Published November 6, 2015

RESEARCH

# Statistical Sampling Properties of the Coefficients of Three Phenotypic Selection Indices

J. Jesus Cerón-Rojas, José Crossa,* and Jaime Sahagún-Castellanos


## ABSTRACT

The aim of the Smith phenotypic selection index (SPSI), the restricted phenotypic selection index (RPSI), and the predetermined proportional gains phenotypic selection index (PPG-PSI) is to maximize the response to selection and provide the breeder with an objective rule for evaluating and selecting several traits. When the phenotypic and genotypic variances and covariances are known, these three indices are the best linear predictors. When these parameters are estimated, the three indices will be optimal only if the estimators of the index weights are unbiased and have minimal variance. There are many methods for determining the sampling properties of the SPSI but there is no method for determining the sampling properties of RPSI and PPG-PSI coefficients. Using the canonical correlation theory, we proposed an asymptotic method for determining the statistical sampling properties of the estimators of the coefficients of the three phenotypic selection indices. We showed that under some assumptions, the sampling properties of the RPSI and PPG-PSI coefficient estimators could be obtained using the sampling properties of the SPSI coefficient estimator. We validated the theoretical results using two real datasets. The theoretical and numerical results indicated that the three estimators of the weights for the three indices were unbiased with minimal variances. We concluded that when the number of genotypes is large, the proposed method could be used to find the sampling properties of the coefficient of the three indices.


J.J. Cerón-Rojas, and J. Sahagún-Castellanos, Instituto de Horticultura, Departamento de Fitotecnia, Universidad Autónoma Chapingo, Chapingo, México, C.P. 56230; J. Crossa, Biometrics and Statistics Unit (BSU), International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600, México D.F., México. Received 29 Mar. 2015. Accepted 23 July 2015. *Corresponding author (j.crossa@cgiar.org).

**Abbreviations:** EHT, ear height; GY, grain yield; MAS, marker-assisted selection; MESIM, molecular eigen selection index method; MOI, grain moisture content; PHT, plant height; PPG-PSI, predetermined proportional gains phenotypic selection index; PSI, phenotypic selection indices; RESIM, restricted eigen selection index; RPSI, restricted phenotypic selection index; SPSI, Smith phenotypic selection index.

In animal and plant breeding, phenotypic selection indices (PSI) are used for combining selection of several traits; they provide animal and plant breeders with objective rules for maximizing overall genetic gains. The aim of PSI is to maximize the selection response and provide the breeder with an objective rule for simultaneously evaluating and selecting several traits (Baker, 1974). One of the most efficient methods for predicting the net genetic merit of plants and animals is the standard Smith (1936) phenotypic selection index (SPSI) under the assumption that the net genetic merit of the candidates for selection is a linear combination of the additive genetic values of several traits. When the index parameters are known, the SPSI (i) is the best linear predictor of the net genetic merits of the candidates for selection, and (ii) has maximum correlation with the true net genetic merit (Bulmer, 1980).

One of the main problems of the SPSI is that, when used to select individuals, the mean of the traits can change in a positive or negative direction without control. This was the main reason why Kempthorne and Nordskog (1959) developed the basic theory of the RPSI that allows imposing restrictions equal to

Published in Crop Sci. 56:51–58 (2016).
doi: 10.2135/cropsci2015.03.0189
Freely available online through the author-supported open-access option.

zero on the expected genetic advance of some traits while the expected genetic advance of other traits increases (or decreases) without imposing any restrictions.

Based on the ideas of the Kempthorne and Nordskog (1959) RPSI, Tallis (1962) proposed a selection index called PPG-PSI that attempts to make some traits change their values based on a predetermined level while the rest of the traits remain without restrictions. Mallard (1972) pointed out that the predetermined proportional gains PSI of Tallis (1962) does not provide optimum genetic advances and was the first to propose an optimum PPG-PSI based on a slight modification of the RPSI.

Another PPG-PSI was proposed by Harville (1975); it maximized the correlation between PSI and the net genetic merit subject to the restriction that the covariance between the index and some linear functions of the genotypes is different from zero. Tallis (1985) modified his original selection index (Tallis, 1962) and minimized the variance of the difference between the index and the net genetic merit using restrictions similar to those of Harville (1975). Itoh and Yamada (1987) mentioned that the PPG-PSI proposed by Harville (1975) and that of Tallis (1985) are equivalent and indicated some problems associated with the proportionality constants used in the Harville (1975) and Tallis (1985) indices. Later, Lin (2005) demonstrated that the PPG-PSI of Tallis (1985) could be extended to cases where more than one predetermined proportional gain is imposed on the genetic gain per selection cycle and that such an index can be constructed in one stage. In practice, the Mallard (1972) and Tallis (1985) PPG-PSI produce the same results when two or more traits are restricted.

The SPSI, RPSI, and Mallard (1972) PPG-PSI are now the standard selection indices used in plant and animal breeding programs for choosing candidates for selection with or without predetermined restrictions on the PSI traits. When the phenotypic and genotypic variance and covariances of these three indices are known, then (i) the correlation between the net genetic merits of the candidates for selection and the SPSI, RPSI, and PPG-PSI is maximized, (ii) the mean prediction error is minimized, and (iii) the SPSI, RPSI, and PPG-PSI are the best linear predictors of the net genetic merit of the candidates for selection and the ones with the highest relative efficiency compared with other selection procedures and are easy to use.

When the phenotypic and genotypic variance and covariances are estimated, it is not known if the sampling properties of the SPSI, RPSI, and PPG-PSI coefficients are indeed optimal. Tallis (1960) derived a large sample variance of index weights for individually selecting any number of traits and the predicted response when phenotypic and genetic parameters are estimated in a half-sib analysis; however, the expressions are complicated and do not allow identifying situations where selection indices are likely to be inefficient. Williams (1962) obtained an exact formula for the sampling variance of the index weights for only two variables of a specific experimental design. Harris (1964) used the delta method to determine the sampling properties of the SPSI; however, the results are confusing and the author did not present a general formula for the sampling statistical properties of the SPSI coefficients. Hayes and Hill (1980) proposed a transformation of the trait variables used for constructing genetic selection indices such that the sampling properties of the SPSI weights can be easily computed using a general formula; however, the formula depends on the transformation of the trait variables.

The selection response of SPSI, RPSI, and PPG-PSI will be optimal when the estimator index weights are unbiased and have minimal variances, but its efficiency is likely to decrease if the estimator index weights are biased with no minimal variances. The sampling variance of the index weights will therefore provide some idea of the likely loss of efficiency; if the variances are high, the index is probably far from optimal (Hayes and Hill, 1980).

The aim of this study was to demonstrate that, in the asymptotic context, it is possible to generate a general formula for determining the sampling properties of the estimators of the SPSI, RPSI, and PPG-PSI coefficients using the canonical correlation theory (Anderson, 1999, 2003). The formula proposed does not depend on any transformation of trait variables or on a specific experimental design.

# MATERIALS AND METHODS

## Theory of Phenotypic Selection Indices

### The Vector of Coefficients of the SPSI

Let $\mathbf{g}' = [\ g_1 \quad g_2 \quad \ldots \quad g_t\ ]$ and $\mathbf{w}' = [\ w_1 \quad w_2 \quad \ldots \quad w_t\ ]$ be vectors of true breeding values and economic weights, respectively, of $t$ traits under selection. The objective of the Smith (1936) PSI is to predict and select the net genetic merit $H = \mathbf{w}'\mathbf{g}$ using the PSI: $I = \mathbf{b}'\mathbf{p}'$, where $\mathbf{p}' = [\ p_1 \quad p_2 \quad \cdots \quad p_t\ ]$ is a vector of trait phenotypic values and $\mathbf{b}' = [\ b_1 \quad b_2 \quad \ldots \quad b_t\ ]$ is a vector of coefficients of the PSI. Let $\mathbf{P}^{-1}$ be the inverse of the phenotypic covariance matrix ($\mathbf{P}$) and $\mathbf{G}$, the matrix covariance of the vector of true breeding values $\mathbf{g}$; then the vector

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{w} \qquad [1]$$

maximizes the PSI expected genetic advance per selection cycle for each trait, that is,

$$\mathbf{E}_{\mathrm{PSI}} = k\frac{\mathbf{G}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}} \qquad [2],$$

where $k$ is the standardized selection differential or selection intensity.

### The Vector of Coefficients of the Restricted Phenotypic Selection Index

Suppose that the breeder is interested in improving only $r$ of $t$ ($r < t$) traits, leaving ($t - r$) unchanged. Kempthorne and Nordskog

(1959) solved this problem by assuming that $\mathbf{b'Pb} = 1$ and imposing restrictions on the SPSI expected genetic advance per selection cycle for each trait (Eq. [2]). Kempthorne and Nordskog (1959) maximized the function: $\Psi = \mathbf{w'Gb} - 0.5\lambda(\mathbf{b'Pb} - 1) - \mathbf{v'C'b}$ with respect to $\mathbf{b}$, where $\mathbf{C'} = \mathbf{U'G}$, $\mathbf{C'b} = 0$, and $\mathbf{U'}$ is a matrix of ones and zeros (1 indicates that the trait is restricted and 0 that the trait has no restriction); $0.5\lambda$ and $\mathbf{v'} = [\ v_1 \quad v_2 \quad \cdots \quad v_r\ ]$ are Lagrange multipliers. The rest of the parameters were defined in Eq. [1]. The vector that maximizes $\Psi$ and $\mathbf{E}_{PSI}$ under the given restrictions is

$$\mathbf{b}_{KN} = \mathbf{Kb} \qquad [3],$$

where $\mathbf{K} = [\mathbf{I} - \mathbf{Q}]$, $\mathbf{Q} = \mathbf{P}^{-1}\mathbf{C}(\mathbf{C'P}^{-1}\mathbf{C})^{-1}\mathbf{C'}$, and $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$ is the vector of coefficients of the SPSI; $\mathbf{I}$ is an identity matrix of order $t \times t$; $\mathbf{P}^{-1}$ and $\mathbf{G}$ were defined in Eq. [1]. When $\mathbf{U'}$ is a null matrix (no restrictions on any traits), $\mathbf{b}_{KN} = \mathbf{b}$. The RPSI expected genetic advance per selection cycle for each trait could be written as

$$\mathbf{E}_{KN} = k\frac{\mathbf{Gb}_{KN}}{\sqrt{\mathbf{b}'_{KN}\mathbf{Pb}_{KN}}} \qquad [4],$$

where $k$ was defined in Eq. [2].

### The Vector of Coefficients of the Predetermined Proportional Gains Phenotypic Selection Index

Mallard (1972) extended the idea of Kempthorne and Nordskog (1959) by considering that if $\mu_q$ is the population mean of the $q$th trait before selection, one objective could be to change $\mu_q$ to $\mu_q + d_q$, where $d_q$ is the predetermined change in $\mu_q$ in one selection cycle (in Kempthorne and Nordskog [1959], $d_q = 0$, $q = 1,2,\cdots,r$); the rest of the traits change with no restrictions.

Let $\mathbf{D'} = \begin{bmatrix} d_r & 0 & \cdots & 0 & -d_1 \\ 0 & d_r & \cdots & 0 & -d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_r & -d_{r-1} \end{bmatrix}$ be a matrix of predetermined changes, where $r$ is the number of predetermined proportional gains and $d_q$ ($q = 1,2,\ldots,r$) is the $q$th element of the vector of predetermined restrictions $\mathbf{d'} = [\ d_1 \quad d_2 \quad \cdots \quad d_r\ ]$, imposed by the researcher. Let $\mathbf{M'} = \mathbf{D'C'}$ be a new matrix of restrictions, where $\mathbf{C'} = \mathbf{U'G}$; then it is possible to impose the desired predetermined proportional gain restrictions on Eq. [2] as $\mathbf{M'b} = 0$ and maximize $\Psi_M = \mathbf{w'Gb} - 0.5\lambda(\mathbf{b'Pb} - 1) - \mathbf{v'M'b}$ with respect to $\mathbf{b}$. The solution is the vector

$$\mathbf{b}_M = \mathbf{K}_M\mathbf{b} \qquad [5],$$

where $\mathbf{K}_M = [\mathbf{I} - \mathbf{Q}_M]$, $\mathbf{Q}_M = \mathbf{P}^{-1}\mathbf{M}(\mathbf{M'P}^{-1}\mathbf{M})^{-1}\mathbf{M'}$, and $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$; $\mathbf{I}$, $0.5\lambda$, and $\mathbf{v'} = [\ v_1 \quad v_2 \quad \cdots \quad v_r\ ]$ were defined in Eq. [3], respectively. When $\mathbf{D} = \mathbf{I}$, $\mathbf{b}_M = \mathbf{b}_{KN}$ (the vector of RPSI), and when $\mathbf{D} = \mathbf{I}$ and $\mathbf{U'}$ is a null matrix, $\mathbf{b}_M = \mathbf{b}$ (the vector of the SPSI). That is, the Mallard (1972) index is more general than the Kempthorne and Nordskog (1959) RPSI and is an optimum PPG-PSI. The PPG-PSI expected genetic advance per selection cycle for each trait could be written as

$$\mathbf{E}_M = k\frac{\mathbf{Gb}_M}{\sqrt{\mathbf{b}'_M\mathbf{Pb}_M}} \qquad [6],$$

where $k$ was defined in Eq. [2].

### Canonical Correlations Between p and g

Let $\mathbf{p}$ and $\mathbf{g}$ be vectors of trait phenotypic values and true breeding values, respectively, of $t$ traits under selection, as defined in Eq. [1]. In addition, suppose that $\mathbf{p}$ and $\mathbf{g}$ have a joint normal distribution. We can define a new vector as $\mathbf{x'} = [\ \mathbf{p'} \quad \mathbf{g'}\ ]$; in this case, the covariance matrix of $\mathbf{x}$ will be $\begin{bmatrix} \mathbf{P} & \mathbf{G} \\ \mathbf{G} & \mathbf{G} \end{bmatrix}$. Matrices $\mathbf{P}$ and $\mathbf{G}$ were defined in Eq. [1]. From the covariance matrix of $\mathbf{x}$, we can find one measure of the association between $\mathbf{p}$ and $\mathbf{g}$ using the canonical correlation theory (Anderson, 1999, 2003; Cerón-Rojas et al., 2008a). In this case, the solution is

$$(\mathbf{P}^{-1}\mathbf{G} - \lambda_j^2\mathbf{I})\boldsymbol{\beta}_j = \mathbf{0} \qquad [7],$$

where $\boldsymbol{\beta}_j$, the $j$th eigenvector, and $\lambda_j$, the $j$th ($j = 1,2,\cdots,t$) canonical correlation of $\mathbf{p}$ and $\mathbf{g}$ can be used to find the sampling statistical properties of Eq. [1], [3], and [5].

In addition, $\boldsymbol{\beta}_j$ and $\lambda_j$ can be estimated from

$$(\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}} - \hat{\lambda}_j^2\mathbf{I})\hat{\boldsymbol{\beta}}_j = \mathbf{0} \qquad [8],$$

where $\hat{\mathbf{P}}$, $\hat{\mathbf{G}}$, $\hat{\boldsymbol{\beta}}_j$, and $\hat{\lambda}_j^2$ are maximum likelihood estimates (Anderson, 2003; Cerón-Rojas et al., 2008b) of $\mathbf{P}$, $\mathbf{G}$, $\boldsymbol{\beta}_j$, and $\lambda_j^2$, respectively. Note that $\hat{\lambda}_j^2$ is positive only if $\hat{\mathbf{P}}$ is positive definite (all eigenvalues positive) and $\hat{\mathbf{G}}$ is positive semidefinite (no negative eigenvalues). Since $\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}$ is an asymmetric matrix, the values of $\hat{\boldsymbol{\beta}}_j$ and $\hat{\lambda}_j^2$ can be obtained using singular value decomposition (Cerón-Rojas et al., 2008b).

## Datasets

### Dataset 1

These data are from commercial egg poultry lines and were obtained from Akbar et al. (1984). The estimated phenotypic ($\hat{\mathbf{P}}$) and genetic ($\hat{\mathbf{G}}$) covariance matrices among rate of lay (number of eggs), age at sexual maturity (d), egg weight (kg), and body weight (kg) were:

$$\hat{\mathbf{P}} = \begin{bmatrix} 240.57 & -95.62 & 2.07 & 54.40 \\ -95.62 & 167.20 & 4.58 & 15.36 \\ 2.07 & 4.58 & 22.80 & 37.20 \\ 54.4 & 15.36 & 37.20 & 516.11 \end{bmatrix} \text{ and}$$

$$\hat{\mathbf{G}} = \begin{bmatrix} 29.86 & -17.9 & -4.13 & -1.75 \\ -17.9 & 18.56 & 1.49 & -4.88 \\ -4.13 & 1.49 & 9.24 & 16.66 \\ -1.75 & -4.88 & 16.66 & 179.73 \end{bmatrix}.$$

The number of genotypes was $n$ = 3330, and the vector of economic values was $\mathbf{w}' = [\ 19.54 \quad -3.56 \quad 17.01 \quad -2.51\ ]$. We restricted three traits in RPSI and imposed three predetermined proportional gains in PPG-PSI, that is, $\mathbf{d}' = [\ 3 \quad -1 \quad 2\ ]$. Then matrices $\mathbf{U}'$ and $\mathbf{D}'$ for this dataset were

$$\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{D}' = \begin{bmatrix} 2 & 0 & -3 \\ 0 & 2 & 1 \end{bmatrix}.$$

The $\mathbf{d}'$ values were taken from Lin (2005), who used data from Akbar et al. (1984) to illustrate results in his paper. In addition, matrices $\mathbf{C} = \hat{\mathbf{G}}\mathbf{U}$ and $\mathbf{M} = \mathbf{CD}$ were equal to

$$\mathbf{C} = \begin{bmatrix} 26.86 & -17.90 & -4.13 \\ -17.90 & 18.56 & 1.49 \\ -4.13 & 1.49 & 9.24 \\ -1.75 & -4.88 & 16.66 \end{bmatrix} \text{ and}$$

$$\mathbf{M} = \begin{bmatrix} 72.11 & -39.93 \\ -40.27 & 38.61 \\ 35.98 & 12.22 \\ -53.48 & 6.90 \end{bmatrix}, \text{ respectively.}$$

### Dataset 2

This is a CIMMYT maize (*Zea mays* L.) $F_2$ population comprising four traits: grain yield (GY, t ha$^{-1}$), plant height (PHT, cm), ear height (EHT, cm), and grain moisture content (MOI). The estimated phenotypic ($\hat{\mathbf{P}}$) and genetic ($\hat{\mathbf{G}}$) covariance matrices among traits GY, PHT, EHT and MOI were:

$$\hat{\mathbf{P}} = \begin{bmatrix} 1.29 & 3.98 & 2.16 & 0.34 \\ 3.98 & 198.87 & 136.56 & 1.13 \\ 2.16 & 136.56 & 184.02 & 1.74 \\ 0.34 & 1.13 & 1.74 & 1.02 \end{bmatrix} \text{ and}$$

$$\hat{\mathbf{G}} = \begin{bmatrix} 0.40 & 2.16 & 1.18 & 0.22 \\ 2.16 & 66.17 & 57.45 & 1.91 \\ 1.18 & 57.45 & 62.36 & 2.1 \\ 0.22 & 1.91 & 2.1 & 0.5 \end{bmatrix},$$

where the number of genotypes or individuals was $n$ = 250 and the vector of economic values was $\mathbf{w}' = [\ 1 \quad -1 \quad -1 \quad -1\ ]$. We restricted three traits in RPSI and imposed three predetermined proportional gains in the PPG-PSI, that is, $\mathbf{d}' = [\ 2 \quad -1 \quad 10\ ]$. Matrices $\mathbf{U}'$ and $\mathbf{D}'$ for this dataset were

$$\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{D}' = \begin{bmatrix} 10 & 0 & -2 \\ 0 & 10 & 1 \end{bmatrix}.$$

Vector $\mathbf{d}'$ was obtained from a paper by Itoh and Yamada (1987). Then, matrices $\mathbf{C} = \hat{\mathbf{G}}\mathbf{U}$ and $\mathbf{M} = \mathbf{CD}$ were equal to

$$\mathbf{C} = \begin{bmatrix} 0.40 & 2.16 & 1.18 \\ 2.16 & 66.17 & 57.45 \\ 1.18 & 57.45 & 62.36 \\ 0.22 & 1.91 & 2.10 \end{bmatrix} \text{ and}$$

$$\mathbf{M} = \begin{bmatrix} 1.64 & 22.78 \\ -93.30 & 719.15 \\ -112.92 & 636.86 \\ -2.00 & 21.20 \end{bmatrix}, \text{ respectively.}$$

## RESULTS AND DISCUSSION

### Expectation and Variance of Vector b

Let $\mathbf{B} = \{\boldsymbol{\beta}_j\}$ ($j = 1,2,\cdots,t$, $t$ = number of traits) be the matrix of the eigenvectors of matrix $\mathbf{T} = \mathbf{P}^{-1}\mathbf{G}$ (Eq. [7]), then $\mathbf{T} = \mathbf{BLB}'$, where $\mathbf{L} = \text{Diag}\{\lambda_j^2\}$ is a diagonal matrix with $\mathbf{T}$ eigenvalues. Suppose that $\mathbf{b}$ is in the space generated by the $\mathbf{B}$ eigenvectors, then $\mathbf{b}$ can be written as

$$\mathbf{b} = \mathbf{B}\boldsymbol{\alpha} = \sum_{j=1}^{t} \alpha_j \boldsymbol{\beta}_j \qquad [9],$$

where $\boldsymbol{\alpha}' = [\ \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_t\ ]$ is a vector of unknown constants (Rao, 2002; Cerón-Rojas et al., 2008b; Crossa and Cerón-Rojas, 2011). By Eq. [9], the expectation and variance of $\mathbf{b}$ can be denoted as

$$E(\mathbf{b}) = \sum_{j=1}^{t} \alpha_j E(\boldsymbol{\beta}_j) \qquad [10] \text{ and}$$

$$\text{Var}(\mathbf{b}) = \sum_{j=1}^{t} \alpha_j^2 \text{Var}(\boldsymbol{\beta}_j) + 2\sum_{i=1}^{t-1}\sum_{q=i+1}^{t} \alpha_i \alpha_q \text{Cov}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_q) \qquad [11],$$

respectively, where $\text{Var}(\boldsymbol{\beta}_j)$ and $\text{Cov}(\boldsymbol{\beta}_i, \boldsymbol{\beta}_q)$ denote the variance of the $j$th eigenvector and the covariance between the $i$th and the $q$th $\mathbf{T}$ eigenvectors. In Eq. [10] and [11], the $\mathbf{T}$ eigenvectors are random, but the $\boldsymbol{\alpha}' = [\ \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_t\ ]$ values are fixed.

### Expectation and Variance of Vectors $\mathbf{b}_{KN}$ and $\mathbf{b}_M$

By Eq. [10] and [11], the expectation and variance of $\mathbf{b}_{KN}$ are

$$E(\mathbf{b}_{KN}) = \mathbf{K}E(\mathbf{b}) \qquad [12] \text{ and}$$

$$\text{Var}(\mathbf{b}_{KN}) = \mathbf{K}\text{Var}(\mathbf{b})\mathbf{K}' \qquad [13],$$

respectively, and the expectation and variance of $\mathbf{b}_M$ are

$$E(\mathbf{b}_M) = \mathbf{K}_M E(\mathbf{b}) \qquad [14] \text{ and}$$

$$\text{Var}(\mathbf{b}_M) = \mathbf{K}_M \text{Var}(\mathbf{b})\mathbf{K}'_M \qquad [15],$$

where $E(\mathbf{b})$ and $\text{Var}(\mathbf{b})$ are the expectation and variance of $\mathbf{b}$. This means that for finding the expectations of $\mathbf{b}_{KN}$ and $\mathbf{b}_M$, we need only $E(\mathbf{b})$ and to find the variances of $\mathbf{b}_{KN}$ and $\mathbf{b}_M$, we only need to find $\text{Var}(\mathbf{b})$.

## Estimator of Vector b

We denoted the estimators of $\mathbf{b}$ by $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{w}$, where $\hat{\mathbf{P}}^{-1}$ and $\hat{\mathbf{G}}$ are estimators of the inverse phenotypic covariance matrix ($\mathbf{P}^{-1}$), and of the covariance matrix of true breeding values ($\mathbf{G}$), both defined in Eq. [1]; $\mathbf{w}$ is the vector of economic values. According to Eq. [9], $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{w}$ can be written as

$$\hat{\mathbf{b}} = \sum_{j=1}^{t} \hat{\alpha}_j \hat{\boldsymbol{\beta}}_j \qquad [16],$$

where $\hat{\alpha}_j$ is the $j$th element of the vector $\hat{\boldsymbol{\alpha}} = \hat{\mathbf{B}}'\hat{\mathbf{b}}$, which is a least square estimator (unbiased and with minimum variance) of $\hat{\boldsymbol{\alpha}} = [\ \alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_t\ ]$ (Crossa and Cerón-Rojas, 2011), and $\hat{\boldsymbol{\beta}}_j$ (Eq. [8]) is a maximum likelihood estimator (asymptotically unbiased) of the eigenvector $\boldsymbol{\beta}_j$ of Eq. [7].

## Expectation and Variance of the Estimator of Vector b

Suppose that vectors $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{b}}$ are independent, then the expectation of $\hat{\mathbf{b}}$ can be written as

$$E(\hat{\mathbf{b}}) = \sum_{j=1}^{t} E(\hat{\alpha}_j \hat{\boldsymbol{\beta}}_j) = \sum_{j=1}^{t} \alpha_j \boldsymbol{\beta}_j = \mathbf{b} \qquad [17],$$

where $t$ = number of traits. That is, $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{w}$ is an asymptotic unbiased estimator of $\mathbf{b}$.

In the asymptotic context, the variance of $\hat{\mathbf{b}}$ can be written as

$$\mathrm{Var}(\hat{\mathbf{b}}) = \sum_{j=1}^{t} \hat{\alpha}_j^2 \mathrm{Var}(\hat{\boldsymbol{\beta}}_j) + 2\sum_{i=1}^{t-1}\sum_{q=i+1}^{t} \hat{\alpha}_i\hat{\alpha}_q \mathrm{Cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_q) \qquad [18],$$

where, by the results obtained by Anderson (1999) in the context of canonical correlations, the right terms of Eq. [18] associated with the eigenvectors of Eq. [8] can be written as

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_i) = \frac{1}{2n}\boldsymbol{\beta}_i\boldsymbol{\beta}_i' + \frac{1}{n}(1-\lambda_i^2)\sum_{j\neq i}^{t} \frac{\lambda_i^2 + \lambda_j^2 - 2\lambda_i^2\lambda_j^2}{(\lambda_i^2 - \lambda_j^2)^2}\boldsymbol{\beta}_i\boldsymbol{\beta}_j' \qquad [19]$$

and, for $i \neq q$,

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_q) = \frac{(1-\lambda_q^2)(1-\lambda_i^2)(\lambda_i^2 + \lambda_q^2)}{n(\lambda_i^2 - \lambda_q^2)^2}\boldsymbol{\beta}_q\boldsymbol{\beta}_i' \qquad [20],$$

where $n$ is the number of individuals or genotypes. Then, because Eq. [19] and [20] are divided by $n$, the estimator $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{w}$ has minimum variance in the asymptotic context, and when $n$ tends to infinity, $\mathrm{Var}(\hat{\mathbf{b}})$ (Eq. [18]) tends to the null matrix. So $\hat{\mathbf{b}}$ is a good estimator of $\mathbf{b}$, which implies that the SPSI is a good predictor of the net genetic merit of plants and animals.

Note that Eq. [20] converges more quickly to the one null matrix than Eq. [19] because the latter equation contains more terms than Eq. [20]. Then, when the number of traits and genotypes is very high, a good approximation to Eq. [18] shall be

$$\mathrm{Var}(\hat{\mathbf{b}}) \approx \sum_{j=1}^{t} \hat{\alpha}_j^2 \mathrm{Var}(\hat{\boldsymbol{\beta}}_j) \qquad [21].$$

In Eq. [21] we have written the variance of $\hat{\mathbf{b}}$ only in terms of the variances of the eigenvectors of Eq. [8]. In practice, Eq. [21] is a good option for obtaining $\mathrm{Var}(\hat{\mathbf{b}})$ because it is a symmetric matrix, while Eq. [18] can be an asymmetric matrix because $\sum_{i=1}^{t-1}\sum_{k=i+1}^{t} \hat{\alpha}_i\hat{\alpha}_k \mathrm{Cov}(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_k)$ is generally an asymmetric matrix. When the number of genotypes is low, this can substantially affect $\sum_{j=1}^{t} \hat{\alpha}_j^2 \mathrm{Var}(\hat{\boldsymbol{\beta}}_j)$, then $\mathrm{Var}(\hat{\mathbf{b}})$ will be an asymmetric matrix, that is, it will not be a covariance matrix.

### Expectation and Variance of the Estimator of Vectors $\mathbf{b}_{KN}$ and $\mathbf{b}_{M}$

The estimators of $\mathbf{b}_{KN}$ and $\mathbf{b}_{M}$ were denoted by $\hat{\mathbf{b}}_{KN} = \mathbf{K}\hat{\mathbf{b}}$ and $\hat{\mathbf{b}}_{M} = \mathbf{K}_{M}\hat{\mathbf{b}}$, respectively, and their expectation and variance as

$$E(\hat{\mathbf{b}}_{KN}) = \mathbf{K}E(\hat{\mathbf{b}}) \text{ and } \mathrm{Var}(\hat{\mathbf{b}}_{KN}) = \mathbf{K}\mathrm{Var}(\hat{\mathbf{b}})\mathbf{K}' \qquad [22]$$

and

$$E(\hat{\mathbf{b}}_{M}) = \mathbf{K}_{M}E(\hat{\mathbf{b}}) \text{ and } \mathrm{Var}(\hat{\mathbf{b}}_{M}) = \mathbf{K}_{M}\mathrm{Var}(\hat{\mathbf{b}})\mathbf{K}_{M}' \qquad [23],$$

where $E(\hat{\mathbf{b}})$ and $\mathrm{Var}(\hat{\mathbf{b}})$ are the expectation and variance of $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{w}$. In Eq. [22] and [23], $\hat{\mathbf{b}}$ is random, but matrices $\mathbf{K}$ and $\mathbf{K}_{M}$ are fixed. Note that because $\mathrm{Var}(\hat{\mathbf{b}})$ is divided by $n$ (Eq. [19] and [20]), $\mathrm{Var}(\hat{\mathbf{b}}_{KN})$ and $\mathrm{Var}(\hat{\mathbf{b}}_{M})$ have minimum variance in the asymptotic context; when $n$ tends to infinity, $\mathrm{Var}(\hat{\mathbf{b}}_{KN})$ and $\mathrm{Var}(\hat{\mathbf{b}}_{M})$ tend to the null matrix.

## Numerical Examples

### Dataset 1

The estimated phenotypic ($\hat{\mathbf{P}}$) and genetic ($\hat{\mathbf{G}}$) covariance matrix values were presented in the materials and methods section. In that section, we presented two matrices, $\mathbf{C} = \hat{\mathbf{G}}\mathbf{U}$ and $\mathbf{M} = \mathbf{CD}$, the vector of economic weight ($\mathbf{w}$) and the vector of predetermined restrictions ($\mathbf{d}$). From these data,

$$\hat{\mathbf{b}}' = \begin{bmatrix} 2.15 & -1.03 & 2.51 & -0.73 \end{bmatrix},$$

$$\hat{\mathbf{b}}_{KN}' = \begin{bmatrix} -0.09 & -0.38 & 1.31 & -0.72 \end{bmatrix}, \text{ and}$$

$$\hat{\mathbf{b}}_{M}' = \begin{bmatrix} 1.83 & 0.68 & 4.42 & -1.01 \end{bmatrix}.$$

Suppose that the selection intensity is 10% ($k = 1.75$), then the estimated values of Eq. [2], [4], and [6] are

$$\hat{\mathbf{E}}'_{PSI} = \begin{bmatrix} 3.00 & -2.05 & 0.02 & -3.62 \end{bmatrix},$$

$$\hat{\mathbf{E}}'_{KN} = \begin{bmatrix} 0 & 0 & 0 & -11.34 \end{bmatrix}, \text{ and}$$

$$\hat{\mathbf{E}}'_{M} = \begin{bmatrix} 1.36 & -0.45 & 0.91 & -6.02 \end{bmatrix}, \text{ respectively.}$$

### Estimates of the Expectation and Variance of the Estimators of Vectors b, $b_{KN}$, and $b_M$

For Dataset 1, the values of vector $\hat{\boldsymbol{\alpha}} = \hat{\mathbf{B}}'\hat{\mathbf{b}}$ were $\hat{\boldsymbol{\alpha}} = \begin{bmatrix} 0.61 & -2.90 & 1.54 & 1.17 \end{bmatrix}$, and the estimated expectations of $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$ and $\hat{\mathbf{b}}_{M}$ were

$$\hat{E}(\hat{\mathbf{b}})' = \begin{bmatrix} 2.15 & -1.03 & 2.51 & -0.73 \end{bmatrix},$$

$$\hat{E}(\hat{\mathbf{b}}_{KN})' = \begin{bmatrix} -0.09 & -0.38 & 1.31 & -0.72 \end{bmatrix}, \text{ and}$$

$$\hat{E}(\hat{\mathbf{b}}_{M})' = \begin{bmatrix} 1.83 & 0.68 & 4.42 & -1.01 \end{bmatrix}, \text{ respectively.}$$

The estimated variances of $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$ and $\hat{\mathbf{b}}_{M}$ were

$$\hat{\text{Var}}(\hat{\mathbf{b}}) = \frac{1}{3330}\begin{bmatrix} 79.18 & -4.66 & -0.39 & 6.16 \\ -4.66 & 73.00 & -9.23 & 15.31 \\ -0.39 & -9.23 & 56.32 & 15.97 \\ 6.16 & 15.31 & 15.97 & 49.66 \end{bmatrix},$$

$$\hat{\text{Var}}(\hat{\mathbf{b}}_{KN}) = \frac{1}{3330}\begin{bmatrix} 0.69 & 3.00 & -10.47 & 5.71 \\ 3.00 & 13.13 & -45.75 & 24.95 \\ -10.47 & -45.75 & 159.42 & -86.92 \\ 5.71 & 24.95 & -86.92 & 47.39 \end{bmatrix}, \text{ and}$$

$$\hat{\text{Var}}(\hat{\mathbf{b}}_{M}) = \frac{1}{3330}\begin{bmatrix} 58.54 & 43.04 & 46.81 & 15.03 \\ 43.04 & 39.65 & -0.90 & 28.78 \\ 46.81 & -0.90 & 193.14 & -66.15 \\ 15.03 & 28.78 & -66.15 & 43.10 \end{bmatrix}.$$

## Dataset 2

The estimated phenotypic ($\hat{\mathbf{P}}$) and genetic ($\hat{\mathbf{G}}$) covariance matrix values and the matrices $\mathbf{C} = \hat{\mathbf{G}}\mathbf{U}$ and $\mathbf{M} = \mathbf{CD}$ for Dataset 2 were presented in the Materials and methods section along with the vector of economic weight ($\mathbf{w}$) and the vector of predetermined restrictions ($\mathbf{d}$). The estimated coefficient vector values were

$$\hat{\mathbf{b}}' = \begin{bmatrix} 0.10 & -0.35 & -0.36 & -3.23 \end{bmatrix},$$

$$\hat{\mathbf{b}}'_{KN} = \begin{bmatrix} 0.15 & -0.01 & 0.02 & -0.23 \end{bmatrix}, \text{ and}$$

$$\hat{\mathbf{b}}'_{M} = \begin{bmatrix} -0.09 & 0.05 & -0.03 & -0.66 \end{bmatrix}.$$

The selection intensity was 10% ($k = 1.75$) and the estimated values of Eq. [2], [4], and [6] were

$$\hat{\mathbf{E}}'_{PSI} = \begin{bmatrix} -0.33 & -8.77 & -8.67 & -0.53 \end{bmatrix},$$

$$\hat{\mathbf{E}}'_{KN} = \begin{bmatrix} 0 & 0 & 0 & -0.48 \end{bmatrix}, \text{ and}$$

$$\hat{\mathbf{E}}'_{M} = \begin{bmatrix} -0.23 & 0.12 & -1.16 & -0.67 \end{bmatrix}, \text{ respectively.}$$

### Estimates of the Expectation and Variance of the Estimators of Vectors b, $b_{KN}$, and $b_M$

The vector values of $\hat{\boldsymbol{\alpha}} = \hat{\mathbf{B}}'\hat{\mathbf{b}}$ were $\hat{\boldsymbol{\alpha}}' = \begin{bmatrix} -0.82 & 1.18 & -2.91 & 0.34 \end{bmatrix}$, and the estimated expectation values of $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$, and $\hat{\mathbf{b}}_{M}$ were

$$\hat{E}(\hat{\mathbf{b}})' = \begin{bmatrix} 0.10 & -0.35 & -0.36 & -3.23 \end{bmatrix},$$

$$\hat{E}(\hat{\mathbf{b}}_{KN})' = \begin{bmatrix} 0.15 & -0.01 & 0.02 & -0.23 \end{bmatrix}, \text{ and}$$

$$\hat{E}(\hat{\mathbf{b}}_{M})' = \begin{bmatrix} -0.09 & 0.05 & -0.03 & -0.66 \end{bmatrix}, \text{ respectively.}$$

The estimated variances of $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$, and $\hat{\mathbf{b}}_{M}$ were

$$\hat{\text{Var}}(\hat{\mathbf{b}}) = \frac{1}{250}\begin{bmatrix} 28.35 & -28.21 & 52.37 & -3.77 \\ -28.21 & 99.91 & -160.63 & 9.24 \\ 52.37 & -160.63 & 301.25 & -19.05 \\ -3.77 & 9.24 & -19.05 & 5.23 \end{bmatrix},$$

$$\hat{\text{Var}}(\hat{\mathbf{b}}_{KN}) = \frac{1}{250}\begin{bmatrix} 1763.29 & -148.09 & 197.79 & -2812.88 \\ -148.09 & 12.44 & -16.61 & 236.24 \\ 197.79 & -16.61 & 22.19 & -315.52 \\ -2812.88 & 236.24 & -315.52 & 4487.24 \end{bmatrix},$$

and

$$\hat{\text{Var}}(\hat{\mathbf{b}}_{M}) = \frac{1}{250}\begin{bmatrix} 4486.35 & -671.01 & 644.98 & -1434.25 \\ -671.01 & 100.92 & -96.74 & 203.54 \\ 644.98 & -96.74 & 92.86 & -200.91 \\ -1434.25 & 203.54 & -200.91 & 671.93 \end{bmatrix},$$

respectively.

### Why are Matrices K and $K_M$ Fixed?

In Eq. [21] and [22], we assumed that matrices $\mathbf{K}$ and $\mathbf{K}_M$ were fixed. We made this assumption because matrices $\mathbf{K}$ and $\mathbf{K}_M$ are projectors. First, note that matrices $\mathbf{K} = [\mathbf{I} - \mathbf{Q}]$ and $\mathbf{Q} = \mathbf{P}^{-1}\mathbf{C}(\mathbf{C}'\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}'$ are idempotent ($\mathbf{K} = \mathbf{K}^2$ and $\mathbf{Q} = \mathbf{Q}^2$) and unique (Searle, 1966); they are also orthogonal, that is, $\mathbf{KQ} = \mathbf{QK} = 0$. Matrix $\mathbf{Q}$ projects vector $\mathbf{b}$ into a space generated by the columns of matrix $\mathbf{C}$ because of the restriction $\mathbf{C}'\mathbf{b} = 0$ used when $\Psi$ is maximized with respect to $\mathbf{b}$. Matrix $\mathbf{K}$ projects $\mathbf{b}$ into a space perpendicular to the space generated by the $\mathbf{C}$ matrix columns (Rao, 2002). Furthermore, because of the restriction $\mathbf{C}'\mathbf{b} = 0$, matrix $\mathbf{K}$ projects $\mathbf{b}$ to a space smaller than the original space of $\mathbf{b}$. The space reduction

where matrix $\mathbf{K}$ projects $\mathbf{b}$ is equal to the number of zeros that appear in Eq. [4]. In the two numerical examples, we found that $\hat{\mathbf{E}}'_{KN} = \begin{bmatrix} 0 & 0 & 0 & -11.34 \end{bmatrix}$, for Dataset 1, and $\hat{\mathbf{E}}'_{KN} = \begin{bmatrix} 0 & 0 & 0 & -0.48 \end{bmatrix}$, for Dataset 2, because we imposed three restrictions on both datasets.

Additionally, note that matrices $\mathbf{Q}_M = \mathbf{P}^{-1}\mathbf{M}(\mathbf{M}'\mathbf{P}^{-1}\mathbf{M})^{-1}\mathbf{M}'$ and $\mathbf{K}_M = [\mathbf{I} - \mathbf{Q}_M]$ of Eq. [5] had the same function as matrices $\mathbf{Q} = \mathbf{P}^{-1}\mathbf{C}(\mathbf{C}'\mathbf{P}^{-1}\mathbf{C})^{-1}\mathbf{C}'$ and $\mathbf{K} = [\mathbf{I} - \mathbf{Q}]$. However, in this case, matrix $\mathbf{Q}_M$ projects $\mathbf{b}$ into a space generated by the columns of matrix $\mathbf{M}$ because of the restriction $\mathbf{M}'\mathbf{b} = 0$ that is introduced when $\Psi_M$ is maximized with respect to $\mathbf{b}$. Matrix $\mathbf{K}_M$ projects $\mathbf{b}$ to a space that is perpendicular to the space generated by the columns of matrix $\mathbf{M}$. Then, matrices $\mathbf{Q}_M$ and $\mathbf{K}_M$ are both projectors, that is, they are idempotent ($\mathbf{K}_M = \mathbf{K}_M{}^2$ and $\mathbf{Q}_M = \mathbf{Q}_M{}^2$) and unique ($\mathbf{K}_M\mathbf{Q}_M = \mathbf{Q}_M\mathbf{K}_M = 0$).

Then, as $\mathbf{K}$ and $\mathbf{K}_M$ are projectors, their main function is to transform vector $\mathbf{b}$ into vectors $\mathbf{b}_{KN}$ and $\mathbf{b}_M$, and to determine the sampling statistical properties of $\hat{\mathbf{b}}_{KN}$ and $\hat{\mathbf{b}}_M$, we can assume that $\mathbf{K}$ and $\mathbf{K}_M$ are fixed.

## Statistical Sampling Properties of $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{w}$

The sampling properties of $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{w}$ were derived directly, and the only condition is that $\hat{\mathbf{b}}$ can be written as Eq. [16], that is, assuming that $\hat{\mathbf{b}}$ is in the space generated by the eigenvectors of matrix $\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}$ (Eq. [8]). When this is the case, the sampling properties of the estimator of eigenvalues and eigenvectors of matrix $\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}$ can be applied to find the sampling properties of $\hat{\mathbf{b}}$.

In both datasets, we used Eq. [21] to obtain $\hat{\mathrm{Var}}(\hat{\mathbf{b}})$, $\hat{\mathrm{Var}}(\hat{\mathbf{b}}_{KN})$, and $\hat{\mathrm{Var}}(\hat{\mathbf{b}}_M)$ because the estimated values of matrix $\sum_{i=1}^{t-1}\sum_{k=i+1}^{t}\hat{\alpha}_i\hat{\alpha}_k\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\beta}}_k)$ for Datasets 1 and 2 were

$$\begin{bmatrix} 0.0093 & -0.0160 & -0.0056 & 0.0097 \\ 0.0132 & -0.0075 & 0.0044 & -0.0075 \\ 0.0020 & -0.0020 & -0.0038 & 0.0007 \\ 0.0077 & -0.0096 & 0.0048 & 0.0019 \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} 0.0088 & -0.0508 & -0.0288 & 0.0143 \\ -0.0430 & 0.1515 & 0.0894 & -0.0356 \\ 0.0191 & -0.2856 & -0.1585 & 0.0462 \\ 0.0023 & 0.0186 & 0.0097 & -0.0019 \end{bmatrix},$$

respectively. In Dataset 2, the values 0.1515, −0.2856, and −0.1585 were relatively high; however, the rest of the values were low. In Dataset 1, all the values were very low. The values of $\sum_{i=1}^{t-1}\sum_{k=i+1}^{t}\hat{\alpha}_i\hat{\alpha}_k\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i,\hat{\boldsymbol{\beta}}_k)$ will be similar to those in Dataset 1 if the numbers of traits and genotypes are very large; in this case, we can assume that Eq. [21] is a good approximation to the values of $\hat{\mathrm{Var}}(\hat{\mathbf{b}})$, $\hat{\mathrm{Var}}(\hat{\mathbf{b}}_{KN})$, and $\hat{\mathrm{Var}}(\hat{\mathbf{b}}_M)$.

## Importance of the Sampling Properties of Vectors $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$ and $\hat{\mathbf{b}}_M$

The method we proposed for finding the sampling properties of $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$, and $\hat{\mathbf{b}}_M$ is very simple, very general, and easy to program. The variances of $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$, and $\hat{\mathbf{b}}_M$ are useful because they can be used to construct confidence intervals or confidence regions for $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$, and $\hat{\mathbf{b}}_M$, which is important to complete the analysis of a selection process. However, because the formulas were developed in the asymptotic context, they require a large number of genotypes.

## Relationship of Vectors b and $\mathbf{b}_{KN}$ to the Restrictive Eigen Selection Index Method

The statistical sampling properties of coefficients $\mathbf{b}$ and $\mathbf{b}_{KN}$ are related to the coefficient of the restricted eigen selection index (RESIM) developed by Cerón-Rojas et al. (2008b) in the canonical correlation theory context. The authors showed that $\mathbf{b}$ is in the space of the $\mathbf{T} = \mathbf{P}^{-1}\mathbf{G}$ eigenvectors (Eq. [7]) and can be written as in Eq. [16]. Cerón-Rojas et al. (2008b) also showed that, under certain assumptions, $\mathbf{b}_{KN}$ could be equal to the RESIM vector of coefficient. The basic theory of RESIM was developed within the framework of the canonical correlation theory, and that is why the statistical properties of its coefficient are known. The RESIM uses the elements of the first eigenvector for determining the proportion of each trait contributing to the selection index, and the square root of the first eigenvalue (singular value) is the selection response. The original ideas presented in this article were inspired by the theory developed in RESIM.

In addition to the relationships of $\mathbf{b}$ and $\mathbf{b}_{KN}$ with the RESIM presented in this paper, we would like to point out the relationship of two indices developed in the molecular marker-assisted selection (MAS) context: the Lande and Thompson (1990) and Dekkers (2007) selection indices with the molecular eigen selection index method (MESIM) of Cerón-Rojas et al. (2008a) which, besides the phenotypic information, also uses molecular information to predict the net genetic merit. In theory, MESIM is very similar to RESIM. Because the indices of Lande and Thompson (1990) and Dekkers (2007) are a direct application of the SPSI to MAS, and the MESIM was developed within the canonical correlation theory context, it is possible to use the method developed in this paper for determining the statistical sampling properties of the estimator of the coefficients of the Lande and Thompson (1990) and Dekkers (2007) selection index using the MESIM theory in a similar manner as we did it for $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$, and $\hat{\mathbf{b}}_M$ using the RESIM and the canonical correlation theory. This shows that, in effect, the method we proposed for finding the sampling properties of $\hat{\mathbf{b}}$, $\hat{\mathbf{b}}_{KN}$, and $\hat{\mathbf{b}}_M$ is general.

As for the economic weights, these should be determined by the breeder or by some mathematical estimation. When the economic weights are known, there is no problem. Furthermore, so far there is no mathematical method for determining the economic weights for the three selection indices described in this paper; a possible solution would be to use selection indices such as RESIM and the eigen selection index method (Cerón-Rojas et al., 2008a,b), which do not require economic weights.

## CONCLUSIONS

Using the canonical correlation theory, we developed an asymptotic method for determining the statistical sampling properties of the estimator of the coefficients of the Smith phenotypic selection index. Also, under certain assumptions applied to the Smith phenotypic selection index coefficient estimator, we obtained the sampling properties of the estimator of the restricted phenotypic selection index, and the predetermined proportional gains phenotypic selection index. The method indicated that when the number of genotypes is large, the estimators of the coefficients of the three indices were unbiased and with minimal variance. We concluded that this method could be used to obtain the sampling properties of the estimator of the coefficients of the three indices.


### Acknowledgments

The authors are grateful to México's Consejo Nacional de Ciencia y Tecnología (CONACYT) for partially funding this research.


### 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

Anderson, T.W. 1999. Asymptotic theory for canonical correlation analysis. J. Multivariate Anal. 70:1–29. doi:10.1006/jmva.1999.1810

Anderson, T.W. 2003. An introduction to multivariate statistical analysis. 3rd ed. John Wiley & Sons, New Jersey.

Baker, R.J. 1974. Selection indexes without economic weights for animal breeding. Can. J. Anim. Sci. 54:1–8. doi:10.4141/cjas74-001

Bulmer, M.G. 1980. The mathematical theory of quantitative genetics. Lectures in biomathematics. University of Oxford, Clarendon Press, England.

Cerón-Rojas, J.J., J. Sahagún-Castellanos, F. Castillo-González, A. Santacruz-Varela, I. Benítez-Riquelme, and J. Crossa. 2008a. A molecular selection index method based on eigenanalysis. Genetics 180:547–557. doi:10.1534/genetics.108.087387

Cerón-Rojas, J.J., J. Sahagún-Castellanos, F. Castillo-González, A. Santacruz-Varela, and J. Crossa. 2008b. A restricted selection index method based on eigenanalysis. J. Agric. Biol. Environ. Stat. 13:421–438. doi:10.1198/108571108X378911

Crossa, J., and J.J. Cerón-Rojas. 2011. Multi-trait multi-environment genome-wide molecular marker selection indices. J. Indian Soc. Agric. Stat. 62:125–142.

Dekkers, J.C.M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331–341. doi:10.1111/j.1439-0388.2007.00701.x

Harris, D.L. 1964. Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20:46–72. doi:10.2307/2527617

Harville, D.A. 1975. Index selection with proportionality constrains. Biometrics 31:223–225. doi:10.2307/2529722

Hayes, J.F., and W.G. Hill. 1980. A reparameterization of a genetic selection index to locate its sampling properties. Biometrics 36:237–248. doi:10.2307/2529975

Itoh, Y., and Y. Yamada. 1987. Comparisons of selection indices achieving predetermined proportional gains. Genet. Sel. Evol. 19:69–82. doi:10.1186/1297-9686-19-1-69

Kempthorne, O., and A.W. Nordskog. 1959. Restricted selection indices. Biometrics 15:10–19. doi:10.2307/2527598

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756.

Lin, C.Y. 2005. A simultaneous procedure for deriving selection indexes with multiple restrictions. J. Anim. Sci. 83:531–536.

Mallard, J. 1972. The theory and computation of selection indices with constraints: A critical synthesis. Biometrics 28:713–735. doi:10.2307/2528758

Rao, C.R. 2002. Linear statistical inference and its applications. 2nd ed. John Wiley & Sons, New York.

Searle, S.R. 1966. Matrix algebra for the biological sciences. John Wiley & Sons, New York.

Smith, H.F. 1936. A discriminant function for plant selection In: Papers on quantitative genetics and related topics. Department of Genetics, North Carolina State College, Raleigh, NC. p. 466–476.

Tallis, G.M. 1960. The sampling errors of estimated genetic regression coefficients and the error of predicted genetic gains. Aust. J. Stat. 2:66–77. doi:10.1111/j.1467-842X.1960.tb00127.x

Tallis, G.M. 1962. A selection index for optimum genotype. Biometrics 18:120–122. doi:10.2307/2527716

Tallis, G.M. 1985. Constrained selection. Jap. J. Genet. 60:151–155. doi:10.1266/jjg.60.151

Williams, J.S. 1962. Some statistical properties of a genetic selection index. Biometrika 9:325–337. doi:10.1093/biomet/49.3-4.325