\mathbf{b}'_i = [b_{i1} \ b_{i2} \ \dots \ b_{in_i}]$  is the index vector of coefficients at Stage  $i$ , and  $\mathbf{x}'_i$  was defined above. Let

$$\mathbf{B}'_0 = \begin{bmatrix} \mathbf{b}'_1 & 0 & \dots & 0 \\ \mathbf{b}'_1 & \mathbf{b}'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}'_1 & \mathbf{b}'_2 & \dots & \mathbf{b}'_N \end{bmatrix}$$

be a transforming matrix, then for each stage we can construct a multistage index as

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} = \begin{bmatrix} \mathbf{b}'_1 & 0 & \dots & 0 \\ \mathbf{b}'_1 & \mathbf{b}'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}'_1 & \mathbf{b}'_2 & \dots & \mathbf{b}'_N \end{bmatrix} \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_N \end{bmatrix}$$

The total selection index is  $I_N = \sum_{i=1}^N I_i = \mathbf{b}' \mathbf{y}$ , where  $\mathbf{b}' = [\mathbf{b}'_1 \ \mathbf{b}'_2 \ \dots \ \mathbf{b}'_N]$  and  $\mathbf{y}$  was defined above (see Appendix A for additional details).

The above procedure uses a partial index until Stage  $N - 1$ , but at Stage  $N$ , it uses a complete index (see Appendix A, subsection “Phenotypic and Genotypic Matrices for Two Stages” for additional details). This approach should be more efficient than the usual independent culling selection method because, at each stage, the OMLPSI (DMLPSI) uses all available information (Young, 1964; Saxton, 1983).

## The Multistage Linear Phenotypic Selection Index Phenotypic and Genotypic Covariance Matrices

Let  $\mathbf{g}' = [g_1 \ g_2 \ \dots \ g_n]$ ,  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2 \ \dots \ \mathbf{x}'_N]$ , and  $\mathbf{x}'_i = [y_1 \ y_2 \ \dots \ y_{n_i}]$ , as defined in the subsection above; then,  $\mathbf{P} = \{\mathbf{P}_{ij}\}$  is the covariance matrix of vector  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2 \ \dots \ \mathbf{x}'_N]$ , where  $\mathbf{P}_{ij} = \text{Cov}(\mathbf{x}'_i, \mathbf{x}'_j)$  is the  $ij$ th ( $i, j = 1, 2, \dots, N$ ) submatrix of  $\mathbf{P}$  and  $\mathbf{G}' = \{\text{Cov}(\mathbf{x}'_i, \mathbf{g})\}' = [\mathbf{G}_1 \ \mathbf{G}_2 \ \dots \ \mathbf{G}_N]$  is the covariance matrix between  $\mathbf{x}'_i$  and  $\mathbf{g}$ , where  $\text{Cov}(\mathbf{x}'_i, \mathbf{g}) = \mathbf{G}_i$  is a submatrix of  $\mathbf{G}$  at Stage  $i$  ( $i = 1, 2, \dots, N$ ). In addition, let

$$\mathbf{Q}_{ij} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & \mathbf{P}_{1j} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \dots & \mathbf{P}_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{i1} & \mathbf{P}_{i2} & \dots & \mathbf{P}_{ij} \end{bmatrix} \text{ and } \mathbf{A}_i = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_i \end{bmatrix} \quad [1a]$$

be submatrices of  $\mathbf{P}$  and  $\mathbf{G}$ , respectively; when  $i = j$ ,

$$\mathbf{Q}_{ij} = \mathbf{Q}_{ii} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & \mathbf{P}_{1i} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \dots & \mathbf{P}_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{i1} & \mathbf{P}_{i2} & \dots & \mathbf{P}_{ii} \end{bmatrix}$$

In Appendix A (Eq. [A1a] and [A1b]), we give additional details associated with matrices  $\mathbf{P}$  and  $\mathbf{G}$ .

Now suppose that the number of traits selected until Stage  $i - 1$  is  $n_{i-1}$  and that at Stage  $i$ , we select  $n_i$  traits, such that  $n_i \leq n_{i-1}$ . By the results in subsection “The Multistage Linear Phenotypic Selection Index at Stage  $i$ ” above, at Stage  $i$ , we shall have  $n_{i-1} + n_i$  traits. This means that matrix  $\mathbf{Q}_{(i-1)i}$  is of size  $n_{i-1}(n_{i-1} + n_i)$  and can be written as

$$\mathbf{Q}_{(i-1)i} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n_{i-1}} & \dots & \sigma_{1(n_{i-1}+n_i)} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n_{i-1}} & \dots & \sigma_{2(n_{i-1}+n_i)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{n_{i-1}1} & \sigma_{n_{i-1}2} & \dots & \dots & \dots & \sigma_{n_{i-1}(n_{i-1}+n_i)} \end{bmatrix} \quad [1b]$$

Equation [1b] indicates that  $\mathbf{Q}_{(i-1)i}$  is a nonsquare and nonsymmetric phenotypic variance–covariance matrix.

## Selection Response at Stage $i$

The selection response ( $R_i$ ) at Stage  $i$  can be written as

$$R_i = k_i \sigma_H \rho_{HI_i} \quad [2]$$

where  $k_i$  (Appendix B) is the selection intensity,  $\sigma_H = \sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}$  is the standard deviation of the variance of the net genetic merit ( $H = \mathbf{w}'\mathbf{g}$ ),  $\mathbf{w}' = [w_1 \ w_2 \ \dots \ w_n]$  and  $\mathbf{g}' = [g_1 \ g_2 \ \dots \ g_n]$  were defined earlier,  $\text{Var}(\mathbf{g}) = \mathbf{C}$  is the covariance matrix of  $\mathbf{g}$ , and  $\rho_{HI_i}$  is the correlation between  $H = \mathbf{w}'\mathbf{g}$  and the index at Stage  $i$  ( $I_i = \mathbf{b}'_i \mathbf{x}'_i$ ).

The second part of Eq. [2] ( $k_i \sigma_H \rho_{HI_i}$ ) indicates that, at each stage, the genetic change due to selection is proportional to  $k_i$ ,  $\sigma_H$ , and  $\rho_{HI_i}$  (Kempthorne and Nordskog, 1959). Thus the genetic gain that can be achieved at each stage by selecting for several traits simultaneously within a population of animals or plants is the product of the selection differential ( $k_i$ ), the standard deviation of  $H = \mathbf{w}'\mathbf{g}$  ( $\sigma_H$ ), and the correlation between  $H = \mathbf{w}'\mathbf{g}$  and  $I_i = \mathbf{b}'_i \mathbf{x}'_i$  ( $\rho_{HI_i}$ ). Selection intensity  $k_i$  is limited by the rate of reproduction of each species, whereas  $\sigma_H$  is relatively beyond man’s control; hence, the best opportunity for increasing selection progress is by ensuring that  $\rho_{HI_i}$  is as large as possible (Hazel, 1943). In general, it is assumed that  $k_i$  and  $\sigma_H$  are fixed and  $\mathbf{w}$  is known and fixed; hence,  $R_i$  will be maximized when  $\rho_{HI_i}$  is maximized only with respect to the vector of coefficients  $\mathbf{b}'_i$  ( $i = 1, 2, \dots, N$ ). To maximize  $R_i$ , we can either maximize  $\rho_{HI_i}$  or minimize the mean squared difference between  $I_i = \mathbf{b}'_i \mathbf{x}'_i$  and  $H = \mathbf{w}'\mathbf{g}$  with respect to the vector of coefficients  $\mathbf{b}'_i$ .

## The Maximized Multistage Linear Phenotypic Selection Index Parameters

According to Cerón-Rojas and Crossa (2018, Chapter 9), the maximized selection response and accuracy at Stage  $i$  are

$$R_i = k_i \sqrt{\mathbf{b}_i' \mathbf{Q}_{ii} \mathbf{b}_i} \quad [3]$$

and

$$\rho_i = \frac{\sqrt{\mathbf{b}_i' \mathbf{Q}_{ii} \mathbf{b}_i}}{\sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad [4]$$

respectively, whereas the total selection response is  $R_t = \sum_{i=1}^N R_i$  (Cochran, 1951; Young, 1964). Matrices  $\mathbf{Q}_{ii}$  and  $\mathbf{A}_i$  are phenotypic and genetic covariance matrices, respectively, defined in Eq. [1a]. Note that Eq. [4] gives the maximum value of  $\rho_{H_i}$ . Thus, although  $\rho_{H_i}$  can take any value,  $\rho_i$  is its maximum value.

Equation [3] and [4] values will change depending on how the vector of coefficients ( $\mathbf{b}_i$ ) is obtained for each index. The OMLPSI and DMLPSI vectors of coefficients will be obtained according to the OMLPSI theory (Young, 1964; Cerón-Rojas and Crossa, 2018, Chapter 9) and the Xu and Muir (1992) method, respectively (see Appendix A, Eq. [A2a] to [A9]).

## The OMLPSI Vector of Coefficients

Cerón-Rojas and Crossa (2018, Chapter 9) have shown that the OMLPSI vector of coefficients that maximizes Eq. [2] at Stage  $i$  is

$$\mathbf{b}_i = \mathbf{Q}_{ii}^{-1} \mathbf{A}_i \mathbf{w} \quad [5]$$

where  $\mathbf{Q}_{ii}^{-1}$  is the inverse of matrix  $\mathbf{Q}_{ii}$ ; matrix  $\mathbf{A}_i$  was defined in Eq. [1a], and  $\mathbf{w}$  is the vector of economic weights. To obtain the maximized OMLPSI selection response and accuracy, Eq. [5] must be used in Eq. [3] and [4].

## The DMLPSI Vector of Coefficients

In Appendix A (Eq. [A2a] to [A9]), we showed that the DMLPSI vector of coefficients at Stage  $i$  is

$$\beta_i = \mathbf{K}_i \mathbf{b}_i \quad [6]$$

where  $\mathbf{K}_i = [\mathbf{I} - \Psi_i]$ ,  $\Psi_i = \mathbf{Q}_{ii}^{-1} \mathbf{S}_{i(i-1)} [\mathbf{S}_{(i-1)i} \mathbf{Q}_{ii}^{-1} \mathbf{S}_{i(i-1)}]^{-1} \mathbf{S}_{(i-1)i}$  (Appendix A, Eq. [A4] to [A8] for details), and  $\mathbf{b}_i = \mathbf{Q}_{ii}^{-1} \mathbf{A}_i \mathbf{w}$  is the OMLPSI vector of coefficients (Eq. [5]), whereas  $\mathbf{Q}_{ii}^{-1}$  and  $\mathbf{A}_i$  were defined earlier, and  $\mathbf{I}$  is an identity matrix of the same size as matrix  $\mathbf{Q}_{ii}$ . When  $\mathbf{S}_{i(i-1)} = 0$ ,  $\mathbf{K}_i = \mathbf{I}$ , and  $\beta_i = \mathbf{b}_i = \mathbf{Q}_{ii}^{-1} \mathbf{A}_i \mathbf{w}$ , which occurs at Stage 1, when the DMLPSI and OMLPSI vectors of coefficients are the same. To obtain the maximized DMLPSI parameters, Eq. [6] should be used in Eq. [3] and [4]. Matrix  $\mathbf{K}_i$  is the only difference between Eq. [5] and [6], and it transforms the OMLPSI vector of coefficients into the DMLPSI vector of coefficients.

## Criteria for Comparing the Efficiency of Each Index to Predict the Net Genetic Merit

The relative efficiency of predicting the net genetic merit at each stage of OMLPSI with respect to DMLPSI efficiency in percentage terms is

$$\varphi = 100(\pi - 1) \quad [7]$$

where  $\pi = \rho_i / r_i$ ;  $\rho_i = \sqrt{\mathbf{b}_i' \mathbf{Q}_{ii} \mathbf{b}_i} / \sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}$  and  $r_i = \sqrt{\beta_i' \mathbf{Q}_{ii} \beta_i} / \sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}$  are OMLPSI and DMLPSI accuracy at Stage  $i$ , respectively. In addition,  $\sqrt{\mathbf{b}_i' \mathbf{Q}_{ii} \mathbf{b}_i}$  and  $\sqrt{\beta_i' \mathbf{Q}_{ii} \beta_i}$  are the standard deviations

of OMLPSI and DMLPSI at Stage  $i$ , respectively, whereas  $\sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}$  is the standard deviation of the net genetic merit ( $H = \mathbf{w}' \mathbf{g}$ ). When  $\varphi = 0$ , the efficiency of both indices is the same; when  $\varphi > 0$ , the efficiency of OMLPSI is higher than DMLPSI efficiency, and when  $\varphi < 0$ , DMLPSI efficiency is higher than OMLPSI efficiency. An additional criterion for comparing the indices' efficiency is that the total selection response  $R_t = R_1 + R_2$  of each index should be lower than or equal to the single-stage LPSI selection response ( $R = k\sigma_p$ ; i.e.,  $R_t \leq R$ ).

We show that  $R_t \leq R$  for OMLPSI only. Suppose that the total proportion retained is  $p = q_1 q_2$  (Fig. 1); then, LPSI selection intensity  $k$  (Fig. 2) is fixed. In addition, since the LPSI standard deviation ( $\sigma_p$ ) is fixed in the target population, the maximum LPSI selection response in the selected population is  $R = k\sigma_p$ , which is the maximum value that is possible to attain for  $p = q_1 q_2$ . In the two-stage context, suppose again that  $p = q_1 q_2$ . Then, at Stage 1,  $R_1 = k_1 \sigma_{I_1}$ , and at Stage 2,  $R_2 = k_2 \sigma_{I_2}$ , where the  $k_1$  and  $k_2$  values are associated with the  $q_1$  and  $q_2$  values, respectively. That is,  $R_t = R_1 + R_2$  is the total OMLPSI selection response. However,  $R = k\sigma_p$  is the maximum value that can be attained when  $p = q_1 q_2$  is fixed; thus,  $R_t \leq R$ . This criterion allows breeders to know the maximum value of  $R_t$  in the breeding context.

## Real Dataset

The number of genotypes in this real dataset was 3330, and the vector of economic weights ( $\mathbf{w}$ ) was  $\mathbf{w}' = [19.54 \ -3.56 \ 17.01 \ -2.51]$ . This dataset comes from a commercial egg poultry line (Akbar et al., 1984), and we will use it to illustrate the indices' theoretical results obtained in this work. The estimated phenotypic ( $\hat{\mathbf{P}}$ ) and genotypic ( $\hat{\mathbf{C}}$ ) covariance matrices among the rate of lay (RL, number of eggs), age at sexual maturity (SM, days), egg weight (EW, kg), and body weight (BW, kg) were

$$\hat{\mathbf{P}} = \begin{bmatrix} 240.57 & -95.62 & 2.06 & 54.40 \\ -95.62 & 167.20 & 4.58 & 15.36 \\ 2.07 & 4.58 & 22.80 & 37.20 \\ 54.40 & 15.36 & 37.20 & 516.11 \end{bmatrix}$$

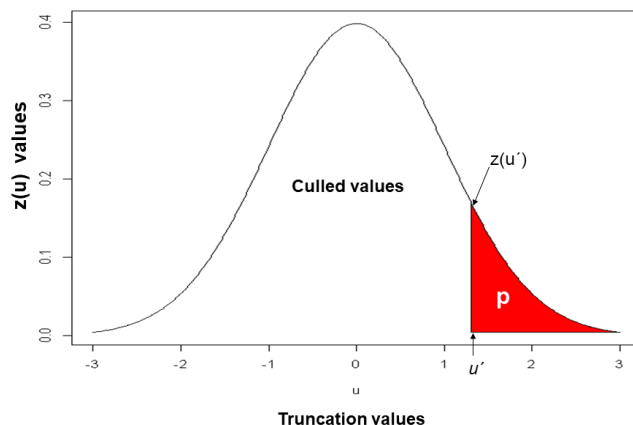


Fig. 1. Theoretical relationship between the truncation points ( $u$ ), the proportion retained ( $p$ ), and the density values  $[z(u)]$  of the truncation points.



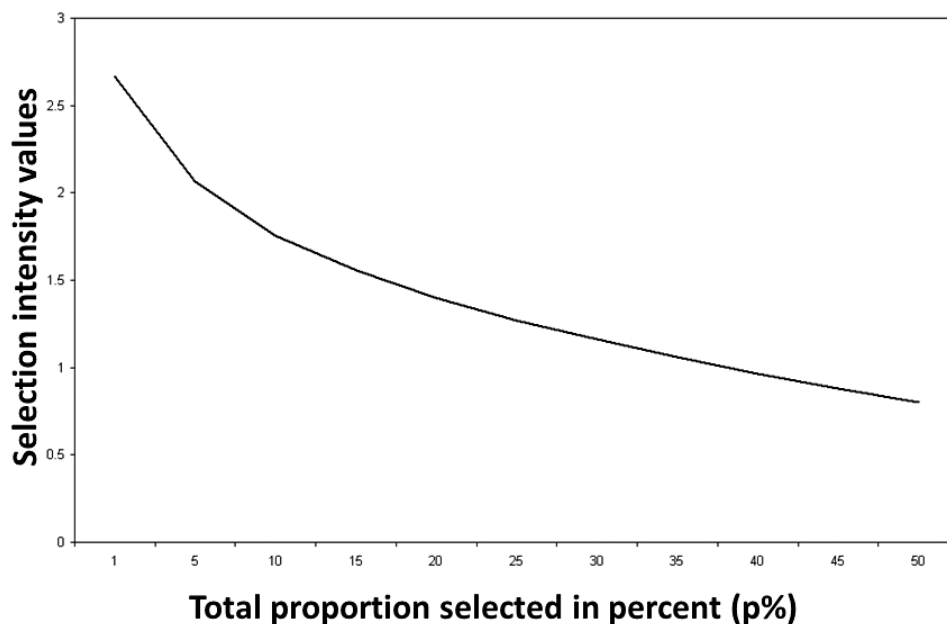


Fig. 2. Values of the selection intensity for different fixed total proportion ( $p = q_1 q_2$ ) selected values.

and

$$\hat{\mathbf{C}} = \begin{bmatrix} 29.86 & -17.90 & -4.13 & -1.75 \\ -17.90 & 18.56 & 1.49 & -4.88 \\ -4.13 & 1.49 & 9.24 & 16.66 \\ -1.75 & -4.88 & 16.66 & 179.73 \end{bmatrix}$$

The total proportions ( $p$ ) of retained values for this dataset were  $p = 0.05, 0.10, 0.20$ , and  $0.30$  for both indices.

For illustration purposes only, at Stage 1, we selected traits RL and SM, whereas at Stage 2, we selected the four traits, according to the Young (1964) method. The vectors of phenotypic values at Stages 1 and 2 were  $\mathbf{x}'_1 = [\text{RL SM}]$  and  $\mathbf{y}' = [\mathbf{x}'_1 \mathbf{x}'_2]$ , respectively, where  $\mathbf{x}'_2 = [\text{EW BW}]$ . At Stage 1, the estimated phenotypic ( $\hat{\mathbf{Q}}_{11}$ ) and genotypic ( $\hat{\mathbf{A}}_1$ ) covariance matrices were

$$\hat{\mathbf{Q}}_{11} = \begin{bmatrix} 240.57 & -95.62 \\ -95.62 & 167.20 \end{bmatrix}$$

and

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} 29.86 & -17.90 & -4.13 & -1.75 \\ -17.90 & 18.56 & 1.49 & -4.88 \end{bmatrix}$$

whereas the estimated covariance matrix of the traits at Stage 1 with traits at Stage 2 ( $\hat{\mathbf{Q}}_{12}$ ) was

$$\hat{\mathbf{Q}}_{12} = \begin{bmatrix} 240.57 & -95.62 & 2.06 & 54.40 \\ -95.62 & 167.20 & 4.58 & 15.36 \end{bmatrix}$$

(see Appendix A, subsection “Phenotypic and Genotypic Matrices for Two Stages” for additional details). At Stage 2,  $\hat{\mathbf{Q}}_{22} = \hat{\mathbf{P}}$  and  $\hat{\mathbf{A}}_2 = \hat{\mathbf{C}}$ .

## Simulated Dataset

This dataset is available in the “Application of a Genomics Selection Index to Real and Simulated Data” repository at <http://hdl.handle.net/11529/10199>. It was simulated (Cerón-Rojas et al., 2015) for eight phenotypic selection cycles (C0–C7), each with four traits ( $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ ), 500 genotypes, and four replicates for each genotype. Data were generated with QU-GENE software (Podlich and Cooper, 1998) using 2500 molecular markers and 315 quantitative trait loci (QTLs). The markers were distributed uniformly across 10 chromosomes, whereas the QTLs were randomly allocated across the 10 chromosomes to simulate one maize (*Zea mays* L.) population. For each trait, the phenotypic value for each of four replications of each plant was obtained by setting the per-plot heritability of  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$  at 0.4, 0.6, 0.6, and 0.8, respectively. A different number of QTLs affected each of the four traits: 300, 100, 60, and 40, respectively. The common QTLs affecting the traits generated genotypic correlations of  $-0.5$ ,  $0.4$ ,  $0.3$ ,  $-0.3$ ,  $-0.2$ , and  $0.1$  between  $T_1$  and  $T_2$ ,  $T_1$  and  $T_3$ ,  $T_1$  and  $T_4$ ,  $T_2$  and  $T_3$ ,  $T_2$  and  $T_4$ , and  $T_3$  and  $T_4$ , respectively. The economic weights for  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$  were 1,  $-1$ , 1, and 1, respectively. Only for illustration purposes, in this work, we used four selection cycles (C1–C4) of the simulated data to illustrate the theoretical results and the efficiency of both indices. At Stage 1, we selected traits  $T_1$  and  $T_2$ , and at Stage 2, we selected all four traits.

## RESULTS

### Real Data

#### Estimated OMLPSI Parameters for Stages 1 and 2

The estimated vectors of coefficients for both stages were  $\mathbf{b}'_1 = [1.936 \ -1.139]$  and  $\mathbf{b}'_2 = [2.149 \ -1.034 \ 2.508 \ -0.733]$ . Thus,  $\hat{I}_1 = \hat{\mathbf{b}}_1 \mathbf{x}_1$  and  $\hat{I}_2 = \hat{\mathbf{b}}_2 \mathbf{y}$  were the estimated OMLPSI values for both stages. The standard deviations of  $\hat{I}_1$  and  $\hat{I}_2$  were  $\sqrt{\hat{\mathbf{b}}_1' \hat{\mathbf{Q}}_{11} \hat{\mathbf{b}}_1} = 39.647$  and  $\sqrt{\hat{\mathbf{b}}_2' \hat{\mathbf{P}} \hat{\mathbf{b}}_2} = 25.862$ , respectively, where  $\hat{\mathbf{P}}^*(\hat{\mathbf{C}}^*)$  was matrix  $\hat{\mathbf{P}}(\hat{\mathbf{C}})$  adjusted for prior selection on  $\hat{I}_1 = \hat{\mathbf{b}}_1 \mathbf{x}_1$ . Matrix  $\hat{\mathbf{P}}^*(\hat{\mathbf{C}}^*)$  should be used

in the estimated selection response and accuracy at Stage 2, but not to estimate the vector of coefficients (Cerón-Rojas and Crossa, 2018, Chapter 9, Eq. [9.5] and [9.6]).

Figure 1 shows the relationship among the truncation points ( $u_1$  and  $u_2$ ), the proportion retained ( $q_1$  and  $q_2$ ) and the heights of the ordinate of the normal curve  $z(u_1) = e^{-0.5u_1^2}/\sqrt{2\pi}$  and  $z(u_2) = e^{-0.5u_2^2}/\sqrt{2\pi}$  (Appendix B, Eq. [A9]). For Stages 1 and 2, we found the selection intensities  $k_1 = z(u_1)/q_1$  and  $k_2 = z(u_2)/q_2$  (Appendix B, Eq. [A9]), as follows. For a fixed value of  $p = q_1q_2$  (e.g.,  $p = 0.05$ ), we used an iterative process with an R code. By successively changing the possible values of  $q_1$  ( $q_2 = p/q_1$ ),  $u_1$ , and  $u_2$  in Eq. [A9] (Appendix B), we found the maximum value of the total selection response  $\hat{R}_t = \hat{R}_1 + \hat{R}_2 = 39.647k_1 + 25.862k_2 = 90.33$  (Fig. 3), where  $\hat{R}_1 = k_1\hat{\sigma}_{I_1}$  and  $\hat{R}_2 = k_2\hat{\sigma}_{I_2}$  were the estimated selection responses at each stage, whereas  $\hat{\sigma}_{I_1}$  (39.647) and  $\hat{\sigma}_{I_2}$  (25.862) were the estimated standard deviations of the variance of  $\hat{I}_1$  and  $\hat{I}_2$ , respectively. Thus, the values of the truncation points ( $u_1 = 1.21$  and  $u_2 = 0.15$ ), proportions retained ( $q_1 = 0.11$  and  $q_2 = 0.44$ ), and selection intensity ( $k_1 = 1.69$  and  $k_2 = 0.90$ ) at Stages 1 and 2, respectively, were those associated with the maximum  $\hat{R}_t = 90.33$  value (Fig. 3).

In the one-stage case, the selection intensity for  $p = 0.05$  was  $k = 2.06$ , and the single-stage estimated LPSI selection response was  $\hat{R} = 88.72$ . According to Young (1964) and Saxton (1983), the maximum estimated total OMLPSI selection response ( $\hat{R}_t = 90.33$ ) value should be lower than or equal to the single-stage estimated LPSI selection response ( $\hat{R} = 88.72$ ). For this dataset, the

estimated total OMLPSI selection response was 1.81% higher than the estimated LPSI selection response.

In Table 1, we present additional truncation points, proportions retained, selection intensities, and maximum estimated selection response values for  $p = q_1q_2 = 0.10, 0.20$ , and  $0.30$ . For each of the latter three values, the maximum estimated total OMLPSI selection responses were 2.01, 2.33, and 2.69%, respectively, greater than the estimated single-stage LPSI selection response (Table 1). Thus, for this real dataset, the estimated total OMLPSI selection response and the estimated LPSI selection response were very similar.

Figure 4 presents the estimated total OMLPSI selection response values ( $\hat{R}_t$ ) when we obtained the selection intensities using the Cochran (1951) and Young (1964) method for 0.05 and 0.10. Cochran (1951) and Young (1964) found the expectations of a bivariate left truncated normal distribution and used them as selection intensities. When we used the Cochran (1951) method, the maximum  $\hat{R}_t$  values were 393.90 and 249.98 for  $p = 0.05$  and 0.10, respectively, and when we used the Young (1964) method, the maximum  $\hat{R}_t$  values were 535.27 and 293.27, respectively. This is because both methods overestimated the selection intensities at both stages. These results were inadmissible when we compared them with the single-stage estimated LPSI selection response values for  $p = 0.05$  and 0.10 (88.72 and 75.47, respectively). We obtained similar results for other values. Thus, according to the results of this section, the Cochran (1951) and Young (1964) method should not be used to obtain the selection intensities and total selection responses in the two-stage context.

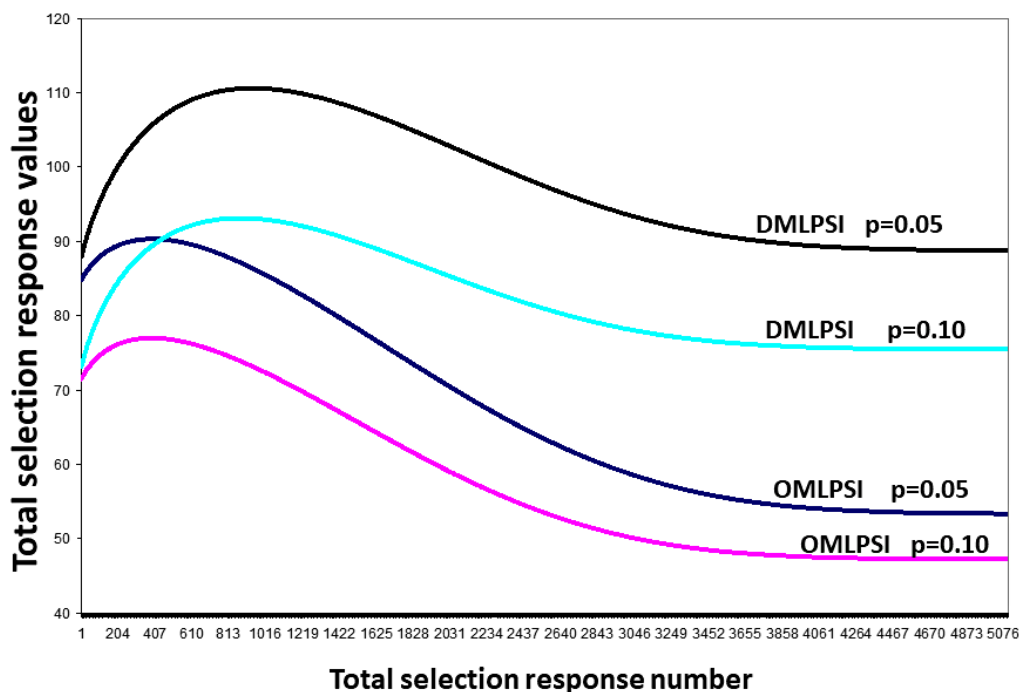


Fig. 3. Distribution of the total estimated optimum (OMLPSI) and decorrelated multistage linear phenotypic selection index (DMLPSI) selection response values, under a two-stage breeding scheme, for a real dataset.

**Table 1.** Real data for total proportion ( $p$ ) retained, estimated optimum and decorrelated multistage linear phenotypic selection index (OMLPSI and DMLPSI, respectively) truncation points ( $u_1$  and  $u_2$ ), proportions retained ( $q_1$  and  $q_2$ ), selection intensities ( $k_1$  and  $k_2$ ), and estimated selection responses ( $\hat{R}_1$ ,  $\hat{R}_2$ , and  $\hat{R}_t = \hat{R}_1 + \hat{R}_2$ ) for Stages 1 and 2. Values of  $\hat{R}$  correspond to single-stage linear phenotypic selection index.

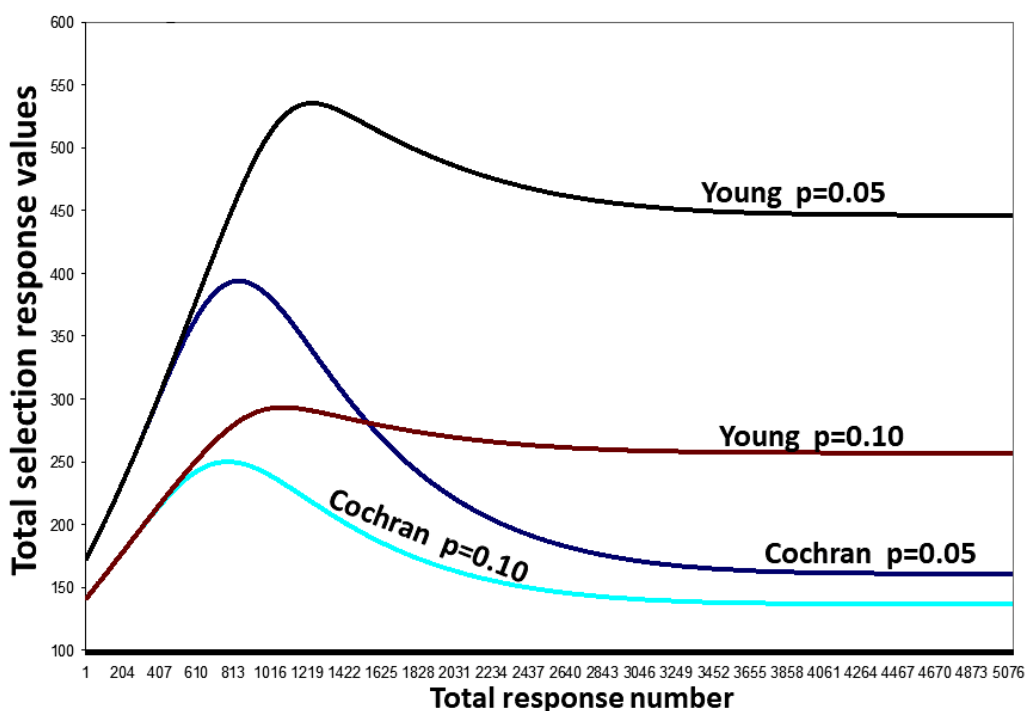
Index	$p$	$u_1$	$u_2$	$q_1$	$q_2$	$k_1$	$k_2$	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$\hat{R}$
OMLPSI	0.05	1.21	0.15	0.11	0.44	1.69	0.90	67.14	23.19	90.33	88.72
	0.10	0.87	-0.04	0.19	0.52	1.42	0.77	56.27	20.72	76.99	75.47
	0.20	0.45	-0.28	0.33	0.61	1.10	0.63	43.77	17.84	61.61	60.21
	0.30	0.15	-0.46	0.44	0.68	0.89	0.53	35.38	15.79	51.18	49.84
DMLPSI	0.05	0.66	0.86	0.26	0.20	1.26	1.41	49.81	60.79	110.60	88.72
	0.10	0.38	0.57	0.35	0.28	1.05	1.19	41.77	51.35	93.12	75.47
	0.20	0.04	0.22	0.48	0.41	0.82	0.94	32.62	40.57	73.19	60.21
	0.30	-0.21	-0.04	0.58	0.51	0.67	0.78	26.51	33.35	59.86	49.84

The estimated OMLPSI accuracies for predicting the net genetic merit at Stages 1 and 2 were obtained as  $\hat{\rho}_1 = \sqrt{\hat{\mathbf{b}}_1' \hat{\mathbf{Q}}_{11} \hat{\mathbf{b}}_1} / \sqrt{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}} = 39.647/116.953 = 0.339$  and  $\hat{\rho}_2 = \sqrt{\hat{\mathbf{b}}_2' \hat{\mathbf{P}}^* \hat{\mathbf{b}}_2} / \sqrt{\mathbf{w}' \hat{\mathbf{C}}^* \mathbf{w}} = 25.862/111.246 = 0.232$ , respectively, where  $\sqrt{\hat{\mathbf{b}}_1' \hat{\mathbf{Q}}_{11} \hat{\mathbf{b}}_1} = 39.647$  and  $\sqrt{\hat{\mathbf{b}}_2' \hat{\mathbf{P}}^* \hat{\mathbf{b}}_2} = 25.862$  were the estimated standard deviations of  $\hat{I}_1$  and  $\hat{I}_2$ , respectively, whereas  $\sqrt{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}} = 116.953$  and  $\sqrt{\mathbf{w}' \hat{\mathbf{C}}^* \mathbf{w}} = 111.246$  were the estimated standard deviations of the net genetic merit without adjusting matrix  $\hat{\mathbf{C}}$  and after matrix  $\hat{\mathbf{C}}$  was adjusted ( $\hat{\mathbf{C}}^*$ ). That is, the estimated OMLPSI accuracy decreased from Stage 1 to Stage 2. This means that the adjusted matrices ( $\hat{\mathbf{P}}^*$  and  $\hat{\mathbf{C}}^*$ ) for prior selection on  $\hat{I}_1 = \hat{\mathbf{b}}_1 \mathbf{x}_1$  (Cerón-Rojas and Crossa, 2018, Chapter 9, Eq. [9.5] and [9.6]) negatively affected OMLPSI accuracy at Stage 2.

### Estimated DMLPSI Parameters

At Stages 1 and 2, the estimated vectors of coefficients were  $\hat{\beta}'_1 = [1.963 \ -1.139]$  and  $\hat{\beta}'_2 = [0.186 \ 0.105 \ 2.508 \ -0.733]$ , respectively. Thus, the estimated DMLPSIs were  $\hat{I}_1 = \hat{\beta}'_1 \mathbf{x}_1$  and  $\hat{I}_2 = \hat{\beta}'_2 \mathbf{y}$ , where  $\mathbf{x}'_1 = [\text{RL} \ \text{SM}]$  and  $\mathbf{y}' = [\mathbf{x}'_1 \ \text{EW} \ \text{BW}]$  were the vectors of observations. The standard deviations of the variance of  $\hat{I}_1 = \hat{\beta}'_1 \mathbf{x}_1$  and  $\hat{I}_2 = \hat{\beta}'_2 \mathbf{y}$  were  $\sqrt{\hat{\beta}'_1 \hat{\mathbf{Q}}_{11} \hat{\beta}_1} = 39.647$  and  $\sqrt{\hat{\beta}'_2 \hat{\mathbf{P}}^* \hat{\beta}_2} = 16.662$ . Because the DMLPSI values were independent between stages, to estimate the DMLPSI parameters at Stage 2, we did not adjust matrices  $\hat{\mathbf{P}}$  and  $\hat{\mathbf{C}}$  for prior selection on  $\hat{I}_1 = \hat{\beta}'_1 \mathbf{x}_1$ .

The values for truncation points ( $u_1 = 0.66$  and  $u_2 = 0.86$ ), proportions retained ( $q_1 = 0.26$  and  $q_2 = 0.20$ ) and selection intensities ( $k_1 = 1.26$  and  $k_2 = 1.41$ ) at Stages 1 and 2 for  $p = q_1 q_2 = 0.05$ , were those associated with the maximum  $\hat{R}_t = 110.60$  value (Fig. 3), which were obtained following the iterative process described by Xu



**Fig. 4.** Distribution of the total estimated optimum multistage linear phenotypic selection index (OMLPSI) response values, for a real dataset, when the selection intensities were obtained with the Cochran (1951) and Young (1964) method.

and Muir (1992). In this case, the maximum estimated total DMLPSI selection response (110.60) was 24.66% higher than the estimated single-stage LPSI selection response (88.72). Thus, for this real dataset, the maximum estimated total DMLPSI selection response was different from the estimated single-stage LPSI selection response.

In Table 1, we present additional truncation points, proportions retained, selection intensities, and estimated selection response values for  $p = q_1q_2 = 0.10, 0.20$ , and  $0.30$  obtained at Stages 1 and 2. In these cases, the maximum estimated total DMLPSI selection responses were 23.39, 21.56, and 20.10%, respectively, higher than the estimated LPSI selection responses (Table 1).

The foregoing results indicated that for this real dataset, the average of the estimated total DMLPSI selection response was 22.80% higher than the average of the estimated single-stage LPSI selection response (68.56), whereas the average of the estimated total OMLPSI selection response was only 2.21% higher than the estimated single-stage LPSI selection response for  $p = 0.05, 0.10, 0.20$ , and  $0.30$ . Xu and Muir (1991, 1992) indicated that the loss of efficiency in the DMLPSI response is justified because their method for obtaining the selection intensities and total responses gives the breeder the opportunity to implement an unlimited number of selection stages, which otherwise would be very difficult or impossible to do.

The estimated accuracies for predicting the net genetic merit for both stages were  $\hat{r}_1 = \sqrt{\hat{\mathbf{b}}_1' \hat{\mathbf{Q}}_{11} \hat{\mathbf{b}}_1} / \sqrt{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}} = 39.647/116.953 = 0.339$  and  $\hat{r}_2 = \sqrt{\hat{\beta}_2' \hat{\mathbf{P}} \hat{\beta}_2} / \sqrt{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}} = 16.662/116.953 = 0.143$  where  $\sqrt{\hat{\beta}_1' \hat{\mathbf{Q}}_{11} \hat{\beta}_1} = 39.647$  and  $\sqrt{\hat{\beta}_1' \hat{\mathbf{Q}}_{11} \hat{\beta}_1} = 39.647$  were the estimated standard deviations of the DMLPSI at Stages 1 and 2, respectively, whereas  $\sqrt{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}} = 116.953$  was the estimated standard deviation of the net genetic merit when matrix  $\hat{\mathbf{C}}$  was not adjusted. At Stage 1,  $\mathbf{S}_{i(i-1)} = \mathbf{Q}_{i(i-1)} \mathbf{B}_{i-1} = 0$  (Appendix A, Eq. [A6] to [A8]) and  $\beta_i = \mathbf{b}_i = \mathbf{Q}_{ii}^{-1} \mathbf{A}_i \mathbf{w}$ , meaning the estimated DMLPSI and OMLPSI vectors of coefficients were the same, from where the estimated OMLPSI and DMLPSI accuracies were also the same because this parameter was not affected by the selection intensity. However, at Stage 2, DMLPSI accuracy decreased due to the restriction imposed on DMLPSI to make its values independent among stages (Appendix A, Eq. [A6] to [A8], for details).

### Accuracy for Predicting the Net Genetic Merit

The estimated OMLPSI accuracies for both stages were  $\hat{\rho}_1 = 0.339$  and  $\hat{\rho}_2 = 0.232$ , whereas  $\hat{r}_1 = 0.339$  and  $\hat{r}_2 = 0.143$  were the estimated DMLPSI accuracies. The estimated accuracies of both indices decreased from Stage 1 to Stage 2. The estimated OMLPSI accuracies decreased because matrices  $\hat{\mathbf{P}}$  and  $\hat{\mathbf{C}}$  were adjusted for prior selection (Cerón-Rojas and Crossa, 2018, Chapter 9, Eq. [9.5] and [9.6]); however, DMLPSI accuracy decreased due to

the restriction imposed on the DMLPSI vector of coefficients at Stage 2.

To compare the efficiency of the indices, we used Eq. [7]:  $\varphi_i = 100[(\hat{\rho}_i/\hat{r}_i) - 1]$ ,  $i = 1, 2$ . The OMLPSI and DMLPSI efficiencies at Stage 1 were the same; however, at Stage 2, OMLPSI efficiency was 62.24% higher than DMLPSI efficiency for predicting the net genetic merit. This means that the OMLPSI was a better predictor of the net genetic merit than the DMLPSI at Stage 2.

### Truncation Points, Proportion Retained, and Selection Intensities

Table 1 presents the OMLPSI and DMLPSI truncation points ( $u_1$  and  $u_2$ ), proportions retained ( $q_1$  and  $q_2$ ), and selection intensities ( $k_1$  and  $k_2$ ) for  $p = q_1q_2 = 0.05, 0.10, 0.20$ , and  $0.30$  in a two-stage context. Because the  $u_1$ ,  $u_2$ ,  $q_1$ ,  $q_2$ ,  $k_1$ , and  $k_2$  and values for OMLPSI and DMLPSI were obtained with a different method, those values were different for both indices. Nevertheless, when the  $p = q_1q_2$  values changed from 0.05 to 0.30, the  $u_1$  and  $u_2$  values decreased, the  $q_1$  and  $q_2$  values increased, and the  $k_1$  and  $k_2$  values decreased in both indices, as we would expect. That is, when the total proportion retained increased from 0.05 to 0.30, the selection intensity decreased, also as we would expect (Fig. 2). In addition, while the DMLPSI  $q_1$  and  $q_2$  values were very similar at both stages, the OMLPSI  $q_1$  values were lower than the  $q_2$  values, which implies that the OMLPSI selection intensity and selection response were different from the DMLPSI selection intensity and selection response (Table 1).

### Simulated Data

#### Predicting the True Selection Response

Table 2 presents the estimated OMLPSI and DMLPSI selection responses ( $\hat{R}_1$ ,  $\hat{R}_2$ , and  $\hat{R}_t = \hat{R}_1 + \hat{R}_2$ ) and the true LPSI selection response ( $R$ ) values obtained in a two-stage context for four simulated selection cycles and  $p = q_1q_2 = 0.01, 0.10$ , and  $0.20$ . For  $p = 0.01$  and  $0.10$ , the average of the estimated total OMLPSI selection response ( $\hat{R}_t = \hat{R}_1 + \hat{R}_2$ ) was 61.70 and 6.18%, respectively, higher than the average of the true selection response (15.22), whereas for  $p = 0.20$ , the average of the estimated total OMLPSI selection response was 15.37% lower than the average of the true selection response. Thus, for this dataset, the best OMLPSI prediction of the mean of the net genetic merit resulted when  $p = 0.10$ , and wrong OMLPSI predictions resulted when  $p = 0.01$  or  $0.20$ .

For  $p = 0.01$ , the average of the estimated total DMLPSI selection response was 42.71% higher than the average of the true selection response; however, for  $p = 0.10$  and  $0.20$ , the average of the estimated total DMLPSI selection response was 5.32 and 27.0% lower than the average of the true selection response. Therefore, for this dataset, the best DMLPSI prediction of the mean of the net genetic merit resulted when  $p = 0.10$ , and wrong



**Table 2. Simulated data for estimated optimum and decorrelated multistage linear phenotypic selection index (OMLPSI and DMLPSI, respectively) selection responses ( $\hat{R}_1$ ,  $\hat{R}_2$ , and  $\hat{R}_t = \hat{R}_1 + \hat{R}_2$ ) for four simulated selection cycles and  $p = q_1q_2 = 0.01$ , 0.10, and 0.20 under a two-stage breeding scheme, and true linear phenotypic selection index selection response ( $R$ ) values.**

Index	Cycle	$p = 0.01$			$p = 0.10$			$p = 0.20$			True response
		$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$\hat{R}_1$	$\hat{R}_2$	$\hat{R}_t$	$R$
OMLPSI	1	18.26	10.70	28.96	11.58	7.38	18.96	9.01	6.09	15.09	17.84
	2	15.53	9.28	24.81	9.80	6.47	16.28	7.60	5.38	12.98	15.66
	3	13.59	9.26	22.85	8.58	6.38	14.96	6.65	5.26	11.91	14.44
	4	16.22	5.62	21.84	10.28	4.15	14.43	7.98	3.58	11.56	14.29
	Avg.	15.90	8.71	24.61	10.06	6.10	16.16	7.81	5.08	12.88	15.22
DMLPSI	1	21.76	3.87	25.63	14.49	2.12	16.61	11.66	1.49	13.15	17.84
	2	17.86	4.60	22.45	11.86	2.64	14.50	9.53	1.92	11.45	15.66
	3	15.98	4.35	20.33	10.61	2.52	13.13	8.52	1.84	10.36	14.44
	4	17.63	3.03	20.66	11.74	1.65	13.39	9.45	1.15	10.60	14.29
	Avg.	17.77	3.95	21.72	12.18	2.23	14.41	9.50	1.61	11.11	15.22

DMLPSI predictions resulted when  $p = 0.01$  or 0.20. That is, DMLPSI and OMLPSI were similar when  $p = 0.10$ .

Based on the foregoing results, the best OMLPSI and DMLPSI predictions of the true selection response resulted when  $p = 0.10$ , and wrong OMLPSI and DMLPSI predictions resulted when  $p = 0.01$ . This means that for the simulated data, the indices' efficiency depended on the total proportion retained ( $p$ ). Thus, for the simulated data, when  $p = 0.10$ , both indices were more efficient for predicting the true mean of the net genetic merit than when  $p = 0.01$  or 0.20. We believe that the results obtained with the simulated data were due to the number of genotypes used to estimate the parameter. Thus, whereas in the real dataset the number of genotypes was 3330, in the simulated the number of genotypes was 500. That is, in the simulated dataset, we used only 15% of the genotypes used in the real dataset to estimate the index parameters.

### Accuracies for Predicting the Net Genetic Merit

Table 3 presents the estimated values of the OMLPSI ( $\hat{\rho}_1$  and  $\hat{\rho}_2$ ) and DMLPSI ( $\hat{r}_1$  and  $\hat{r}_2$ ) accuracies for predicting the net genetic merit in a two-stage context for four simulated cycles. The averages of the estimated OMLPSI and DMLPSI accuracies at Stage 1 were the same. At Stage 2, however, due to the restriction imposed on the covariance between the DMLPSI values (Appendix A,

**Table 3. Estimated values of optimum ( $\hat{\rho}_1$  and  $\hat{\rho}_2$ ) and decorrelated ( $\hat{r}_1$  and  $\hat{r}_2$ ) multistage linear phenotypic selection index (OMLPSI and DMLPSI, respectively) accuracies to predict the net genetic merit under a two-stage (each stage denoted by 1 and 2) breeding scheme for four simulated cycles.**

Cycle	OMLPSI		DMLPSI	
	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{r}_1$	$\hat{r}_2$
1	0.81	0.97	0.81	0.40
2	0.77	0.88	0.77	0.43
3	0.75	0.88	0.75	0.43
4	0.78	0.87	0.78	0.38
Avg.	0.78	0.90	0.78	0.41

Eq. [A6] to [A8]), the average of the estimated DMLPSI accuracies was lower than the average of the estimated OMLPSI accuracies (Table 3). According to Eq. [7], at Stage 2 the average of the estimated OMLPSI accuracies was 119.512% higher than the average of the estimated DMLPSI accuracies for four simulated selection cycles. Thus, for this dataset, the OMLPSI was a better predictor of the net genetic merit than the DMLPSI.

## DISCUSSION

### The Criteria of the Relative Efficiency of the Indices

The estimated OMLPSI and DMLPSI accuracies for predicting the genetic merit and the assumption that the estimated total OMLPSI and DMLPSI selection response must be lower than or equal to the estimated LPSI selection response (Young 1964; Saxton, 1983) were the criteria for evaluating the relative efficiency of both indices. The two criteria were dependent on the method used to estimate the vector of coefficients of each index. The estimated total selection response predicts the mean value of the net genetic merit in the progeny population, whereas the estimated accuracy indicates how close the estimated index values were to the unknown net genetic merit values.

At Stage 2, the OMLPSI variances reduced their size due to the correction imposed on the variance-covariance matrices **P** and **G**, which affected not only the selection responses, but also their accuracies. However, DMLPSI efficiency at Stage 2 was affected not by the adjusted covariance matrices, but by the restrictions imposed on their covariance values among stages.

### Total OMLPSI and DMLPSI Selection Responses

For the real dataset, the average of the estimated total OMLPSI selection responses was only 2.21% higher than the average of the estimated LPSI selection response for all  $p$  values. In addition, the average of the estimated total DMLPSI selection responses was 22.80% higher than the

average of the estimated LPSI selection responses, for all  $p$  values. Thus, for this real dataset, we can expect the total estimated OMLPSI selection response to be a better estimator of the mean of the net genetic merit than the total estimated DMLPSI selection responses.

For the simulated dataset, the best OMLPSI and DMLPSI predictions of the true selection response resulted when  $p = 0.10$ , and the wrong OMLPSI and DMLPSI predictions resulted when  $p = 0.01$ . However, note that although the average of the estimated total OMLPSI selection response overestimated the true selection response by 6.18%, the average of the estimated total DMLPSI selection response underestimated the true selection response by 5.32%. However, for  $p = 0.01$ , both indices overestimated the true response, but while the OMLPSI overestimated the true selection response by 61.70%, the DMLPSI overestimated the true selection response by 42.71%. Thus, for the simulated data, the total estimated selection response of both indices depended on the total proportion retained. We attributed the results obtained with the simulated data to the number of genotypes used to estimate the parameters. That is, in the real dataset, the number of genotypes was 3330, but in the simulated data, the number of genotypes was only 500, which represents only 15% of the size of the genotypes used in the real dataset to estimate the parameters of the indices. This means that the number of genotypes used to estimate the indices' parameters was an important factor for both indices in the real and simulated data.

### OMLPSI and DMLPSI Accuracies

In this case, for the real and simulated datasets, the estimated OMLPSI accuracies were higher than the estimated DMLPSI accuracies at Stage 2. For the real and simulated datasets, at Stage 1, the estimated accuracy of both indices was the same. At Stage 2, however, for the real data, the estimated OMLPSI accuracy was 62.24% higher than the estimated DMLPSI accuracy for predicting the net genetic merit at Stage 2, whereas for the simulated data, the average of the estimated OMLPSI accuracies was 119.51% higher than the average of the estimated DMLPSI accuracies at Stage 2. Thus, based on the estimated accuracies of both indices, we can expect the OMLPSI to be a better predictor of the net genetic merit than the DMLPSI after Stage 1.

### The Method for Obtaining the OMLPSI Selection Intensity

The method used in this work to obtain the OMLPSI selection intensities in a two-stage context is simple and can be programmed in a computer using an R code. This method did not overestimate the selection intensities, as the Cochran (1951) and Young (1964) methods did. Thus,

the proposed method was useful for obtaining the selection intensity values of OMLPSI in a two-stage context.

### The Restrictions DMLPSI Imposed on the Covariance Values

The DMLPSI imposed the restriction that the covariance between DMLPSI values at different stages be zero. This restriction was to ensure the existence of solutions for the truncation points at different stages without resorting to numerical multiple integration (Xu and Muir, 1991, 1992; Xie et al., 1997). However, the restriction decreased the estimated DMLPSI accuracy and could overestimate the DMLPSI selection response after Stage 1. Xu and Muir (1991, 1992) indicated that the loss of DMLPSI efficiency after Stage 1 is justified because their method for obtaining the selection intensities and total responses gives the breeder the opportunity to implement an unlimited number of selection stages, which would otherwise be very difficult or impossible to do.

Xu and Muir (1991) indicated that the restriction imposed on the covariance between DMLPSI values is similar to the Kempthorne and Nordskog (1959) restriction imposed on the expected genetic gain per trait, which prevents some traits from changing their mean values while the rest of the trait means remain without restrictions (Cerón-Rojas and Crossa, 2018, Chapter 3). In effect, Xu and Muir (1991, 1992) and Kempthorne and Nordskog (1959) used a projector matrix (e.g.,  $\mathbf{K}$ ) to project the OMLPSI (LPSI) vector of coefficients ( $\mathbf{b}$ ) into a space smaller than the original space of  $\mathbf{b}$ . The reduction of the space into which the Kempthorne and Nordskog (1959) matrix projects  $\mathbf{b}$  is equal to the number of zeros that appears on the expected genetic gain per trait, and the selection response and accuracy decrease as the number of restrictions increases (Cerón-Rojas and Crossa 2018, Chapter 3). However, it is not clear if under the Xu and Muir (1991, 1992) restrictions the selection response and accuracy decrease as the number of stages increases. If this were true, the Xu and Muir (1992) method would not give the breeder the opportunity to implement an unlimited number of stages, because the selection response and accuracy would decrease as the number of stages increases and soon would be null. For example, Xie et al. (1997) compared the estimated single-stage LPSI selection response with the estimated DMLPSI selection response for two and three stages and found that at Stages 2 and 3, the estimated total DMLPSI selection response explained only 92 and 87%, respectively, of the estimated LPSI selection response. That is, at Stage 3, the estimated total DMLPSI selection response was lower (5%) than at Stage 2. In addition, Xie et al. (1997) also indicated that under certain circumstances (they did not specify which), the estimated total DMLPSI selection response could be higher than the estimated single-stage LPSI selection response, as we found in this work when we used the

simulated data for  $p = 0.01$ . However, this is not a good result because the DMLPSI overestimated the true LPSI selection response.

### Additional Approaches of the OMLPSI and DMLPSI

Saxton (1983) and Ayyagari et al. (1985) applied the OMLPSI to five pig and poultry traits, respectively, in a similar manner as we did in this work; however, they obtained the selection intensities in a different way. Saxton (1983) applied a two-stage selection scheme in two ways: first, by selecting three traits and then two traits; and second, by first selecting the last two traits and later the first three traits. Under the first scheme, Saxton (1983) found that the estimated total selection response overestimated the single-stage LPSI response by 3.8%, but under the second, he found that the estimated total selection response overestimated the single-stage LPSI response by only 1.5%. These results were very similar to the results obtained with the OMLPSI in this work when we used real data. Ayyagari et al. (1985) developed six selection schemes with five poultry traits and, in all cases, they underestimated the single-stage LPSI response. That is, the average of the total selection response explained only 70.5% of the estimated LPSI response. These results were not in agreement with the OMLPSI results when we used real data. We believe that these results were due to the way Ayyagari et al. (1985) obtained the selection intensities. Cerón-Rojas and Crossa (2018, Chapter 9) applied the OMLPSI to six chicken traits (Hicks et al., 1998) in a two-stage context. However, these authors used the Young (1964) method to obtain the selection intensities for two stages; thus, the estimated selection response and expected genetic gain per trait values were only approximated.

Xu and Muir (1992) applied the DMLPSI to four poultry traits in a two-stage context, by first selecting one trait and then three traits. They found that the estimated total DMLPSI selection explained 90% of the estimated single-stage LPSI response. Xie et al. (1997) found similar results when they compared the estimated single-stage LPSI selection response with the estimated DMLPSI selection response at Stages 2 and 3. They found that at Stages 2 and 3, the estimated total DMLPSI selection response explained 92 and 87%, respectively, of the estimated LPSI selection response. That is, in all cases, the estimated total DMLPSI selection response could not explain all the estimated single-stage LPSI selection response.

Results of this study are the first ones comparing (with real and simulated data) the relative efficiency of the OMLPSI with DMLPSI efficiency using the total selection response and accuracy as the main criteria to compare the efficiency of the two indices.

## CONCLUSIONS

We evaluated the relative efficiency of two multistage linear phenotypic selection indices. We determined the efficiency of both indices based on the estimated total selection response and accuracy of each index using a real and a simulated dataset. In both datasets, we found that the OMLPSI was a better predictor of the net genetic merit than the DMLPSI. Therefore, breeders should not use the DMLPSI when performing multistage phenotypic selection.

### Author Contribution Statement

J.J. Cerón-Rojas evaluated the statistical methods, wrote the R codes used to estimate all the parameters, conducted the data analysis, and wrote the manuscript. J. Crossa and F.H. Toledo reviewed and corrected the manuscript according to reviewer and editor suggestions. All authors read and approved the final manuscript.

### References

- Akbar, M.K., C.Y. Lin, N.R. Gyles, J.S. Gavora, and C.J. Brown. 1984. Some aspects of selection indices with constraints. *Poult. Sci.* 63:1899–1905. doi:10.3382/ps.0631899
- Ayyagari, V., S.C. Mohapatra, G.S. Bisht, T.S. Thiyyagasundaram, and D.C. Johari. 1985. Efficiency of one and two-stage selection indexes for improvement of net economic merit in white Leghorns. *Theor. Appl. Genet.* 70:166–171. doi:10.1007/BF00275317
- Börner, V., and N. Reinsch. 2012. Optimising multistage dairy cattle breeding schemes including genomic selection using decorrelated or optimum selection indices. *Genet. Sel. Evol.* 44:1–11. doi:10.1186/1297-9686-44-1
- Cerón-Rojas, J.J., J. Crossa, V.N. Arief, K. Basford, J. Rutkoski, D. Jarquín, et al. 2015. A genomic selection index applied to simulated and real data. *G3: Genes, Genomes, Genet.* 5:2155–2164. doi:10.1534/g3.115.019869
- Cerón-Rojas, J.J., and J. Crossa. 2018. Linear selection indices in modern plant breeding. Springer, Cham, the Netherlands. doi:10.1007/978-3-319-91223-3
- Cochran, W.G. 1951. Improvement by means of selection In: J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA. 31 July–12 Aug. 1950. Univ. California Press, Berkeley. p. 449–470. <https://projecteuclid.org/euclid.bsmsp/1200500247> (accessed 2 Apr. 2019).
- Hazel, L.N. 1943. The genetic basis for constructing selection indexes. *Genetics* 8:476–490.
- Hazel, L.N., and J.L. Lush. 1942. The efficiency of three methods of selection. *J. Hered.* 33:393–399. doi:10.1093/oxfordjournals.jhered.a105102
- Hicks, C., W.M. Muir, and D.A. Stick. 1998. Selection index updating for maximizing rate of annual genetic gain in laying hens. *Poult. Sci.* 77:1–7. doi:10.1093/ps/77.1.1
- Kempthorne, O., and A.W. Nordskog. 1959. Restricted selection indices. *Biometrics* 15:10–19. doi:10.2307/2527598
- Podlich, D.W., and M. Cooper. 1998. QU-GENE: A simulation platform for quantitative analysis of genetic models. *Bioinformatics* 14:632–653. doi:10.1093/bioinformatics/14.7.632

- Saxton, A.M. 1983. A comparison of exact and sequential methods in multi-stage index selection. *Theor. Appl. Genet.* 66:23–28. doi:10.1007/BF00281843
- Schott, J.R. 2005. *Matrix analysis for statistics*. 2nd edition. Wiley, Hoboken, NJ.
- Smith, F.H. 1936. A discriminant function for plant selection. *Ann. Eugen.* 7:240–250. doi:10.1111/j.1469-1809.1936.tb02143.x
- Springer, M.D. 1979. *The algebra of random variables*. John Wiley & Sons, New York.
- Xie, C., S. Xu, and J.A. Mosjidis. 1997. Multistage selection indices for maximum genetic gain and economic efficiency in red clover. *Euphytica* 98:75–82. doi:10.1023/A:1003074814916
- Xu, S., T.G. Martin, and W.M. Muir. 1995. Multistage selection for maximum economic return with an application to beef cattle breeding. *J. Anim. Sci.* 73:699–710. doi:10.2527/1995.733699x
- Xu, S., and W.M. Muir. 1991. Multistage selection for genetic gain by orthogonal transformation. *Genetics* 129:963–974.
- Xu, S., and W.M. Muir. 1992. Selection index updating. *Theor. Appl. Genet.* 83:451–458. doi:10.1007/BF00226533
- Young, S.S.Y. 1964. Multi-stage selection for genetic gain. *Heredity* 19:131–145. doi:10.1038/hdy.1964.11

## APPENDIX A

### Phenotypic and Genotypic Matrices for Two Stages

Suppose that there are four economic traits of interest; then  $\mathbf{y}' = [y_1 \ y_2 \ y_3 \ y_4]$  and  $\mathbf{g}' = [g_1 \ g_2 \ g_3 \ g_4]$  are the vectors of observable phenotypic values and unobservable breeding values, respectively. In addition, suppose that at the first stage we have two traits, and at the second stage we have two additional traits available for selection, then  $n_1 = n_2 = 2$  and vector  $\mathbf{y}'$  can be partitioned as  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2]$ , where  $\mathbf{x}'_1 = [y_1 \ y_2]$  and  $\mathbf{x}'_2 = [y_3 \ y_4]$  are the vectors of two traits that become evident at the first and second stages, respectively. At the first stage, the phenotypic covariance matrix of  $\mathbf{x}_1$  ( $\mathbf{P}_1$ ) is

$$\text{Var}(\mathbf{x}_1) = \begin{bmatrix} \text{Var}(y_1) & \text{Cov}(y_1, y_2) \\ \text{Cov}(y_2, y_1) & \text{Var}(y_2) \end{bmatrix} = \mathbf{P}_1$$

whereas the covariance matrix of  $\mathbf{x}_1$  with the vector of true breeding values  $\mathbf{g}$  ( $\mathbf{G}_1$ ) is

$$\begin{aligned} & \text{Cov}(\mathbf{x}_1, \mathbf{g}) \\ &= \begin{bmatrix} \text{Cov}(y_1, g_1) & \text{Cov}(y_1, g_2) & \text{Cov}(y_1, g_3) & \text{Cov}(y_1, g_4) \\ \text{Cov}(y_2, g_1) & \text{Cov}(y_2, g_2) & \text{Cov}(y_2, g_3) & \text{Cov}(y_2, g_4) \end{bmatrix} = \mathbf{G}_1 \end{aligned}$$

For the second stage, in addition to matrix  $\mathbf{P}_1$ , we need the phenotypic covariance matrix between  $\mathbf{x}_1$  and  $\mathbf{x}_2$  ( $\mathbf{P}_{12}$ ) and the phenotypic covariance matrix of  $\mathbf{x}_2$  ( $\mathbf{P}_2$ ). That is, the covariance matrix of phenotypic values at Stage 2 will be

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_2 \end{bmatrix}$$

In a similar manner, in addition to matrix  $\mathbf{G}_1$ , at Stage 2, we need the covariance between  $\mathbf{x}_2$  and  $\mathbf{g}$  ( $\mathbf{G}_2$ ); that is, at Stage 2, the covariance matrix between phenotypic and breeding values can be written as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$$

Matrices  $\mathbf{G}$  and  $\mathbf{C}$  are not exactly the same, because while  $\mathbf{C} = \text{Var}(\mathbf{g})$ ,

$$\mathbf{G} = \begin{bmatrix} \text{Cov}(\mathbf{x}_1, \mathbf{g}) \\ \text{Cov}(\mathbf{x}_2, \mathbf{g}) \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$$

and this last matrix changes at each stage.

### Phenotypic and Genotypic Matrices for $N$ Stages

Now suppose that traits are measured in  $N$  stages ( $n \geq N$ ); then  $\mathbf{y}' = [y_1 \ y_2 \ \dots \ y_n]$  can be partitioned into  $N$  subvectors as  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2 \ \dots \ \mathbf{x}'_N]$ , where  $\mathbf{x}'_1 = [y_1 \ y_2 \ \dots \ y_n]$ ,  $\mathbf{x}'_2 = [y_{(n1+1)} \ y_{(n1+2)} \ \dots \ y_{(n1+n2)}]$ , etc., are subvectors of  $\mathbf{y}'$  for Stages 1, 2, etc. Let  $n_i$  be the number of traits selected at Stage  $i$ , then  $n = \sum_{i=1}^N n_i$  is the total number of traits measured for  $N$  stages. In a similar manner, the phenotypic ( $\mathbf{P}$ ) and genotypic ( $\mathbf{G}$ ) matrices can be partitioned according to the subvectors of  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2 \ \dots \ \mathbf{x}'_N]$  as

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \text{Var}(\mathbf{x}_1) & \text{Cov}(\mathbf{x}_1, \mathbf{x}_2) & \dots & \text{Cov}(\mathbf{x}_1, \mathbf{x}_N) \\ \text{Cov}(\mathbf{x}_2, \mathbf{x}_1) & \text{Var}(\mathbf{x}_2) & \dots & \text{Cov}(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\mathbf{x}_N, \mathbf{x}_1) & \text{Cov}(\mathbf{x}_N, \mathbf{x}_2) & \dots & \text{Var}(\mathbf{x}_N) \end{bmatrix} \quad [\text{A1a}] \\ &= \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & \mathbf{P}_{1N} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \dots & \mathbf{P}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{N1} & \mathbf{P}_{N2} & \dots & \mathbf{P}_{NN} \end{bmatrix} \end{aligned}$$

and

$$\mathbf{G} = \begin{bmatrix} \text{Cov}(\mathbf{x}_1, \mathbf{g}) \\ \text{Cov}(\mathbf{x}_2, \mathbf{g}) \\ \vdots \\ \text{Cov}(\mathbf{x}_N, \mathbf{g}) \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_N \end{bmatrix} \quad [\text{A1b}]$$

respectively, where  $\text{Cov}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{P}_{ij}$  is the  $ij$ th submatrix of  $\mathbf{P}$  and  $\text{Cov}(\mathbf{x}_i, \mathbf{g}) = \mathbf{G}_i$  is the  $i$ th submatrix of  $\mathbf{G}$  at Stage  $i$  (Xu and Muir, 1992).

### The DMLPSI for Selecting One Trait at Each Stage

In this subsection, we present the procedure described originally by Xu and Muir (1991), which is useful to



understand the DMLPSI theory in the multitrait selection context. Let  $\mathbf{y}' = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_n]$  be a  $1 \times n$  vector of trait phenotypic values, where  $n$  denotes the total number of traits in which the breeder is interested. Let  $\mathbf{P}$  be an  $n \times n$  symmetric and positive definite covariance matrix of trait phenotypic values. Then, by the Cholesky decomposition method (Schott, 2005), there is a unique upper triangular matrix  $\mathbf{T}$  such that  $\mathbf{P}$  can be written as  $\mathbf{P} = \mathbf{T}'\mathbf{T}$ , where  $\mathbf{T}'$  is the transposed matrix of  $\mathbf{T}$ . From this result, vector  $\mathbf{y}' = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_n]$  can be transformed into a new vector  $\mathbf{x}' = [x_1 \ x_2 \ \dots \ x_N]$  as

$$\mathbf{x} = (\mathbf{T}')^{-1}\mathbf{y} \quad [\text{A2a}]$$

where

$$(\mathbf{T}')^{-1} = \begin{bmatrix} t_{11} & 0 & \dots & 0 \\ t_{21} & t_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{bmatrix}$$

is the inverse of matrix  $\mathbf{T}'$ . The variance of  $\mathbf{x}$  is  $\text{Var}(\mathbf{x}) = (\mathbf{T}')^{-1}\text{Var}(\mathbf{y})\mathbf{T}^{-1} = \mathbf{I}_n$ , where  $\mathbf{I}_n$  is an identity matrix of size  $n \times n$ . This means that the elements of the transformed vector  $\mathbf{x}$  are independent. Note that Eq. [A2a] is equal to

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = (\mathbf{T}')^{-1} \mathbf{y} = \begin{bmatrix} t_{11}\gamma_1 & 0 & \dots & 0 \\ t_{21}\gamma_1 & t_{22}\gamma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1}\gamma_1 & t_{n2}\gamma_2 & \dots & t_{nn}\gamma_n \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} \quad [\text{A2b}]$$

That is, each element of vector  $\mathbf{x}' = [x_1 \ x_2 \ \dots \ x_N]$  is an index value ( $x_i = I_i$ ). In addition, as  $\text{Var}(\mathbf{x}) = \mathbf{I}_n$ , the covariance of any two indices is null and the index at Stage  $i$  is

$$I_i = \sum_{j=1}^{n_i} t_{ij}\gamma_{ij} \quad [\text{A3}]$$

where  $n_i$  is the number of traits at Stage  $i$ , and  $\gamma_{ij}$  is the  $j$ th ( $j = 1, 2, \dots, n$ ) trait measured at Stage  $i$  ( $i = 1, 2, \dots, N$ ). Thus, the overall procedure described in Eq. [A2a] to [A3] is a type of multistage selection index.

The foregoing selection procedure indicates that, at each stage, the breeder can add only one trait to the index and that matrix  $(\mathbf{T}')^{-1}$  contains the coefficients of the indices at each stage. In addition, note that the index of Eq. [A3] does not maximize its correlation with the net genetic merit ( $H = \mathbf{w}'\mathbf{g}$ ), and matrix  $(\mathbf{T}')^{-1}$  does not include the covariance matrix of genetic values (Eq. A1b). Thus, we need an additional procedure to obtain a DMLPSI similar to the LPSI described by Smith (1936) and Hazel (1943).

## The DMLPSI for Selecting Several Traits at Each Stage

Now we extend Eq. [A2a] and [A2b] to select several traits at each stage according to Xu and Muir (1992). Let

$$\mathbf{B}' = \begin{bmatrix} \beta'_1 & 0 & \dots & 0 \\ \beta'_1 & \beta'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta'_1 & \beta'_2 & \dots & \beta'_N \end{bmatrix}$$

be a transforming matrix similar to matrix  $(\mathbf{T}')^{-1}$ , where  $\beta'_i = [\beta_{i1} \ \beta_{i2} \ \dots \ \beta_{in_i}]$  is the DMLPSI vector of coefficients at Stage  $i$  ( $i = 1, 2, \dots, N$ ). With matrix  $\mathbf{B}'$ , we can construct a DMLPSI for each stage as

$$\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix} = \begin{bmatrix} \beta'_1 & 0 & \dots & 0 \\ \beta'_1 & \beta'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta'_1 & \beta'_2 & \dots & \beta'_N \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \quad [\text{A4}]$$

where vector  $\mathbf{x}'_i = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_{n_i}]$  is the  $i$ th ( $i = 1, 2, \dots, N$ ) subvector of the vector of scores  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2 \ \dots \ \mathbf{x}'_N]$ .

## The DMLPSI Vector of Coefficients for Several Traits at Stage $i$

To obtain the DMLPSI vector of coefficients for several traits at Stage  $i$ , we need to rewrite Eq. [A1a] and [A1b], and matrix  $\mathbf{B}'$ , as follows. Let

$$\mathbf{Q}_{ij} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \dots & \mathbf{P}_{1j} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \dots & \mathbf{P}_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{i1} & \mathbf{P}_{i2} & \dots & \mathbf{P}_{ij} \end{bmatrix}, \quad \mathbf{A}_i = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_i \end{bmatrix}, \quad [\text{A5}]$$

$$\text{and } \mathbf{B}'_i = \begin{bmatrix} \beta'_1 & 0 & \dots & 0 \\ \beta'_1 & \beta'_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta'_1 & \beta'_2 & \dots & \beta'_i \end{bmatrix}$$

( $i, j = 1, 2, \dots, N$ ) be the  $ij$ th submatrix of  $\mathbf{P}(\mathbf{Q}_{ij})$ ,  $\mathbf{A}_i$  a submatrix with the genetic covariance matrices until Stage  $i$ , and  $\mathbf{B}'_i$ , a submatrix of  $\mathbf{B}'$  until Stage  $i$ .

With Eq. [A4] and [A5], we need to obtain the DMLPSI vector of coefficients such that the covariance between the DMLPSI values at Stage  $i$  ( $I_i$ ) and at Stage  $j$  ( $I_j$ ,  $i \neq j$ , is null [i.e.,  $\text{Cov}(I_i, I_j) = 0$ ]. This means that we need to minimize the mean squared difference between the DMLPSI ( $I_i = \beta'_i \mathbf{x}_i$ ) and the net genetic merit ( $H = \mathbf{w}'\mathbf{g}$ ) at Stage  $i$  under the restriction  $\text{Cov}(I_i, I_j) = 0$ . Thus, by restriction  $\text{Cov}(I_i, I_j) = 0$ , if  $\mathbf{J}' = [I_1 \ I_2 \ \dots \ I_N]$  is a vector with the DMLPSIs until Stage  $N$  (Eq. [A4]), then  $\text{Var}(\mathbf{J}) = \text{Diag}\{\sigma_{I_i}\}_{N}$ , where  $\text{Diag}\{\sigma_{I_i}\}_{N}$  is a diagonal matrix of

size  $N \times N$ . Restriction  $\text{Cov}(I_i, I_j) = 0$  is the main difference between the OMLPSI and the DMLPSI.

Let

$$\mathbf{J}_{i-1} = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_{i-1} \end{bmatrix} = \begin{bmatrix} \beta'_1 & 0 & \cdots & 0 \\ \beta'_1 & \beta'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta'_1 & \beta'_2 & \cdots & \beta'_{i-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{i-1} \end{bmatrix}$$

be a subvector of Eq. [A4] until Stage  $i - 1$ , where

$$\mathbf{B}'_{i-1} = \begin{bmatrix} \beta'_1 & 0 & \cdots & 0 \\ \beta'_1 & \beta'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta'_1 & \beta'_2 & \cdots & \beta'_{i-1} \end{bmatrix}$$

is a matrix with the DMLPSI vector of coefficients until Stage  $i - 1$ . We need to find the DMLPSI vector of coefficients ( $\beta_i$ ) that minimizes the mean squared difference between  $I_i$  and  $H = \mathbf{w}'\mathbf{g}$   $\{E[(H - I_i)^2]\}$  at Stage  $i$  under the restriction  $\text{Cov}(I_i, \mathbf{J}_{i-1}) = \beta'_i \mathbf{S}_{i(i-1)} = 0$ , where  $\text{Cov}(I_i, \mathbf{J}_{i-1})$  is the covariance between  $I_i$  and  $\mathbf{J}_{i-1}$ ,  $\mathbf{S}_{i(i-1)} = \mathbf{Q}_{i(i-1)} \mathbf{B}_{i-1}$ , and  $\mathbf{Q}_{i(i-1)}$  (Eq. [1b] and [A5]) is a phenotypic covariance matrix with covariance values between the traits selected until Stage  $i - 1$  and the traits selected at Stage  $i$ . Let  $\mathbf{S}'_{i(i-1)} = \mathbf{S}_{(i-1)i} = \mathbf{B}'_{i-1} \mathbf{Q}_{i(i-1)}$  be the transpose of matrix  $\mathbf{S}_{i(i-1)} = \mathbf{Q}_{i(i-1)} \mathbf{B}_{i-1}$ ; then, assuming that matrices  $\mathbf{Q}_{ii}$ ,  $\mathbf{Q}_{i(i-1)}$ ,  $\mathbf{A}_j$ , and the vector of economic weights ( $\mathbf{w}$ ) are known, we will minimize  $E[(H - I_i)^2]$  under the restriction  $\text{Cov}(I_i, I_j) = \beta'_i \mathbf{Q}_{i(i-1)} \mathbf{B}_{i-1} = \mathbf{S}_{(i-1)i} \beta_i = 0$ . That is, it is necessary to minimize the function

$$f(\beta_i, \mathbf{v}) = \beta'_i \mathbf{Q}_{ii} \beta_i - 2\mathbf{w}' \mathbf{A}_i \beta_i + 2\mathbf{v}' \mathbf{S}_{(i-1)i} \beta_i \quad [\text{A6}]$$

with respect to vectors  $\beta_i$  and  $\mathbf{v}' = [v_1 \ v_2 \ \dots \ v_{i-1}]$ , where  $\mathbf{v}$  is a vector of Lagrange multipliers (Schott, 2005). In matrix notation, the derivative results of  $\beta_i$  and  $\mathbf{v}'$  were

$$\begin{bmatrix} \beta_i \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{ii} & \mathbf{S}_{i(i-1)} \\ \mathbf{S}_{(i-1)i} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}_i \mathbf{w} \\ 0 \end{bmatrix} \quad [\text{A7}]$$

from which the DMLPSI vector of coefficients that minimizes Eq. [A6] at Stage  $i$  can be written as

$$\beta_i = \mathbf{K}_i \mathbf{b}_i \quad [\text{A8}]$$

where  $\mathbf{K}_i = [\mathbf{I} - \mathbf{\Psi}_i]$ ,  $\mathbf{\Psi}_i = \mathbf{Q}_{ii}^{-1} \mathbf{S}_{i(i-1)} [\mathbf{S}_{(i-1)i} \mathbf{Q}_{ii}^{-1} \mathbf{S}_{i(i-1)}]^{-1} \mathbf{S}_{(i-1)i}$ , and  $\mathbf{b}_i = \mathbf{Q}_{ii}^{-1} \mathbf{A}_i \mathbf{w}$  is the OMLPSI vector of coefficients (Eq. [5]);  $\mathbf{Q}_{ii}^{-1}$  is the inverse of matrix  $\mathbf{Q}_{ii}$ , and  $\mathbf{I}$  is an identity matrix of the same size as matrix  $\mathbf{Q}_{ii}$ . When restrictions  $\text{Cov}(I_i, I_j) = \beta'_i \mathbf{Q}_{i(i-1)} \mathbf{B}_{i-1} = 0$  are not imposed on Eq. [A6],  $\beta_i = \mathbf{b}_i$ , which is the OMLPSI vector of coefficients (Eq. [5]). By the restriction  $\text{Cov}(I_i, I_j) =$

$\beta'_i \mathbf{Q}_{i(i-1)} \mathbf{B}_{i-1} = 0$ ,  $I_i$  and  $I_{i-1}$  are not correlated, hence the name *decorrelated multistage linear phenotypic selection index*. In addition, note that at Stage 1,  $\mathbf{S}_{i(i-1)} = \mathbf{S}_{1(0)} = \mathbf{Q}_{1(0)} \mathbf{B}_0 = 0$  (because there are no traits before Stage 1),  $\mathbf{\Psi}_i = \mathbf{\Psi}_1 = \mathbf{Q}_{11}^{-1} \mathbf{S}_{1(0)} [\mathbf{S}_{(0)1} \mathbf{Q}_{11}^{-1} \mathbf{S}_{1(0)}]^{-1} \mathbf{S}_{(0)1} = 0$ , and  $\mathbf{K}_i = \mathbf{K}_1 = (\mathbf{I} - 0) = \mathbf{I}$ , where  $\mathbf{I}$  is an identity matrix of the same size as  $\mathbf{Q}_{11}$ . Thus,  $\beta_1 = \mathbf{K}_1 \mathbf{b}_1 = \mathbf{Q}_{11}^{-1} \mathbf{A}_1 \mathbf{w}$ , where  $\mathbf{Q}_{11}^{-1}$  is the inverse of matrix  $\mathbf{Q}_{11}$ ,  $\mathbf{A}_1 = \mathbf{G}_1$ , and  $\text{Cov}(\mathbf{x}_1, \mathbf{g}) = \mathbf{G}_1$ .

## APPENDIX B

### OMLPSI Selection Intensity for Two Stages

In the multistage selection context, it is usual to fix the total proportion to be selected ( $p = \prod_{i=1}^N q_i$ ) before selection is performed and then determine the unknown proportion  $q_i$  for each stage under the restriction  $p = \prod_{i=1}^N q_i$ , where  $N$  is the number of stages (Cochran, 1951; Young, 1964; Xu and Muir, 1992). In the two-stage selection context, we would have  $p = q_1 q_2$ . In this Appendix, we describe a method to obtain the OMLPSI selection intensity, while we obtained the DMLPSI selection intensity according to the Xu and Muir (1992) method.

Let  $I_1 = \mathbf{b}'_1 \mathbf{x}_1$  and  $I_2 = \mathbf{b}'_2 \mathbf{y}$  be the OMLPSI at Stages 1 and 2, respectively, and assume that the indices have bivariate normal distribution. Let  $I_1$  and  $I_2$  be transformed as  $u_1 = (I_1 - \mu_{I_1})/\sigma_{I_1}$  and  $u_2 = (I_2 - \mu_{I_2})/\sigma_{I_2}$  with mean zero and variance 1.0, where  $\mu_{I_1}$  and  $\mu_{I_2}$  are the means, whereas  $\sigma_{I_1}$  and  $\sigma_{I_2}$  are the standard deviations of  $I_1$  and  $I_2$ , respectively. In the OMLPSI context, the selected population has bivariate left truncated normal distribution with probability density function  $h(u_1, u_2) = f(u_1, u_2)/p$ , where  $p = q_1 q_2$ ,

$$f(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\rho_{12}^2}} \exp \left[ -\frac{1}{2(1-\rho_{12}^2)} (u_1^2 + u_2^2 - 2\rho_{12}u_1u_2) \right],$$

and  $\rho_{12}$  is the correlation between  $u_1$  and  $u_2$  (Young, 1964).

Consider the transformations (Springer, 1979)  $v_1 = u_1$  and  $v_2 = (u_2 - \rho_{12}u_1)/\sqrt{1-\rho_{12}^2}$ , with Jacobian  $\mathbf{j}$ , where

$$\mathbf{j}^{-1} = \begin{vmatrix} \frac{\partial v_1}{\partial u_1} & \frac{\partial v_1}{\partial u_2} \\ \frac{\partial v_2}{\partial u_1} & \frac{\partial v_2}{\partial u_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -\rho_{12} & 1 \end{vmatrix} \frac{1}{\sqrt{1-\rho_{12}^2}} = \frac{1}{\sqrt{1-\rho_{12}^2}};$$

$|\cdot|$  denotes the determinant function and  $\partial$  the partial derivatives of  $v_1$  and  $v_2$  with respect to  $u_1$  and  $u_2$ . Thus,

$$v_1^2 + v_2^2 = \frac{u_1^2 + u_2^2 - 2\rho_{12}u_1u_2}{(1-\rho_{12}^2)}$$

and

$$g(v_1, v_2) = |\mathbf{j}| f(u_1, u_2) = \left[ \frac{1}{\sqrt{2\pi}} e^{-0.5v_1^2} \right] \left[ \frac{1}{\sqrt{2\pi}} e^{-0.5v_2^2} \right]$$

The transformations indicate that variables  $v_1$  and  $v_2$  are independent, each with a standard normal distribution.

Variables  $v_1$  and  $v_2$  are associated with the truncation points as  $v_1 = u_1$  and  $u_2 = v_2\sqrt{1-\rho_{12}^2} + \rho_{12}v_1$ . This means that  $u_1$  and  $u_2$  values should be obtained in two steps. First, we obtained the values of  $v_1$  and  $v_2$  from two independent standard normal distributions; then, we obtained the values of  $v_1 = u_1$ ,  $u_2 = v_2\sqrt{1-\rho_{12}^2} + \rho_{12}v_1$ ,  $q_1$ , and  $q_2$ ; finally, we obtained the OMLPSI selection intensity at Stages 1 and 2 as

$$k_1 = \frac{z(u_1)}{q_1} \text{ and } k_2 = \frac{z(u_2)}{q_2} \quad [\text{A9}]$$

respectively, where  $z(u_1) = e^{-0.5u_1^2}/\sqrt{2\pi}$  and  $z(u_2) = e^{-0.5u_2^2}/\sqrt{2\pi}$  are the height of the ordinate of the normal curve at the

lowest values of  $u_1$  and  $u_2$  retained, whereas  $q_1$  and  $q_2$  are the proportions of the population of animals or plants selected at each stage (Fig. 1). Equation [A9] values should be associated with the maximum  $R_i = R_1 + R_2$  value (Fig. 3 and 4), where  $R_1 = k_1\sigma_{I_1}$  and  $R_2 = k_2\sigma_{I_2}$  are the selection responses, whereas  $\sigma_{I_1}$  and  $\sigma_{I_2}$  are the standard deviations of the variance of  $I_1$  and  $I_2$  at Stages 1 and 2, respectively.

The total proportion  $p = q_1q_2$  retained depends on the reproductive rate and longevity of the species under consideration and on whether the population is expanding, stationary, or declining in number. The ordinate  $z(u_1)$  and  $z(u_2)$  values of the normal curve are determined by the proportions ( $q_1$  and  $q_2$ ) retained. The amount of progress is larger as  $p = q_1q_2$  becomes smaller (Fig. 2)—that is, as selection becomes more intense (Hazel and Lush, 1942).