Received: 6 August 2022 | Accepted: 5 November 2022

DOI: 10.1002/csc2.20882

Crop Science

ORIGINAL ARTICLE

Crop Breeding & Genetics

A linear profit function for economic weights of linear phenotypic selection indices in plant breeding

J. Jesus Cerón-Rojas[1] | Manje Gowda[1] | Fernando Toledo[1] | Yoseph Beyene[1] | Alison R. Bentley[1] | Leo Crespo-Herrera[1] | Keith Gardner[1] | Jose Crossa[1,2]

[1]International Maize and Wheat Improvement Center (CIMMYT), Km 45, Carretera Mexico-Veracruz, Texcoco, Edo. de Mexico, Mexico

[2]Colegio de Post-Graduados, Edo. de Mexico, Montecillos, Mexico

Correspondence
Jose Crossa, International Maize and Wheat Improvement Center (CIMMYT), Km 45, Carretera Mexico-Veracruz, Texcoco, Edo. de Mexico, Mexico.
Email: j.crossa@cgiar.org

Assigned to Associate Editor Marilyn Warburton.

Funding information
Bill and Melinda Gates Foundation, Grant/Award Number: INV-003439

Abstract

The profit function (net returns minus costs) allows breeders to derive trait economic weights to predict the net genetic merit (H) using the linear phenotypic selection index (LPSI). Economic weight is the increase in profit achieved by improving a particular trait by one unit and should reflect the market situation and not only preferences or arbitrary values. In maize (*Zea mays* L.) and wheat (*Triticum aestivum*) breeding programs, only grain yield has a specific market price, which makes application of a profit function difficult. Assuming the traits' phenotypic values have multivariate normal distribution, we used the market price of grain yield and its conditional expectation given all the traits of interest to construct a profit function and derive trait economic weights in maize and wheat breeding. Using simulated and real maize and wheat datasets, we validated the profit function by comparing its results with the results obtained from a set of economic weights from the literature. The criteria to validate the function were the estimated values of the LPSI selection response and the correlation between LPSI and H. For our approach, the maize and wheat selection responses were 1,567.13 and 1,291.5, whereas the correlations were .87 and .85, respectively. For the other economic weights, the selection responses were 0.79 and 2.67, whereas the correlations were .58 and .82, respectively. The simulated dataset results were similar. Thus, the profit function is a good option to assign economic weights in plant breeding.

1 | INTRODUCTION

The linear phenotypic selection index (LPSI) is a linear combination of observable random trait phenotypic values (**y**, i.e., $I = \boldsymbol{\beta}' y$). According to Hazel et al. (1994), the LPSI allows breeders to select and jointly improve several traits that differ in additive variability, heritability, economic importance, and in the correlation among their phenotypes and genotypes in plant and animal breeding. Similarly, the net genetic merit (H) is a random unobservable linear combination of true breeding values (**g**) of the traits weighted by their respective economic weights (**w**, i.e., $H = \mathbf{w}'\mathbf{g}$), and when the phenotypic and genotypic covariance matrices are known, the LPSI is the Best Linear Predictor (BLP) of H (Cerón-Rojas & Crossa, 2022; Cochran, 1951; Searle et al., 2006; Teunissen, 2007).

Abbreviations: AD, anthesis day; BLP, best linear predictor; EH, ear height; FIRA, Fideicomiso Instituido en Relación con la Agricultura; GY, grain yield; HD, heading; LPSI, linear phenotypic selection index; MLE, maximum likelihood estimator; MOI, moisture content; PH, plant height; QT, quantitative trait; QTL, quantitative trait locus.

The main LPSI objectives are to predict H, maximize the LPSI selection response, and the correlation between H and LPSI. When the population means of H and $\mathbf{y}$ are zero, the selection response is the conditional expectation of H given LPSI (Cerón-Rojas & Crossa, 2022; Cochran, 1951). The size of the estimated LPSI selection response and the estimated correlation between H and LPSI are the main criteria to validate and compare the efficiency of any LPSI in plant and animal breeding.

The main assumptions of the LPSI theory are: (a) vector $\mathbf{g}$ is composed entirely of the additive effects of genes; (b) H is the total individual genotypic economic value; and (c), the $\mathbf{y}$ and H values have a joint multivariate normal distribution. Based on these assumptions, the LPSI is the conditional expectation of H given $\mathbf{y}$ (Cerón-Rojas & Crossa, 2022; Cochran, 1951), and to predict H we need to estimate the LPSI vector of coefficients ($\boldsymbol{\beta}$) so that the estimated LPSI values may discriminate those individuals with the highest H values (Smith, 1936). Currently, a linear selection index can be a linear combination of marker scores and phenotypic values (Cerón-Rojas et al., 2008; Lande & Thompson, 1990), genomic estimated breeding values (Cerón-Rojas et al., 2015; Cerón-Rojas & Crossa, 2019), and phenotypic and genomic estimated breeding values jointly (Cerón-Rojas & Crossa, 2018, 2020a, 2022; Dekkers, 2007).

The trait economic weight is the increase in profit (net returns minus costs) achieved by improving a particular trait by one unit while the others remain fixed (Charffeddine & Alenda, 1998; Blasco, 2021). It should reflect the market situation and the marginal benefit from one unit of improvement, as opposed to just preferences or simply arbitrarily fixed values (Magnussen, 1990). Strain and Nordskog (1962) proposed using a profit function to integrate the costs and benefits of a breeding program and compare the profitability of lines and crosses. Later, Moav and Moav (1966), and Moav and Hill (1966), used the partial derivatives of the profit function (evaluated in the trait's mean) as economic weights for within-line selection.

The profit function depends on the traits of interest, the market prices of the traits, the production technology, and market conditions (Charffeddine & Alenda, 1998). An additional feature of profit functions is they may be linear or non-linear. Although a non-linear profit function can be a function of phenotypic values of individual plants or a function of population means, the linear profit function is one of the phenotypic trait values of individual plants or animals (Itoh & Yamada, 1988). Nevertheless, an LPSI can still be used for a non-linear profit function, although the optimum index depends on the selection intensity and the number of generations over which the selection response is to be maximized. According to Goddard (1983), an LPSI achieves the greatest increase in profit when the profit function and economic weights are not linear. Thus, an LPSI will always give the optimum selection response for linear and non-linear profit functions.

Core Ideas

- LPSI is the best linear predictor of the net genetic merit .
- The main LPSI objective is to predict the net genetic merit and maximize the selection response.
- The profit function allows breeders to derive trait economic weights to predict H.
- Economic weight is the increase in profit achieved by improving a particular trait by one unit.

Conditions for applying the profit function theory are: (a) changes in profit function should be due to changes in H and not to environmental conditions or changes in technology; (b) prices and costs are constants; and (c) because the genetic gains in each cycle are low, a linear approach to the profit function is recommended (Charffeddine & Alenda, 1998). In this context, the net genetic merit, the estimated LPSI selection response, and the estimated LPSI values should be interpreted in terms of economic gains and costs (Blasco, 2021).

In this research, we describe a profit function to obtain economic weights in the maize (*Zea mays* L.) and wheat (*Triticum aestivum*) breeding context. Contrary to animal breeding, where the traits of economic interest have a specific price on the market, in maize and wheat breeding only grain yield has a specific market price, which makes it difficult to apply a profit function and obtain economic weights. For this reason, the proposed profit function is based on grain yield market price and on the regression coefficients of grain yield on all the associated traits. In this manner, the grain yield market price is distributed over the other traits as a product of price and the regression coefficient of each trait associated with grain yield. Using seven simulated maize datasets and one real maize and wheat dataset, we validated the profit function by comparing its results with the results obtained from a set of economic weights taken from published literature (Cerón-Rojas et al., 2015; Cerón-Rojas & Crossa, 2019, 2020a, 2020b). The criteria to validate the proposed profit function were the size of the estimated LPSI selection response and the estimated correlation coefficient between LPSI and H. For the simulated and real datasets, we found the profit function described in this work is a good option to assign economic weights in maize and wheat breeding.

As we shall see later, the approach to the profit function theory described in this study is a mathematical formalization of Smith's (1936) idea to assign economic weights to the traits of interest in the wheat breeding context. Finally, the results obtained in this study are the first to use a profit function to derive economic weights in maize and wheat breeding. A short review of the LPSI theory is provided in the Appendix.

2 | MATERIAL AND METHODS

2.1 | Methods

2.1.1 | The Smith (1936) idea for assigning economic weights in wheat breeding

Consider in a wheat selection program we consider the vector $\mathbf{y'} = [Y_1 \ldots Y_t]$ of t traits, where Y_1 denotes grain yield, Y_2 baking quality, Y_3 resistance to flag smut, and so forth. We can evaluate Y_2 and Y_3 in terms of Y_1 as follows. Supposing an advance of 10 in the baking score is equal in value to an advance of one bushel per acre in yield, and that a decrease of 20% infection is worth one bushel of yield, then, taking Y_1 as a standard and units as indicated, $w_1 = 1.0$, $w_2 = 0.1$, and $w_3 = -0.05$ will be the economic values of each trait. This is the Smith's (1936) idea for deriving economic weight in wheat breeding programs. Now, let $\mathbf{y'}_{-1} = [Y_2 \ldots Y_t]$ be a vector of $t-1$ traits that does not include Y_1 and assume that Y_1 is the dependent random variable, whereas $\mathbf{y'}_{-1}$ is the vector of random explanatory variables. To derive economic weights for maize and wheat breeding, we will adapt the above-mentioned Smith's idea to the multiple regression context, that is, $Y_1 = b_0 + \mathbf{b'y}_{-1} + e$ (where e has normal distribution, null expectation, and variance, σ_e^2), using the profit function and the regression theory.

2.1.2 | The cost function

Let $a_1, a_2, \ldots, a_N$ be the N input prices or costs of the N input variables $X_1, X_2, \ldots, X_N$ such as fertilizers, number of cultivated hectares, number of workers per hectare, worker wages per hectare, etc.; then, the cost function is:

$$C = C_0 + a_1 X_1 + a_2 X_2 + \cdots + a_N X_N \quad (1)$$

where C_0 is the fixed cost in the plant breeding program.

2.1.3 | The profit function for grain yield

The main objective for maize and wheat breeding programs is to increase grain yield (Y_1) and to decrease traits such as plant height, days to maturity, and plant diseases, and others. In these programs, only Y_1 has a market price, so we will define the profit function as follows: let π be the price of Y_1 in tons per hectare, and let N_H be the number of hectares cultivated by the breeder; then, the profit function associated with Y_1 is:

$$P = N_H \left(\pi Y_1 - C\right), \quad (2)$$

where $N_H \pi Y_1$ is the net economic return and $N_H C$ is the selection cycle cost of the breeding program. Equation 2 only includes Y_1.

2.1.4 | The profit function for all traits under selection

Suppose each random vector of t traits $\mathbf{y'} = [Y_1 \ldots Y_t]$ (which include the grain yield, Y_1, and all traits associated with Y_1) is independent and identically distributed as a multivariate normal distribution with vector of means $\boldsymbol{\mu'} = [\mu_1 \ldots \mu_t]$ and covariance matrix $\mathbf{P}$, where matrix $\mathbf{P}$ indicates that the elements of $\mathbf{y}$ may be correlated and have different variance. We will include vector $\mathbf{y'}_{-1} = [Y_2 \ldots Y_t]$ (which does not include Y_1) in Equation 2 through the conditional expectation of Y_1 given $\mathbf{y}_{-1}$. Thus, let:

$$Y_1 = b_0 + \mathbf{b'y}_{-1} + e \quad (3)$$

be the multiple linear regression model where e has a normal distribution, null expectation and variance $\sigma_e^2 = \sigma_1^2 - cov'(Y_1, \mathbf{y}_{-1})\mathbf{S}^{-1}cov(Y_1, \mathbf{y}_{-1})$, where σ_1^2 is the variance of Y_1, $cov(Y_1, \mathbf{y}_{-1})$ is the covariance between Y_1 and $\mathbf{y}_{-1}$, and $\mathbf{S}^{-1}$ is the inverse of the covariance matrix of $\mathbf{y}_{-1}$ (S) (Rencher & Schaalje, 2008). In addition, we assumed that the covariance between any pairs of e is 0. Therefore, with Equation 3, the conditional expectation of Y_1 given $\mathbf{y}_{-1}$ is:

$$E\left(Y_1 | \mathbf{y}_{-1}\right) = \mu_1 + \mathbf{b'}\left(\mathbf{y}_{-1} - \mathbf{m}\right), \quad (4)$$

where μ_1 is the mean of Y_1, $\mathbf{m'} = [\mu_2 \ldots \mu_t]$ is the vector of means of $\mathbf{y}_{-1}$, and $\mathbf{b'} = cov'(Y_1, \mathbf{y}_{-1})\mathbf{S}^{-1} = [b_2 \ldots b_t]$ is the vector of regression coefficients. In Equations 3 and 4, the values of $\mathbf{y}_{-1}$ are not under the control of the experimenter and occur randomly along with Y_1 (Rencher, 2002). Thus, on each plant, we observe Y_1 and $\mathbf{y}_{-1}$ jointly.

2.1.5 | The maximum likelihood estimator (MLE) of Equations 3 and 4 is

$$\hat{Y}_1 = \hat{E}\left(Y_1 | \mathbf{y}_{-1}\right) = \hat{\mu}_1 + \hat{\mathbf{b}}'\left(\mathbf{y}_{-1} - \hat{\mathbf{m}}\right), \quad (5)$$

where $\hat{Y}_1$ is the estimated BLP of Y_1, whereas $\hat{\mu}_1$, $\hat{\mathbf{b}}' = \widehat{cov}'(Y_1, \mathbf{y}_{-1})\mathbf{S}^{-1} = [\hat{b}_2 \ldots \hat{b}_t]$ and $\hat{\mathbf{m}}$ are MLE of μ_1, $\mathbf{b}' = cov'(Y_1, \mathbf{y}_{-1})\mathbf{S}^{-1} = [b_2 \ldots b_t]$, and m, respectively (Rencher & Schaalje, 2008). Equation 5 is a function of $\mathbf{y}_{-1}$ alone and not a function of Y_1 (Christensen, 2011), allowing us to write an approximated profit function for t traits as:

$$P_t \approx N_H\left[0.5\pi Y_1 + 0.5\pi \hat{Y}_1 - C\right] = N_H\left\{0.5\pi Y_1 + 0.5\pi\left[\hat{\mu}_1 + \hat{\mathbf{b}}'\left(y_{-1} - \hat{\mathbf{m}}\right)\right] - C\right\} \quad (6)$$

In Equation 6, the symbol "≈" indicates an approximation.

2.1.6 | Deriving and estimating the economic weights

Suppose C is fixed (Equation 1), the partial derivatives of Equation 6 with respect to Y_1 and each trait associated with $\hat{Y}_1$ (Equation 5) are, respectively, $\frac{\partial}{\partial Y_1} P_t = \frac{\pi N_H}{2}$ and $\frac{\partial}{\partial Y_j} P_t = \frac{\pi N_H}{2}\hat{b}_j, j = 2,3,\ldots,t,$ from where the estimated economic weights for Y_1 and Y_j are:

$$\hat{w}_1 = \frac{\pi N_H}{2} \text{ and } \hat{w}_j = \frac{\pi N_H}{2}\hat{b}_j,\ j = 2, 3, \ldots, t, \quad (7)$$

respectively (Goddard, 1983; Moav & Hill, 1966; Moav & Moav, 1966). Therefore, according to Equation 7, an MLE of the vector of economic weights $\mathbf{w}' = [w_1 \ldots w_t]$ (Equation A1 and A2) is:

$$\hat{\mathbf{w}} = \frac{\pi N_H}{2}\begin{bmatrix}1\\ \hat{\mathbf{b}}\end{bmatrix}, \quad (8)$$

where all the components of Equation 8 were defined earlier.

2.1.7 | Estimation of the LPSI and its parameters

According to Pawitan (2013, p. 45), the invariance property of the MLE says: "*If* $\hat{\mathbf{w}}$ *is the MLE of* $\mathbf{w}$ *and* $g(\mathbf{w})$ *is a function of* $\mathbf{w}$*, then* $g(\hat{\mathbf{w}})$ *is the MLE of* $g(\mathbf{w})$." Thus, by this property and by Equation 8, we can assume that an MLE estimator of the LPSI vector of coefficients (Equation A3) is:

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{P}}^{-1}\,\hat{\mathbf{G}}\,\hat{\mathbf{w}} = \frac{\pi N_H}{2}\,\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\begin{bmatrix}1\\ \hat{\mathbf{b}}\end{bmatrix} \quad (9)$$

where $\hat{\mathbf{G}}$ and $\mathbf{P}\hat{}^{-1}$ are restricted MLE (Cerón-Rojas & Crossa, 2018) of the genotypic covariance matrix $\mathbf{G}$ (Equation A1) and the inverse phenotypic covariance matrix of $\mathbf{P}$ (Equation A3), respectively. Equation 9 implies the estimated LPSI (Equation A3) is:

$$\hat{I} = \hat{\boldsymbol{\beta}}'\mathbf{y} \quad (10)$$

which should be interpreted in terms of monetary units when breeders predict H (Equations A1 and A2). The estimated LPSI selection response (Equation A4) is:

$$\hat{R} = k\sqrt{\hat{\boldsymbol{\beta}}'\hat{\mathbf{P}}\hat{\boldsymbol{\beta}}}, \quad (11)$$

where k denotes the selection intensity. Furthermore, because $\frac{\pi N_H}{2}$ appears on the numerator and denominator of the correlation (Equation A5) between LPSI and H, this last parameter is not affected by $\frac{\pi N_H}{2}$ and can be estimated as:

$$\hat{\rho}_{HI} = \frac{\sqrt{\hat{\boldsymbol{\beta}}'\hat{\mathbf{P}}\hat{\boldsymbol{\beta}}}}{\sqrt{\hat{\mathbf{w}}'\hat{\mathbf{G}}\hat{\mathbf{w}}}}. \quad (12)$$

All the terms of Equation 12 were defined earlier.

2.1.8 | Estimating the LPSI parameters

For the set of economic weights taken from published literature (Cerón-Rojas et al., 2015) we estimated the LPSI parameter and we made selection with the simulated and real datasets using "RIndSel," an R software for Index Selection (Alvarado et al., 2018; Pacheco et al., 2017; Perez-Elizalde et al., 2014). In addition, to apply the profit function described in this work, we made an R-code to estimate the economic weights and to make LPSI selection. RIndSel is an R software completely automated, thus, users only need to learn how to introduce their data into the program and how to interpret the results. This software, and a complete user manual, can be downloaded from https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10854

2.2 | Materials

2.2.1 | Simulated datasets

These datasets were simulated by Cerón-Rojas et al. (2015) with QU-GENE software using 2,500 molecular markers and 315 quantitative trait loci (QTLs) for eight phenotypic selection cycles (C0 to C7), each with four traits (Y_1, Y_2, Y_3, and Y_4), 500 genotypes, and four replicates for each genotype. The authors distributed the markers uniformly across 10 chromosomes and the QTLs randomly across the 10 chromosomes to simulate maize (*Zea mays* L.) populations. 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 Y_1 and Y_2, Y_1 and Y_3, Y_1 and Y_4, Y_2 and Y_3, Y_2 and Y_4, and Y_3 and Y_4, respectively. The economic weights assigned by Cerón-Rojas et al. (2015) to Y_1, Y_2, Y_3, and Y_4, were 1, −1, 1, and 1, respectively. Using these economic weights and those obtained with the proposed profit function, we compared the results using seven phenotypic selection cycles (C1 to C7) for the selected top 10% ($k = 1.755$) of the estimated LPSI values in each cycle. In the selection process, we assumed that Y_1 denotes grain yield in kilograms, whereas Y_2, Y_3, and Y_4 may denote plant height, ear height, etc.

2.2.2 | Real datasets

We used one maize and one wheat dataset from CIMMYT breeders' experimental research. The maize traits (grain yield, GY in t ha^{-1}; anthesis day, AD, in days; moisture content, MOI, %; plant height, PH, in cm; and ear height, EH, in cm) were evaluated in five sites. The number of maize genotypes was 68, each with two repetitions, whereas the environment was optimal. The economic weight for each trait was 5, −0.3, −0.3, −0.3 and −0.3, respectively. Likewise, the wheat traits (GY; heading, HD, days; and PH) were evaluated in one environment. The number of wheat genotypes was 100, each with two repetitions, whereas the economic weight for each trait was 5, −0.3, and −0.3, respectively. We compared the maize and wheat results using the selected top 10% ($k = 1.755$) of the estimated LPSI values.

3 | DATA AVAIL ABILITY

3.1 | Simulated and real data

3.1.1 | Simulated data

Simulated population datasets are available in the Application of a Genomics Selection Index to Real and Simulated Data repository at http://hdl.handle.net/11529/10199. The simulated phenotypic datasets are in a folder named Simulated_Data_GSI, which contains two subfolders: Data_Phenotypes_April-26-15 and Haplotypes_GSI_April-26-15. In turn, folder Data_Phenotypes_April-26-15 contains two subfolders: GSI_Phenotypes-05 and PSI_Phenotypes-05. This last folder contains eight Excel datasets, from which we used the following Excel files for the LPSI selection: C1_PSI_05_Pheno, C2_PSI_05_Pheno, C3_PSI_05_Pheno, C4_PSI_05_Pheno, C5_PSI_05_Pheno, C6_PSI_05_Pheno, and C7_PSI_05_Pheno.

3.1.2 | Real data

The real datasets are available by request at j.crossa@cgiar.org.

3.2 | Estimated costs and prices

3.2.1 | Simulated data

Fideicomiso Instituido en Relación con la Agricultura (FIRA), or Trust Established in Relation to Agriculture, is a Mexican government institution that predicts prices and costs associated with agriculture each year. According to FIRA, for all Mexican states in 2022, the average cost per hectare of cultivated maize grain was MX$38,392.75 (Mexican pesos), whereas the average estimated price for each ton of maize grain yield was MX$3,182.00. For all Mexican states, FIRA estimated a maize grain yield of 12 tons per hectare for 2022. Complete FIRA information is available at the link: https://www.fira.gob.mx/Nd/Agrocostos.jsp

Based on FIRA information, we adapted the Cerón-Rojas et al. (2015) simulated maize datasets described earlier, if the harvest corresponded to 1% of a hectare, that is, a plot or cultivated area of 10 by 10 = 100 m^2. Note that 1 ha is equal to 100 by 100 = 10,000 m^2; therefore, to obtain the cost for the simulated datasets, we divided MX$38,392.75 by 100, where the cost for each simulated selection cycle was MX$383.93. In addition, the average of simulated grain yield for Cycle 1 was 161.88 kg; thus, because for 1,000 kg the price is MX$3,182.00, by using the rule of three, we found that for 161.88 kg the grain yield price is MX$515.101. We used a similar approach to obtain prices and costs for the other selection cycles.

3.2.2 | Real data

According to FIRA data, we used MX$3,182.00 for each ton of maize. Nevertheless, for each ton of wheat, the average price estimated by FIRA was MX$5,265.00. To estimate the vector of economic weights (Equation 9) for both real datasets, $N_H = 1.0$ (1 ha).

4 | RESULTS

4.1 | Theoretical results

4.1.1 | Generalized profit function

Note that Equation 6 can be written as

$$P_t \approx N_H \left[\frac{\pi}{n} Y_1 + \frac{(n-1)\pi}{n} \hat{Y}_1 - C \right].$$

This means that when $n = 1$ we shall have Equation 2, and for $n = 2$ we shall have Equation 6. However, when $n = 3$ or $n = 4$, we shall have a profit function that assigns more weight to $\hat{Y}_1$ than to Y_1. Thus, by the above equation, breeders are free to assign weights to $\hat{Y}_1$ and Y_1 according to their interest.

4.1.2 | Some statistical properties of the estimated LPSI parameters

Matrices $\hat{\mathbf{G}}$ and $\mathbf{P}^{\hat{}-1}$ are MLE, and because $\hat{\mathbf{w}}$ is also an MLE, by the MLE invariance property earlier indicated, we

TABLE 1 Cost and price (π, in Mexican pesos), and π time percent of cultivated hectares (1% of N_H) divided by 2 ($\frac{\pi N_H}{2}$), for seven simulated selection cycles

Cycle	Cost	Price	$\frac{\pi N_H}{2}$
1	383.93	515.101	2.576
2	383.93	542.415	2.712
3	383.93	564.924	2.825
4	383.93	587.793	2.939
5	383.93	611.757	3.059
6	383.93	634.598	3.173
7	383.93	656.082	3.280
Average	383.93	587.524	2.938

can assume that the estimated LPSI vector of coefficients ($\hat{\boldsymbol{\beta}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\hat{\mathbf{w}}$, Equation 9) is MLE. The same is true for the estimated correlation between the estimated LPSI ($\hat{\rho}_{HI}$) and the net genetic merit (Equation 12; Table 1), as well as the estimated LPSI selection response ($\hat{R}$, Equation 11; Table 3). Cerón-Rojas and Crossa (2020b) have shown $\hat{\rho}_{HI}$ and $\hat{R}$ are asymptotically unbiased estimators with minimum variance, and they gave methods to obtain confidence intervals for the expectations of $\hat{R}$ and ρ_{HI}; however, until now, the statistical properties of $\hat{\boldsymbol{\beta}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\hat{\mathbf{w}}$ have not been shown completely. By the above results, all the estimated parameters associated with the LPSI are maximum likelihood estimators with minimum variance, and they are asymptotic unbiased estimators for R and ρ_{HI}.

4.1.3 | Three asymptotic statistical properties of the estimator of LPSI

In the Introduction, we indicated some statistical properties of the LPSI when the phenotypic (**P**) and genotypic (**G**) covariance matrices are known. Let $\epsilon = H - \hat{I}$ be the error of prediction of $\hat{I}$ (Equation 10); then: (a) ϵ and $\hat{I}$ are independent; (b) by the central limit theorem (Kollo, 2005), $\hat{I}$ converges in distribution with the normal distribution $N[0, 2(\sigma_I^2)^2/(n-1)]$ (Cerón-Rojas & Crossa, 2020b); and (c) $\hat{I}$ is an asymptotically efficient predictor. This last result allows us to construct a confidence interval for the conditional expectation of H [E(H|I)] as:

$$\hat{I} \pm Z_{\frac{\alpha}{2}} \frac{\sqrt{2}}{\sqrt{n-1}} \hat{\sigma}_I^2$$

where $\hat{I}$ denotes the estimated LPSI values, $Z_{\alpha/2}$ is the upper $100\frac{\alpha}{2}$ percentage point of the standard normal distribution, and $0 \leq \alpha \leq 1$ is the level of confidence. Thus, to establish a $100(1-\alpha) = 95\%$ confidence interval for E(H/I), the value of $Z_{\alpha/2}$ is equal to 1.96. To construct the above confidence interval, it should be convenient to omit the $\frac{\pi N_H}{2}$ from the estimated LPSI parameters because $\frac{\pi N_H}{2}$ increases the length of the interval. The same is true for the variance of the error of prediction. That is, the $\frac{\pi N_H}{2}$ values increase the estimate of the prediction error variance. According to these results, the LPSI theory gives breeders a viable statistical method to make multi-trait selection.

4.2 | Numerical results

4.2.1 | Cost, prices, and economic weights for simulated datasets

Table 1 presents cost (Equation 1), grain yield price (π), and π times the number of hectares cultivated ($N_H = 0.01$ or 1% of a cultivated hectare) and divided by 2 ($\frac{\pi N_H}{2}$; e.g., [(515.101)(0.01)]/2 = 2.576). Although the cost for each selection cycle was the same, the price changed, as this depends on the harvested grain yield in each selection cycle. Similar results were obtained for the $\frac{\pi N_H}{2}$ values because this depends on π.

According to Equations 7–9, the estimated economic weights presented in Table 2 ($\hat{w}_1$ to $\hat{w}_4$) were the product of $\frac{\pi N_H}{2}$ times the coefficient of grain yield Y_1 (1.0 or $\hat{b}_1$) and the estimated coefficients of regression of Y_1 on Y_2, Y_3, and Y_4 ($\hat{b}_2$ to $\hat{b}_4$). For this reason, whereas the economic weight of Y_1 was equal to $\frac{\pi N_H}{2}$ (Table 2), the economic weights for traits Y_2 to Y_4 were $\hat{w}_2 = \frac{\pi N_H}{2}\hat{b}_2$, $\hat{w}_3 = \frac{\pi N_H}{2}\hat{b}_3$, and $\hat{w}_4 = \frac{\pi N_H}{2}\hat{b}_4$, respectively. In addition, because the estimated regression coefficients $\hat{b}_2$ to $\hat{b}_4$ are MLE, by the MLE invariance property (Pawitan, 2013), the estimators $\hat{w}_2$, $\hat{w}_2$, and $\hat{w}_4$ were MLE. This means we can assume the estimator of the vector of economic weights (Equation 8) was a minimum variance and asymptotic unbiased estimator.

4.2.2 | Estimated LPSI selection response for simulated datasets

For $k = 1.755$ (top 10% of the estimated LPSI values), Table 3 presets the estimated selection response obtained when using the profit function ($\hat{R}$) and the estimated selection response ($\hat{R}^*$) when the economic weight were obtained from the published literature. In a similar manner, Table 3 presents the estimated correlation coefficient ($\hat{\rho}_{\mathrm{HI}}$, $\hat{\rho}^*_{\mathrm{HI}}$) between the LPSI and H for seven simulated selection cycles. Although the average of the $\hat{R}$ values was 29.382, the average of the $\hat{R}^*$ values was 14.175 for the seven simulated selection cycles. This

TABLE 2 Coefficient of grain yield Y_1 ($\hat{b}_1$); estimated coefficients of regression of grain yield on traits Y_2, Y_3, and Y_4 ($\hat{b}_2$ to $\hat{b}_4$); estimated economic weights ($\hat{w}_1$ to $\hat{w}_4$) for all four traits; and estimated index coefficients for traits Y_1 to Y_4 ($\hat{\beta}_1$ to $\hat{\beta}_4$) for seven simulated selection cycles

Cycle	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	$\hat{w}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
1	1	−0.60	0.52	0.17	2.58	−1.55	1.35	0.45	1.34	−1.96	1.80	0.73
2	1	−0.28	0.55	0.08	2.71	−0.75	1.50	0.22	1.53	−0.95	1.83	0.35
3	1	−0.43	0.28	0.31	2.82	−1.20	0.78	0.87	1.52	−1.44	1.01	1.16
4	1	−0.44	0.57	0.02	2.94	−1.31	1.67	0.07	1.69	−1.54	1.99	0.15
5	1	−0.45	0.45	0.25	3.06	−1.37	1.38	0.77	1.68	−1.56	1.66	1.09
6	1	−0.30	0.34	0.09	3.17	−0.96	1.08	0.28	1.70	−1.14	1.24	0.42
7	1	−0.29	0.62	−0.04	3.28	−0.94	2.03	−0.12	1.48	−1.19	2.50	−0.06
Average	1	−0.39	0.48	0.13	2.94	−1.15	1.40	0.36	1.56	−1.40	1.70	0.55

TABLE 3 Estimated selection index response ($\hat{R}$, $\hat{R}^*$), for k = 1.755, and estimated correlation coefficient ($\hat{\rho}_{HI}$,$\hat{\rho}^*_{HI}$) between the index and the net genetic merit for seven simulated selection cycles

Cycle	$\hat{R}$	$\hat{R}^*$	$\hat{\rho}_{HI}$	$\hat{\rho}^*_{HI}$
1	33.617	17.808	0.880	0.906
2	28.847	15.690	0.830	0.883
3	27.309	14.219	0.810	0.866
4	31.731	14.219	0.830	0.866
5	30.761	13.638	0.820	0.855
6	26.588	12.038	0.780	0.830
7	26.821	11.612	0.780	0.832
Average	29.382	14.175	0.820	0.863

*This parameter was estimated using the vector of economic weights $\mathbf{w}' = [1\ -1\ 1\ 1]$.

means that the estimated vector of economic weights (Equation 8) affected mainly the estimated LPSI selection ($\hat{R}$), as we would expect.

By Equations 8 and 9, the estimated selection response (Equation 11) can be written as $\hat{R} = \frac{\pi N_H}{2} \quad k\sqrt{\begin{bmatrix} 1 & \hat{\mathbf{b}}' \end{bmatrix}' \hat{\mathbf{G}}\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}} \begin{bmatrix} 1 \\ \hat{\mathbf{b}} \end{bmatrix}}$. This means that $\hat{R}$ is proportional to $\frac{\pi N_H}{2}$, k, and to $\sqrt{\begin{bmatrix} 1 & \hat{\mathbf{b}}' \end{bmatrix}' \hat{\mathbf{G}}\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}} \begin{bmatrix} 1 \\ \hat{\mathbf{b}} \end{bmatrix}}$. Note that when $\frac{\pi N_H}{2}$ and $\sqrt{\begin{bmatrix} 1 & \hat{\mathbf{b}}' \end{bmatrix}' \hat{\mathbf{G}}\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}} \begin{bmatrix} 1 \\ \hat{\mathbf{b}} \end{bmatrix}}$ tend to zero, $\hat{R}$ tends to zero. Therefore, the higher value of $\hat{R}$ will be when $\frac{\pi N_H}{2}$ and $\begin{bmatrix} 1 & \hat{\mathbf{b}}' \end{bmatrix}' \hat{\mathbf{G}}\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}} \begin{bmatrix} 1 \\ \hat{\mathbf{b}} \end{bmatrix}$ will be high. Thus, $\hat{R}$ has one economic part ($\frac{\pi N_H}{2}$) and one genetic part. For the breeding objective, $\begin{bmatrix} 1 & \hat{\mathbf{b}}' \end{bmatrix}' \hat{\mathbf{G}}\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}} \begin{bmatrix} 1 \\ \hat{\mathbf{b}} \end{bmatrix}$ should be high, whereas for the economic objective, $\frac{\pi N_H}{2}$ should be high. For breeding and economic objectives jointly, both parts should be high.

4.2.3 | Estimated LPSI correlation for simulated datasets

The total average of the $\hat{\rho}_{\text{HI}}$ values was 0.820 (Table 3), meaning that the estimated LPSI values predict the H values with high accuracy. However, the total average of the $\hat{\rho}^*_{\text{HI}}$ values was 0.863 (Table 3), that is, the values of $\hat{\rho}^*_{\text{HI}}$ were higher than the $\hat{\rho}_{\text{HI}}$ values. In addition, because $\frac{\pi N_H}{2}$ appears in the numerator and denominator of the correlation between the estimated LPSI and H values (Equation 12), the values of $\hat{\rho}_{\text{HI}}$and $\hat{\rho}^*_{\text{HI}}$ were not affected by the $\frac{\pi N_H}{2}$ values. That is, because $\hat{\rho}_{\text{HI}}$and $\hat{\rho}^*_{\text{HI}}$ are invariant to scale change, they are not affected by changes in economic weights.

4.2.4 | Maize real data results

For this dataset, the ton grain yield price was MX$3182.00, thus, for $N_H = 1.0$, $\frac{\pi N_H}{2} = \frac{(3182)(1)}{2} = 1,591.0$, $\hat{\mathbf{b}}' = [\,1\ 0.174\ -0.030\ 0.030\ 0.013\,]$, $\hat{\mathbf{w}}' = [\,1,591\ 276.23\ -46.65\ 43.76\ 20.30\,]$, and $\hat{\boldsymbol{\beta}}' = [\,276.80\ 512.60\ 55.18\ 2.72\ 52.94\,]$. In addition, for $k = 1.755$, the estimated LPSI selection response was $\hat{R} = 1,567.13$, whereas the estimated correlation between H and the LPSI was $\hat{\rho}_{\text{HI}} = 0.870$, and the estimated LPSI was $\hat{I} = 276.8\text{GY} + 512.6\text{AD} + 55.18\text{MOI} + 2.72\text{PH} + 52.94\text{EH}$. When the economic weights for each trait were 5, −0.3, −0.3, −0.3, and −0.3, respectively, $\hat{\boldsymbol{\beta}}' = [\,0.11\ -0.22\ 0.13\ 0.05\ -0.09\,]$, $\hat{R}^* = 0.79$, and $\hat{\rho}^*_{\text{HI}} = 0.58$. Thus, because $\frac{\pi N_H}{2}$ only affected $\hat{R}$, this was higher than $\hat{R}^*$. Nevertheless, in this case $\hat{\rho}_{\text{HI}}$ was higher than $\hat{\rho}^*_{\text{HI}}$.

4.2.5 | Wheat real data results

For this dataset, the ton grain yield price was MX$5265.00, then, for $N_H = 1.0$, $\frac{\pi N_H}{2} = \frac{(5,265)(1)}{2} = 2,632.50$, $\hat{\mathbf{b}}' = [\,1\ -0.002\ 0.012\,]$, $\hat{\mathbf{w}}' = [\,2,632.5\ -5.30\ 31.60\,]$, $\hat{\boldsymbol{\beta}}' = [\,1,849.30\ -2.30\ 35.12\,]$. In addition, for $k = 1.755$, $\hat{R} = 1291.5$ and $\hat{\rho}_{HI} = 0.85$. When the economic weight for each trait were 5, −0.3, and −0.3, respectively, $\hat{\boldsymbol{\beta}}' = [\,3.30\ -0.25\ 0.18\,]$, $\hat{R}^* = 2.67$, and $\hat{\rho}^*_{\text{HI}} = 0.82$. Once again, because $\frac{\pi N_H}{2}$ affected only $\hat{R}$, this was higher than $\hat{R}^*$, whereas the correlation coefficients were similar.

4.2.6 | The real maize and wheat genotypes selected with LPSI

Table 4 presents the selected maize genotype (with $k = 1.755$), and the means of five selected traits (GY, AD, MOI, PH, and EH) using the economic weights obtained with the profit function and the economic weights 5, −0.3, −0.3, −0.3, and −0.3, respectively. In addition, Table 4 presents the total means of the selected traits, the population mean of the traits, and the selection differential (mean of the selected traits minus the population mean of the traits) for each trait. Table 5 presents the selected wheat genotype (with $k = 1.755$), the means of three selected traits (GY, HD, and PH) using the economic weights obtained with the profit function and the economic weights 5, −0.3 and −0.3, respectively, the total mean of the selected traits, the population mean of the traits, and the selection differential for each trait.

The maize genotypes selected by our approach are different to the maize genotypes selected by the other approach (Table 4). However, six wheat genotypes (6, 10, 12, 39, 45, 71) selected with our approach were the same as those selected by the other approach (Table 5). This means that when the number of traits increases in the LPSI, both approaches tend to select different genotypes, which does not occur when the number of LPSI traits is low, as in the wheat dataset. Likewise, although the maize and wheat estimated LPSI values obtained using the profit function economic weights were all positive (Tables 4 and 5), the estimated maize and wheat LPSI values obtained using the other economic weights described above were all negative. Thus, both sets of economic weights affect the estimated LPSI values in a different way, as we would expect. In addition, note that the traits mean selected with the LPSI using our approach were all higher than the traits mean selected by the LPSI using the other approach. This explains why, in general, the selection differential values for each trait obtained with our approach were mainly positive, whereas for the other approach they were mostly negative.

5 | DISCUSSION

Using a stochastics linear regression model and a profit function, we developed a methodology to enable plant breeders define economic weights for selection indices. Our aim was to obtain economic weights for selection indices that can improve selection decisions in plant breeding when several traits (GY, maturity, HD, PH, etc.) are simultaneously selected. The problem to construct a profit function in maize and wheat breeding is evident: only GY has a market price, as we have indicated in the Introduction of this study. Therefore, our approach is based on GY market price and on the regression coefficients of GY on all the other associated traits.

Our results show that the profit function and the regression theory allow us to estimate the trait economic weights in the maize and wheat breeding context and select genotypes using the LPSI theory. For seven simulated datasets and two real datasets, the estimated LPSI selection responses were higher in all cases when we used the method described in this study to obtain the economic weights. This was not generally true for the estimated correlation between the LPSI and H, thus further research is necessary on this topic.

When we compared the $\hat{R}$ values obtained in this study with the $\hat{R}^*$ values (Table 3) obtained by Cerón-Rojas et al. (2015) and Cerón-Rojas and Crossa (2020b), who used the simulated datasets described in this paper and the LPSI to make selections, we found that in all cases the $\hat{R}$ values were higher than the $\hat{R}^*$ values. The same was true for the maize and wheat real datasets. In addition, the estimated selection responses of this research have an economic interpretation, but this type of interpretation it is not possible for the $\hat{R}^*$ values of the above authors. Thus, the profit function described in this study to obtain economic weights is effective for evaluating the profitability of plant breeding programs.

TABLE 4 Selected maize genotype and selected traits (grain yield, GY; anthesis day, AD; moisture content, MOI; plant height, PH; and ear height, EH) for k = 1.755 using the economic weights obtained with the profit function and the economic weights 5, −0.3, −0.3, −0.3 and −0.3, respectively, and estimated linear phenotypic selection index (LPSI) values

	Results using profit function							Results NOT using profit function					
Genotype	GY	AD	MOI	PH	EH	LPSI	Genotype	GY	AD	MOI	PH	EH	LPSI
	t ha^{-1}	d	%	——cm——				t ha^{-1}	d	%	——cm——		
59	8.7	72.1	16.3	225.3	118.7	17,188.9	67	4.3	64.3	14.6	205.6	91.6	−9.4
62	8.3	72.0	15.5	230.4	114.5	17,052.6	68	5.5	67.0	15.4	210.9	98.7	−10.2
64	8.1	71.3	15.3	239.1	118.5	16,913.1	56	5.2	65.6	15.7	208.0	100.9	−10.2
61	8.5	70.5	14.7	226.2	121.3	16,883.6	55	5.9	66.7	16.1	217.3	104.9	−10.2
63	6.4	72.4	14.0	214.7	104.1	16,859.1	66	7.7	71.3	16.0	212.0	95.4	−10.4
52	7.8	71.3	17.2	231.5	117.4	16,843.9	65	7.2	70.0	14.2	215.4	99.0	−10.5
9	7.4	71.1	16.8	229.4	121.9	16,813.3	53	7.5	70.8	17.7	232.8	113.2	−10.7
SIM[a]	7.9	71.5	15.7	228.1	116.6			6.3	67.4	14.8	210.6	99.0	
PM	7.4	70.2	16.4	226.0	115.8			7.4	70.2	16.4	226.0	115.8	
SD	0.47	1.3	−0.74	2.1	0.81			−1.1	−2.8	−1.6	−15.4	−16.8	

[a]SIM, selected individual mean; PM, population mean; SD, selection differential.

TABLE 5 Selected wheat genotypes and selected traits (grain yield, GY; heading, HD; and plant height, PH) for $k = 1.755$ using the economic weights obtained with the profit function and the economic weights 5, −0.3, and −0.3, respectively, and estimated linear phenotypic selection index (LPSI) values

Results using profit function					Results NOT using profit function				
Genotype	GY	HD	PH	LPSI	Genotype	GY	HD	PH	LPSI
	t ha^{-1}	d	cm			t ha^{-1}	d	cm	
45	9.1	74.9	76.7	15,713.6	45	9.1	74.9	76.8	−2.0
6	8.9	77.5	79.7	15,344.9	71	8.8	74.2	75.4	−2.7
31	8.8	75.9	82.2	15,140.5	39	8.7	70.1	80.3	−2.8
71	8.7	74.1	75.3	15,091.0	11	8.4	70.2	76.4	−3.0
9	8.7	77.3	81.8	15,024.3	10	8.7	76.5	75.4	−3.5
10	8.7	76.4	75.3	14,976.4	26	8.6	75.6	74.1	−3.5
39	8.7	70.1	80.2	14,926.8	12	8.6	73.8	78.7	−3.5
83	8.7	79.4	80.1	14,903.4	6	8.9	77.6	79.7	−3.8
12	8.6	73.7	78.6	14,880.8	37	8.2	70.7	75.8	−3.9
14	8.7	75.7	85.8	14,871.3	100	8.2	71.7	74.5	−3.9
Selected individual mean	8.7	75.5	79.6			8.6	73.5	76.7	
Population mean	8.1	75.8	79.3			8.1	75.8	79.3	
Selection differential	0.63	−0.28	0.29			0.50	−2.32	−2.63	

As we would expect, the profit function (Equations 7 and 8) assigned more weight to Y_1 than to the other traits. In addition, because the economic weight of Y_1 is equal to $\frac{\pi N_H}{2}$, all the seven selection cycles coefficients of Y_1 ($\hat{b}_1$) in Table 2 are equal to 1.0, whereas the values of the other traits' regression coefficients ($\hat{b}_2$ to $\hat{b}_4$) differ from 1.0. This shows the method described in this study is an adaptation of the Smith (1936) idea to the multiple regression context using the profit function and regression theory.

Note that Equations 7 and 8 are linked to the market situation and therefore the trait economic values are neither simply arbitrarily fixed values nor preference values. In addition, the regression coefficients, which are multiplied by $\frac{\pi N_H}{2}$ to obtain the economic weights, are associated with grain yield effects. Thus, the proposed profit function is a good option for obtaining economic weights to make LPSI selection in plant breeding.

5.1 | Why use a linear approach to derive the economic weights?

In the study of maize and wheat quantitative trait (QTs), it is assumed traits such as GY, PH, EH, etc., are the result of an undetermined number of unobservable gene effects distributed across the plant genome that interact among themselves and with the environment to produce the observable characteristic plant phenotypes (Cerón-Rojas & Crossa, 2018). This implies the QTs have continuously distributed phenotypes that do show a complex Mendelian inheritance (Hill, 2010). The QTs are difficult to analyze because heritable variations of these traits are masked by larger non-heritable variations that make it difficult to determine the genotypic values of individual plants (Smith, 1936).

To analyze QTs, we assumed the traits of interest and the net genetic merit have joint multivariate normal distribution. Under this distribution, the means, variances, and covariances completely describe the index and trait values. Moreover, if the trait values are not correlated, they are independent; linear combinations of traits are normal; and even when the trait phenotypic values do not have normal distribution, by the central limit this distribution serves as a useful approximation (Cerón-Rojas & Crossa, 2020b; Rencher, 2002). In addition, under the multivariate normality assumption, the regression of the net genetic merit on any linear function of the phenotypic values is linear (Cerón-Rojas & Crossa, 2022; Kempthorne & Nordskog, 1959).

Using histograms, quantile–quantile plots, and the Shapiro–Wilk and Kolmogorov–Smirnov normality tests, Cerón-Rojas and Crossa (2018, 2020b) showed that the estimated LPSI values, and the average values of traits such as GY, PH, EH, etc., in maize and wheat breeding, approached the normal distribution. One additional criterion to assume the QTs have multivariate normal distribution is based on the infinitesimal model theory (Barton et al., 2017; Fisher, 1918; Turelli, 2017; Walsh & Lynch, 2018). Under this model, (a) in the plant genome there is a very large number of loci, each with very small effects; (b) in a randomly mating population, under no selection, the genotypic distribution is normal, and (c) the genotypic distributions stay at least close to normal after selection (Walsh & Lynch, 2018).

Under the foregoing assumptions, Equations 3–5 are linear. Moreover, any function that is expressible as Equation 4 is linear even if the vector **b** depends on the joint distribution of H and the traits phenotypes (Cerón-Rojas & Crossa, 2022). Based on these reasons, using a linear approach to obtain economic weights in the maize and wheat breeding context seems correct. Finally, note in Equation 3, the independent and dependent variables are random variables, and the same is true for the residuals; then Equation 3 and 4 are stochastics linear model, no determinist model.

Alternatively, Goddard (1983) has analyzed the profit function in the linear and nonlinear context and concluded the better approach to maximize the LPSI selection response is to use a linear profit function. In his research, Goddard (1983) presents examples of why breeders should use a linear profit function in the LPSI context to derive economic weights. Based on that, we believe our approach to obtain economic weights is optimal in the context of maize and wheat.

5.2 | The invariance property of the MLE

The invariance property of the MLE is associated with the invariance principle of the likelihood ratio, which indicates "in the likelihood function the information should be *invariant* to the choice of parameterization" (Pawitan, 2013, p. 45). This means if we do not know where the parameter of interest is, then we should not know where its log is, or where its squared is, or its inverse value. That is, we should be equally ignorant regardless of how we model our problem. Pawitan (2013, p. 44) indicates "the invariant property of the likelihood ratio should be seen only as a convenient axiom, rather than a self-evident truth."

5.3 | Other approximations to the economic weights problem

Economic weights are difficult to assign in plant and animal breeding programs, and some authors have described alternatives to the LPSI. For example, Elston (1963) described a free-economic weights selection index that does not require estimates of the phenotypic and genotypic covariance matrices. Likewise, other authors (Brascamp, 1984; Itoh & Yamada, 1986, 1987; Pesek & Baker, 1969; Yamada et al., 1975) described the desired gains index, which does not use economic weights because it does not predict H, rather it only estimates the expectation of **g**. In turn, Cerón-Rojas et al. (2008b, 2016) described the eigen selection index method, where **w** (the vector of "economic weights") is a linear combination of the first eigen-vector of the matrix of multi-trait heritability. Thus, these last three indices do not use a profit function (net returns minus costs) to obtain economic weights to evaluate breeding programs. However, in this study, we have showed it is possible to assign economic weights to maize and wheat breeding traits using a profit function.

Finally, it is evident that there is the possibility to development non-linear profit functions to obtain economic weights, however, Goddard (1983) have showed that a LPSI will always give the highest selection response including when breeder uses non-linear profit function. Our research represents the first formal attempt to incorporate economic weights for selection indices that can improve selection decisions.

5.4 | Socio economic limitations to selection indices based only on profitability

Farmers and consumers are now placing strong emphasis on sourcing and sustainability, thus other criteria such as carbon footprints, water use, nitrogen emissions, and ecosystem services should be considered in addition to profitability. However, these issues are hard to capture in a simple selection index that attempts to improve few traits. Farms are more complex than simple profit and the need for cultivars that do not maximize profitability of the crop but of the system is an important factor that is not easily incorporated in a selection index.

In addition, not all farmers are risk takers, and often plant high yielding hybrids at low density to avoid failure risks. This gradient in plant population densities used by farmers indicates there is a gradient in risk attitudes. The question is: what is the target group to make selections for? Simple models to account for risk attitude suggest the selection of different crops and cultivars within crops account for the risk aversion differences. The method proposed here for estimating economic weights to be used in a selection index to bring an economic dimension to selection indices is a step towards allowing the breeder to use economic information correctly. Further studies are needed to formalize and bring socioeconomic dimensions to selection decisions by using a framework that considers the many uncertainties and sources of variability among farms and the mixture of farms in the target population of farms.

6 | CONCLUSIONS

Assuming the traits and the net genetic merit have joint multivariate normal distribution, we have described a profit function for obtaining economic weights in maize and wheat breeding programs. Using the profit function and the linear regression theory, we obtained a profit function that extends the Smith's idea to assign economic weights in wheat breeding. In animal breeding programs, all economic traits of interest have specific market prices; however, in the maize and wheat breeding context, only GY has a specific price on the market. For this reason, the proposed profit function is composed of two parts: one associated with the GY and the other linked to the expectation of GY given the values of the other traits. Using the proposed approach, the average of the estimated correlation between the LPSI and the net genetic merit for the seven simulated selection cycles was 0.820, and for the maize and wheat real datasets were 0.87 and 0.85, respectively. Therefore, we concluded the profit function proposed in this study is a strong option for obtaining economic weights in plant breeding.

ACKNOWLEDGMENTS

We thank all CIMMYT scientists, field workers, and lab assistants who collected the real data used in this study. We very much appreciate the fruitful exchange of ideas with Prof. Dr. Agustin Blasco, a source of scientific inspiration. We appreciate the efficient assistance from the associate and technical editors, and the valuable opinions of two anonymous reviewers that enrich the discussion on the value and limitations of selection indices based solely on profitability.

Open Access fees were received from the Bill & Melinda Gates Foundation. We acknowledge the financial support provided by the Bill & Melinda Gates Foundation (INV-003439 BMGF/FCDO Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods, AGG) as well as USAID projects (Amend. No. 9 MTO 069033, USAID-CIMMYT Wheat/AGGMW, AGG-Maize Supplementary Project, AGG, Stress Tolerant Maize for Africa) generated the CIMMYT data analyzed in this study. We are also thankful for the financial support provided by the Foundation for Research Levy on Agricultural Products (FFL) and the Agricultural Agreement Research Fund (JA) through the Research Council of Norway for Grants 301835 (Sustainable Management of Rust Diseases in Wheat) and 320090 (Phenotyping for Healthier and more Productive Wheat Crops)

AUTHOR CONTRIBUTIONS

J. Jesus Cerón-Rojas: Conceptualization; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing. Manje Gowda: Investigation; Validation; Writing – review & editing. Yoseph Beyene: Conceptualization; Methodology; Writing – review & editing. Alison Bentley: Conceptualization; Writing – review & editing. Leo Crespo-Herrera: Conceptualization; Writing – review & editing. Fernando Toledo: Investigation; Methodology; Validation; Writing – review & editing. Keith Gardner: Investigation; Methodology; Writing – review & editing. Jose Crossa: Conceptualization; Investigation; Methodology; Validation; Writing – original draft; Writing – review & editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Alison R. Bentley https://orcid.org/0000-0001-5519-4357
Leo Crespo-Herrera https://orcid.org/0000-0003-0506-4700
Jose Crossa https://orcid.org/0000-0001-9429-5855

REFERENCES

Alvarado, G., Pacheco, A., Perez-Elizalde, S., Burgueño, J., & Rodriguez, F. M. (2018). RinSel: Selection indices with R. In *Linear selection indices in modern plant breeding* (pp. 243–256). Springer.

Barton, N. H., Etheridge, A. M., & Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, *118*, 50–73. https://doi.org/10.1016/j.tpb.2017.06.001

Blasco, A. (2021). *Mejora genética animal*. Síntesis.

Brascamp, E. W. (1984). Selection indices with constraints. *Animal Breeding Abstract*, *52*(9), 645–654.

Cerón-Rojas, J. J., Castillo-González, F., Sahagún-Castellanos, J., Santacruz-Varela, A., Benítez-Riquelme, I., & Crossa, J. (2008). A molecular selection index method based on eigen analysis. *Genetics*, *180*, 547–557. https://doi.org/10.1534/genetics.108.087387

Cerón -Rojas, J. J., Crossa, J., Arief, V. N., Basford, K., Rutkoski, J., Jarquín, D., Alvarado, G., Beyene, Y., Semagn, K., & DeLacy, I. (2015). A genomic selection index applied to simulated and real data. *G3 Genes|Genomes|Genetics*, *5*, 2155–2164. https://doi.org/10.1534/g3.115.019869

Cerón-Rojas, J. J., & Crossa, J. (2018). *Linear selection indices in modern plant breeding*. Springer.

Cerón-Rojas, J. J., & Crossa, J. (2019). Efficiency of a constrained linear genomic selection index to predict the net genetic merit in plants. *G3 Genes|Genomes|Genetics*, *9*, 3981–3994. https://doi.org/10.1534/g3.119.400677

Cerón-Rojas, J. J., & Crossa, J. (2020a). Combined multi-stage linear genomic selection indices to predict the net genetic merit in plant breeding. *G3 Genes|Genomes|Genetics*, *10*, 2087–2101. https://doi.org/10.1534/g3.120.401171

Cerón-Rojas, J. J., & Crossa, J. (2020b). Expectation and variance of the estimator of the maximized selection response of linear selection indices with normal distribution. *Theoretical and Applied Genetics*, *133*(9), 2743–2758. https://doi.org/10.1007/s00122-020-03629-6

Cerón-Rojas, J. J., & Crossa, J. (2022). The statistical theory of linear selection indices from phenotypic to genomic selection. *Crop Science*, *62*, 537–563. https://doi.org/10.1002/csc2.20676

Charffeddine, N., & Alenda, R. (1998). Conceptos económicos en un programa de mejora genética en vacuno de leche. *ITEA*, *94A*(3), 179–205.

Christensen, R. (2011). *Plane answers to complex questions: The theory of linear models* (4th ed.). Springer.

Cochran, W. G. (1951). Improvement by means of selection. In J. Neyman (Ed.), *Proceedings of the second Berkeley symposium on mathematical statistics and probability* (pp. 449–470). University of California Press.

Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics*, *124*, 331–341. https://doi.org/10.1111/j.1439-0388.2007.00701.x

Elston, R. C. (1963). A weight-free index for the purpose of ranking or selection with respect to several traits at a time. *Biometrics*, *19*, 85–97. https://doi.org/10.2307/2527573

Fisher, R. A. (1918). The correlation between relatives under the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, *52*, 399–433. https://doi.org/10.1017/S0080456800012163

Goddard, M. E. (1983). Selection indices for non-linear profit functions. *Theoretical and Applied Genetics*, *64*, 339–344. https://doi.org/10.1007/BF00274177

Hazel, L. N., Dickerson, G. E., & Freeman, A. E. (1994). The selection index: Then, now, and for the future. *Journal of Dairy Science*, *77*, 3236–3251. https://doi.org/10.3168/jds.S0022-0302(94)77265-9

Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 73–85. https://doi.org/10.1098/rstb.2009.0203

Itoh, Y., & Yamanda, Y. (1986). Re-examination of selection index for desired gains. *Genetics, Selection, Evolution*, *18*(4), 499–504. https://doi.org/10.1186/1297-9686-18-4-499

Itoh, Y., & Yamada, Y. (1987). Comparisons of selection indices achieving predetermined proportional gains. *Genetics, Selection, Evolution*, *19*(1), 69–82. https://doi.org/10.1186/1297-9686-19-1-69

Itoh, Y., & Yamada, Y. (1988). Linear selection indices for non-linear profit functions. *Theoretical and Applied Genetics*, *75*, 553–560. https://doi.org/10.1007/BF00289120

Kempthorne, O., & Nordskog, A. W. (1959). Restricted selection indices. *Biometrics*, *15*, 10–19. https://doi.org/10.2307/2527598

Kollo, T. (2005). *Advanced multivariate statistics with matrices*. Springer.

Lande, R., & Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, *124*, 743–756. https://doi.org/10.1093/genetics/124.3.743

Magnussen, S. (1990). Selection index: economic weights for maximum simultaneous genetic gains. *Theoretical and Applied Genetics*, *79*, 289–293. https://doi.org/10.1007/BF01186069

Moav, R., & Moav, J. (1966). Profit in a broiler enterprise as a function of egg production of parent stocks and growth rate of their progeny. *British Poultry Science*, *7*, 5–15. https://doi.org/10.1080/00071666608415600

Moav, R., & Hill, W. G. (1966). Specialized sire and dam lines. IV. Selection within lines. *Animal Production*, *8*, 373–390.

Pacheco, A., Pérez, S., Alvarado, G., Cerón, J., Rodríguez, F., Crossa, J., & Burgueño, J. (2017). *RIndSel: Selection indices for plant breeding*. CIMMYT Research Data & Software Repository Network.

Pawitan, Y. (2013). *In all likelihood: Statistical modelling and inference using likelihood* (2nd ed.). Oxford University Press.

Perez-Elizalde, S., Cerón-Rojas, J. J., Crossa, J., Fleury, D., & Alvarado, G. (2014). Rindsel: An R package for phenotypic and molecular selection indices used in plant breeding. In D. Fleury & R. Whitford (Eds.), Crop breeding: Methods in molecular biology *(*methods and protocols*)* (Vol. 1145, pp. 87–96). Humana Press.

Pesek, J., & Baker, R. J. (1969). Desired improvement in relation to selection indices. *Canadian Journal of Plant Science*, *49*, 803–804. https://doi.org/10.4141/cjps69-137

Rencher, A. C. (2002). *Methods of multivariate analysis* (2nd ed.). Wiley.

Rencher, A. C., & Schaalje, G. B. (2008). *Linear models in statistics* (2nd ed.). John Wiley and Sons.

Searle, S. R., Casella, G., & McCulloch, C. E. (2006). *Variance components*. John Wiley and Sons.

Smith, H. F. (1936). A discriminant function for plant selection. *Annals of Eugenics*, *7*, 240–250. https://doi.org/10.1111/j.1469-1809.1936.tb02143.x

Strain, J. H., & Nordskog, A. W. (1962). Genetic aspects of the profit equation in a broiler enterprise. *Poultry Science*, *41*, 1892–1902. https://doi.org/10.3382/ps.0411892

Teunissen, P. J. G. (2007). Best prediction in linear models with mixed integer/real unknowns: theory and application. *Journal of Geodesy*, *81*, 759–780. https://doi.org/10.1007/s00190-007-0140-6

Turelli, M. (2017). Fisher's infinitesimal model: a story for the ages. *Theoretical Population Biology*, *118*, 46–49. https://doi.org/10.1016/j.tpb.2017.09.003

Walsh, B., & Lynch, M. (2018). *Evolution and Selection of Quantitative Traits*. Oxford University Press.

Yamada, Y., Yokouchi, K., & Nishida, A. (1975). Selection index when genetic gains of individual traits are of primary concern. *The Japanese Journal of Genetics*, *50*(1), 33–41. https://doi.org/10.1266/jjg.50.33

How to cite this article: Cerón-Rojas, J. J., Gowda, M., Toledo, F., Beyene, Y., Bentley, A. R, Crespo-Herrera, L., Gardner, K., & Crossa, J. (2023). A linear profit function for economic weights of linear phenotypic selection indices in plant breeding. *Crop Science*, 1–13. https://doi.org/10.1002/csc2.20882

APPENDIX

Linear Phenotypic Selection Index Theory

The net genetic merit

Let $\mathbf{g}' = [G_1 \; G_2 \; \ldots \; G_t]$ be a vector of true unobservable genotypic random variables for t traits with multivariate normal distribution and null expectation; then the individual net genetic merit is:

$$H = \boldsymbol{w}'\boldsymbol{g} \tag{A1}$$

where $\mathbf{w}' = [w_1 \; w_2 \; \ldots \; w_t]$ is a vector of economic weights. The variance of H is denoted as:

$\sigma_H^2 = \mathbf{w}'\mathbf{G}\mathbf{w}$,

where $\mathbf{G}$ is the genotypic covariance matrix.

It is possible to write Equation A1 as:

$$H = m_H + \boldsymbol{\beta}'(\mathbf{y} - \boldsymbol{\mu}) + \epsilon \tag{A2}$$

where m_H is the mean of H, $\mathbf{y}' = [Y_1 \; Y_2 \; \ldots \; Y_t]$ is a random vector of t traits, $\boldsymbol{\mu}' = [\mu_1 \; \mu_2 \; \ldots \; \mu_t]$ is a vector of t phenotypic means of $\mathbf{y}$, and ϵ is the deviation of $m_H + \boldsymbol{\beta}'(\mathbf{y} - \boldsymbol{\mu})$ from H. We have assumed that ϵ has a normal distribution, null expectation, and variance σ_ϵ^2, and that the covariance between any pairs of ϵ is 0. Equation A2 is a multiple linear regression model where the independent ($\mathbf{y}$) and dependent (H) variables are random and have a joint multivariable normal distribution.

In the profit function context, Equations A1 and A2 denote the total individual genotypic economic value (Kempthorne & Nordskog, 1959) and are expressed in monetary units. For example, if we select for grain yield and grain protein in each selection cycle, H should be written as:

$$H = \frac{\$}{\frac{\text{grain}}{\text{ha}}}\,\text{grain/ha} + \frac{\$}{\frac{\text{protein}}{\text{ha}}}\,\text{protein/ha} = \$$$

where "\$" denotes the costs of grain yield and grain protein (Blasco, 2021).

The LPSI is the BLP of H

Suppose that m_H and $\boldsymbol{\mu}$ are equal to zero, and the covariance matrix $\mathbf{P}$ is finite; then, the LPSI (or the BLP of H) is:

$$\mathrm{E}(\mathrm{H}|\mathbf{y}) = \mathbf{w}'\mathbf{G}\mathbf{P}^{-1}\mathbf{y} = \boldsymbol{\beta}'\mathbf{y} \tag{A3}$$

where $\mathbf{w}'\mathbf{G}$ is the covariance between H and $\mathbf{y}$, $\mathbf{P}^{-1}$ is the inverse of the matrix $\mathbf{P}$ and $\boldsymbol{\beta}' = \mathbf{w}'\mathbf{G}\mathbf{P}^{-1}$. The assumption of multivariate normal distribution of H and $\mathbf{y}$ is a sufficient condition for Equation A3 to be linear.

Selection response and correlation

Because $\boldsymbol{\beta}' = \mathbf{w}'\mathbf{G}\mathbf{P}^{-1}$, the LPSI selection response is:

$$R = k\sqrt{\boldsymbol{\beta}' P \beta}, \tag{A4}$$

where k is the selection intensity and $\sqrt{\boldsymbol{\beta}' P \beta}$ is the standard deviation of I. Equation A4 predicts the mean improvement in H attributable to indirect selection on I only when $\boldsymbol{\beta}' = \mathbf{w}'\mathbf{G}\mathbf{P}^{-1}$. In similar manner, the correlation between H and I can be written as:

$$\rho_{HI} = \frac{\sqrt{\boldsymbol{\beta}' P \beta}}{\sqrt{\mathbf{w}'\mathbf{G}\mathbf{w}}}. \tag{A5}$$

All the other terms were defined earlier. Equation A5 is the proportion of the variance of H that can be attributed to the regression relationship with I.