Chapter 5
Linear Genomic Selection Indices

Abstract The linear genomic selection index (LGSI) is a linear combination of genomic estimated breeding values (GEBVs) used to predict the individual net genetic merit and select individual candidates from a nonphenotyped testing population as parents of the next selection cycle. In the LGSI, phenotypic and marker data from the training population are fitted into a statistical model to estimate all individual available genome marker effects; these estimates can then be used in subsequent selection cycles to obtain GEBVs that are predictors of breeding values in a testing population for which there is only marker information. The GEBVs are obtained by multiplying the estimated marker effects in the training population by the coded marker values obtained in the testing population in each selection cycle. Applying the LGSI in plant or animal breeding requires the candidates to be genotyped for selection to obtain the GEBV, and predicting and ranking the net genetic merit of the candidates for selection using the LGSI. We describe the LGSI and show that it is a direct application of the linear phenotypic selection index theory in the genomic selection context; next, we present the combined LGSI (CLGSI), which uses phenotypic and GEBV information jointly to predict the net genetic merit. The CLGSI can be used only in training populations when there are phenotypic and maker information, whereas the LGSI is used in testing populations where there is only marker information. We validate the theoretical results of the LGSI and CLGSI using real and simulated data.

5.1 The Linear Genomic Selection Index

5.1.1 Basic Conditions for Constructing the LGSI

Conditions described in Chap. 4 (Sect. 4.1.1) for constructing a valid linear molecular selection index (LMSI), are also necessary for the linear genomic selection index (LGSI); however, in addition to those conditions, the LGSI also requires:

1. All marker effects to be estimated simultaneously in the training population.
2. The estimated marker effects to be used in subsequent selection cycles to obtain GEBVs that are predictors of the individual breeding values in the testing population (candidates for selection) for which there is only marker information.

J. J. Céron-Rojas, J. Crossa, *Linear Selection Indices in Modern Plant Breeding*,
https://doi.org/10.1007/978-3-319-91223-3_5

3. The GEBV values to be composed entirely of the additive genetic effects.
4. Phenotypes to be used to estimate all marker effects in the training population, not to make selections in the testing population (Heffner et al. 2009; Lorenz et al. 2011).

5.1.2 Genomic Breeding Values and Marker Effects

The breeding value (g_i) is the average additive effects of the genes an individual receives from both parents; thus, it is a function of the genes transmitted from parents to progeny and is the only component that can be selected and, therefore, the main component of interest in breeding programs (Mrode 2005). The ith phenotypic value (y_i) can be denoted as $y_i = g_i + e_i$, where g_i is the *breeding value* and e_i the residual. Basic assumptions for g_i and e_i are: both g_i and e_i have normal distribution with expectation equal to zero and variance $\sigma^2_{g_i}$ and $\sigma^2_{e_i}$ respectively. This means that $y_i = \mu_i + g_i + e_i$ is a linear mixed model (Mrode 2005; Searle et al. 2006), where μ_i is the mean of y_i.

Let $\mathbf{y}'_i = [y_{i1} \;\; y_{i2} \;\; \cdots \;\; y_{in}]$ be a vector $1 \times n$ of observations in the ith trait and let $\mathbf{g}'_i = [g_{i1} \;\; g_{i2} \;\; \cdots \;\; g_{in}]$ be a vector $1 \times n$ of unobservable breeding values associated with $\mathbf{y}_i$; then $\mathbf{y}_i$ can be written as

$$\mathbf{y}_i = \mathbf{1}\mu_i + \mathbf{Z}\mathbf{g}_i + \mathbf{e}_i, \tag{5.1}$$

where μ_i is the mean of the ith trait, $\mathbf{1}$ is a vector $n \times 1$ of 1s, $\mathbf{Z}$ is a design matrix of 0s and 1s, $\mathbf{g}_i \sim \text{MVN}(\mathbf{0}, \mathbf{A}\sigma^2_{g_i})$ is a vector of breeding values, and $\mathbf{e}_i \sim \text{MVN}(\mathbf{0}, \mathbf{I}_n\sigma^2_{e_i})$ is a vector of residuals; $\mathbf{0}$ is the mean and $\mathbf{A}\sigma^2_{g_i}$ and $\mathbf{I}_n\sigma^2_{e_i}$ the covariance matrix of $\mathbf{g}_i$ and $\mathbf{e}_i$ respectively; $\mathbf{A}$ is the numerical relationship matrix (Mrode 2005) and $\mathbf{I}_n$ an identity matrix $n \times n$; $\sigma^2_{g_i}$ and $\sigma^2_{e_i}$ are the additive and residual variances associated with g_i and e_i; and MVN stands for multivariate normal distribution.

Suppose that $\mathbf{A}$, $\mathbf{Z}$, μ_i, $\sigma^2_{g_i}$, and $\sigma^2_{e_i}$ are known; then, according to Mrode (2005), the best linear unbiased predictor (BLUP) of $\mathbf{g}_i$ can be written as

$$\widehat{\mathbf{g}}_i = \sigma^2_{g_i}\mathbf{A}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{1}\mu_i), \tag{5.2}$$

where $\mathbf{V}^{-1}$ is the inverse matrix of the variance of $\mathbf{y}_i$, i.e., $Var(\mathbf{y}_i) = \sigma^2_{g_i}\mathbf{Z}\mathbf{A}\mathbf{Z}' + \mathbf{I}_n\sigma^2_{e_i} = \mathbf{V}$. In the context of animal breeding, Eq. (5.2) is considered a univariate linear phenotypic selection index (LPSI) (Mrode 2005) and is used to rank and select individuals as parents of the next generation in the context of one trait. Equation (5.2) can be extended to the multi-trait phenotypic selection index case, but to predict the net genetic merit ($H = \mathbf{w}'\mathbf{g}$, see Chap. 2 for details) it would be necessary to construct linear combinations of the predicted values of $\mathbf{g}_i$ associated with the traits of interest as was described in the Foreword of this book.

The vector of the individual genomic breeding values ($\boldsymbol{\gamma}_i$) associated with the ith characteristic ($i = 1, 2, \ldots, t$; t = number of traits) of the candidates for selection can be written as

$$\boldsymbol{\gamma}_i = \mathbf{X}\mathbf{u}_i, \tag{5.3}$$

where $\mathbf{X}$ is an $n \times m$ matrix (n = number of observations and m = number of markers in the population) of coded marker values ($2 - 2p$, $1 - 2p$, and $-2p$ for genotypes AA, Aa, and aa respectively) associated with the additive effects of the quantitative trait loci (QTL) and $\mathbf{u}_i$ is an $m \times 1$ vector of the additive effects of the QTL associated with markers that affect the ith trait. It is assumed that $\boldsymbol{\gamma}_i$ has MVN with mean $\mathbf{0}$ and variance $\mathbf{G}\sigma^2_\gamma$, i.e., $\boldsymbol{\gamma}_i \sim$ MVN ($\mathbf{0}$, $\mathbf{G}\sigma^2_{\gamma_i}$), where $\sigma^2_{\gamma_i}$ is the additive genomic variance of $\boldsymbol{\gamma}_i$ and $\mathbf{G} = \mathbf{X}\mathbf{X}'/c$ is the $n \times n$ additive genomic relationship matrix between genotypes; $c = \sum_{j=1}^{m} 2p_j(1 - p_j)$ in an F_2 population, and $c = \sum_{j=1}^{m} 4p_j(1 - p_j)$ in a double haploid population; p is the frequency of allele A and $1 - p$ is the frequency of allele a in the jth marker ($j = 1, 2, \ldots, m$).

The additive genomic relationship matrix $\mathbf{G} = \mathbf{X}\mathbf{X}'/c$ has special properties. For example, in the asymptotic context, the expectation of matrix $\mathbf{G}$ is equal to the numerical relationship matrix $\mathbf{A}$, i.e., $E(\mathbf{G}) = \mathbf{A}$ (Habier et al. 2007; Van Raden 2008); this means that $\mathbf{G}$ is a particular realization of $\mathbf{A}$ and when the number of markers and genotypes increases in the training population, the value of $\mathbf{G}$ tends to concentrate around $\mathbf{A}$. Thus, it can be assumed that at the limit, when the number of markers and genotypes is very high, $\mathbf{G} = \mathbf{A}$ (Cerón-Rojas and Sahagún-Castellanos 2016).

The vector of genomic breeding values (Eq. 5.3) has a similar function in genomic selection as $\mathbf{g}_i$ in the phenotypic selection context. In addition, $\mathbf{g}_i$ can be written as $\mathbf{g}_i = \boldsymbol{\gamma}_i + \eta_i$, where $\eta_i = \mathbf{g}_i - \boldsymbol{\gamma}_i$ (Gianola et al. 2003). Also, note that

$$Cov(\mathbf{g}_i, \boldsymbol{\gamma}_i) = \sigma^2_{\gamma_i}, \tag{5.4}$$

i.e., the covariance between $\boldsymbol{\gamma}_i$ and $\mathbf{g}_i$ is equal to the variance of $\boldsymbol{\gamma}_i$ (Dekkers 2007).

Let $\mathbf{y}'_i = [y_{i1} \quad y_{i2} \quad \cdots \quad y_{in}]$ be a vector $1 \times n$ of observation of the ith trait in the training population and let $\boldsymbol{\gamma}'_i = [\gamma_{i1} \quad \gamma_{i2} \quad \cdots \quad \gamma_{in}]$ be a vector $1 \times n$ of unobservable genomic breeding values associated with $\mathbf{y}_i$; then, $\mathbf{y}_i$ can also be written as

$$\mathbf{y}_i = \mathbf{1}\mu_i + \mathbf{Z}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{5.5}$$

where μ_i is the mean of the ith trait, $\mathbf{1}$ is a vector $n \times 1$ of 1s, $\mathbf{Z}$ is a design matrix, $\boldsymbol{\gamma}_i \sim$ MVN ($\mathbf{0}$, $\mathbf{G}\sigma^2_\gamma$) and $\boldsymbol{\varepsilon}_i \sim$ MVN ($\mathbf{0}$, $\mathbf{I}_n\sigma^2_{\varepsilon_i}$) are vectors of genomic breeding values and of residuals respectively, and $\sigma^2_{\varepsilon_i}$ is the residual variance. $\mathbf{I}_n$, $\mathbf{G}$, and σ^2_γ were defined in Eqs. (5.2) and (5.3).

According to Eqs. (5.2) and (5.3), when μ_i, σ^2_γ and $\sigma^2_{\varepsilon_i}$ are known, the vector of GEBVs for the individuals with the ith trait can be obtained as

$$\widehat{\boldsymbol{\gamma}}_i = \sigma^2_{\gamma_i}\mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{1}\mu_i), \tag{5.6}$$

where the variance of $\mathbf{y}_i$ should now be written as $\mathbf{V} = \sigma^2_{\gamma_i}\mathbf{ZGZ}' + \mathbf{I}_n\sigma^2_{\varepsilon_i}$. In the context of genomic selection, Eq. (5.6) is considered a univariate LGSI and is used to rank and select individuals as parents of the next generation (Van Raden 2008; Togashi et al. 2011). Equation (5.6) is the BLUP of $\boldsymbol{\gamma}_i$ and can be extended to a multi-trait genomic selection index, but to predict the net genetic merit ($H = \mathbf{w}'\mathbf{g}$), it is necessary to construct an LGSI, which is a linear combination of $\boldsymbol{\gamma}_i$.

Although Eq. (5.6) is theoretically very important in LGSI, in practice we need to estimate the marker effects associated with all the traits of interest and to use these estimates in the testing population to obtain the GEBV of the candidates for selection. Let $\mathbf{u}' = [\mathbf{u}'_1 \quad \mathbf{u}'_2 \quad \cdots \quad \mathbf{u}'_t]$ be a vector $1 \times nt$ associated with t traits. In the univariate context, Van Raden (2008) showed that the ith vector $\mathbf{u}_i$ of marker effects in the training population can be estimated as

$$\widehat{\mathbf{u}}_i = c^{-1}\mathbf{X}'[\mathbf{G} + \upsilon\mathbf{I}_n]^{-1}(\mathbf{y}_i - \mathbf{1}\mu_i), \tag{5.7}$$

where $\upsilon = \dfrac{\sigma^2_{e_i}}{\sigma^2_{g_i}}$; $\sigma^2_{g_i}$, $\sigma^2_{e_i}$ and the other parameters were defined earlier. According to Ceron-Rojas et al. (2015), to estimate the vector $\mathbf{u}' = [\mathbf{u}'_1 \quad \mathbf{u}'_2 \quad \cdots \quad \mathbf{u}'_t]$ in the multi-trait context, Eq. (5.7) can be written as

$$\widehat{\mathbf{u}} = c^{-1}\mathbf{W}'_t[(\mathbf{I}_t \otimes \mathbf{G}) + (\mathbf{N} \otimes \mathbf{I}_n)]^{-1}(\mathbf{y} - \boldsymbol{\mu} \otimes \mathbf{1}), \tag{5.8}$$

where $\mathbf{W}_t = \mathbf{I}_t \otimes \mathbf{X}$, "$\otimes$" denotes the Kronecker product (Schott 2005), c and $\mathbf{X}$ were defined in Eq. (5.3); $\mathbf{N} = \mathbf{RC}^{-1}$, where$\mathbf{R}$ and $\mathbf{C}$ are the residual and breeding value covariance matrices for t traits respectively; $\mathbf{y}' = [\mathbf{y}'_1 \quad \mathbf{y}'_2 \quad \cdots \quad \mathbf{y}'_t] \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V})$ is a vector of size $1 \times tn$, with covariance matrix $\mathbf{V} = \mathbf{C} \otimes \mathbf{G} + \mathbf{R} \otimes \mathbf{I}_n$; $\mathbf{I}_t$ is an identity matrix of size $t \times t$ and $\mathbf{I}_n$ was defined earlier; $\boldsymbol{\mu}' = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_t]$ is a vector $1 \times t$ of means associated with vector $\mathbf{y}$, and $\mathbf{1}$ is a vector $n \times 1$ of 1s. In this case, the estimator of the vector of sub-vectors of genomic breeding values $\boldsymbol{\gamma}' = [\boldsymbol{\gamma}_1 \quad \boldsymbol{\gamma}_2 \quad \ldots \quad \boldsymbol{\gamma}_t]$ in the testing population can be obtained as

$$\widehat{\boldsymbol{\gamma}} = \mathbf{W}_t\widehat{\mathbf{u}}. \tag{5.9}$$

Equation (5.9) is the vector of GEBVs for the multi-trait case. Thus, in the testing population, in Eq. (5.9), only the coded values in matrix $\mathbf{X}$ change, whereas $\widehat{\mathbf{u}}$ is the same in each selection cycle. Note that to obtain Eqs. (5.7) and (5.8), we assumed that $\boldsymbol{\mu}$, $\mathbf{C}$, and $\mathbf{R}$ are known.

We indicated that the genomic breeding values have normal distribution (Eq. 5.5). Using the simulated data described in Chap. 2, Sect. 2.8.1, in Fig. 5.1 we present the distribution of the GEBVs (Eq. 5.9) associated with traits T1 in the first (Fig. 5.1a) and the fifth (Fig. 5.1b) selection cycles in the testing population. In effect, the frequency distribution of the GEBVs approaches normal distribution in both selection cycles.

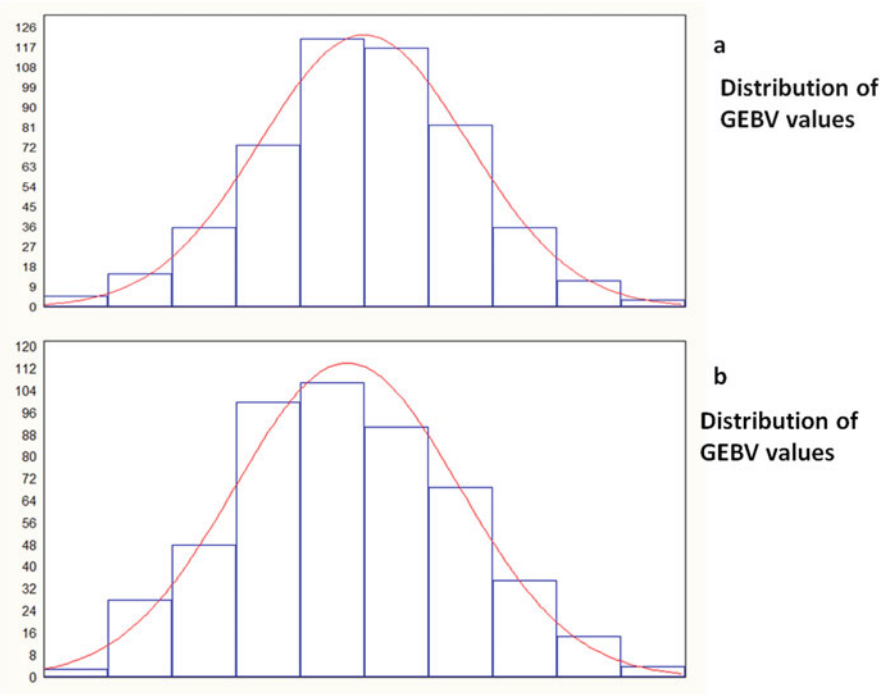

Fig. 5.1 Distribution of the genomic estimated breeding values (GEBVs) associated with traits T1 in (**a**) the first and (**b**) the fifth selection cycles in the testing population

5.1.3 The LGSI and Its Parameters

Similar to the LPSI (Chap. 2), the objective of the LGSI is to predict the net genetic merit $H = \mathbf{w}'\mathbf{g}$, where $\mathbf{g}' = [g_1 \quad g_2 \quad \ldots \quad g_t]$ (t = number of traits) is a vector of unobservable true breeding values and $\mathbf{w}' = [w_1 \quad w_2 \quad \ldots \quad w_t]$ is a vector of economic weights. Suppose that the genomic breeding values $\boldsymbol{\gamma}_i = \mathbf{X}\mathbf{u}_i$ are known; then, the LGSI can be written as

$$I_{\mathrm{G}} = \boldsymbol{\beta}'\boldsymbol{\gamma}, \tag{5.10}$$

where $\boldsymbol{\beta}$ is an unknown vector of weights.

The main advantage of the LGSI over the LPSI lies in the possibility of reducing the intervals between selection cycles (L_G) by more than two thirds (Lorenz et al. 2011); thus, this parameter should be incorporated into the LGSI selection response and the expected genetic gain per trait to reflect the main advantage of the LGSI over the LPSI and the other indices. Assuming that $L_G = 1$, in the LPSI context we

wrote the selection response as $R_I = k_I \sigma_H \rho_{HI}$; however, if $L_G \neq 1$, the LGSI selection response can be written as

$$R_{I_G} = \frac{k_I}{L_G} \frac{\sigma_{HI_G}}{\sigma^2_{I_G}} = \frac{k_I}{L_G} \sigma_H \rho_{HI_G}, \tag{5.11}$$

where k_I is the standardized selection differential (or selection intensity) associated with the LGSI, σ_{HI_G} is the covariance between $H = \mathbf{w}'\mathbf{g}$ and the LGSI, $\sigma^2_{I_G}$ is the variance of the LGSI, σ_H is the standard deviation of H, ρ_{HI_G} is the correlation between H and the LGSI, and L_G denotes the intervals between selection cycles.

Let $\mathbf{C}$ and $\boldsymbol{\Gamma}$ be matrices of covariance of the breeding values ($\mathbf{g}$) and of the genomic breeding values ($\boldsymbol{\gamma}$) respectively; then, the correlation between $H = \mathbf{w}'\mathbf{g}$ and $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$ can be written as

$$\rho_{HI_G} = \frac{\mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}\sqrt{\boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta}}}, \tag{5.12}$$

where $\mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \sigma_{HI_G}$ is the covariance between $H = \mathbf{w}'\mathbf{g}$ and $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$, $\sigma_H = \sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}$ is the standard deviation of the variance of $H = \mathbf{w}'\mathbf{g}$, and $\sigma_{I_G} = \sqrt{\boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta}}$ is the standard deviation of the variance of $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$.

5.1.4 Maximizing LGSI Parameters

To maximize the genomic selection response (Eq. 5.11), suppose that k_I, σ_H and L_G are fixed and take the derivative of the natural logarithm (ln) of the correlation between H and I_G (Eq. 5.12) with respect to vector $\boldsymbol{\beta}$, equate the result of the derivative to the null vector, and isolate $\boldsymbol{\beta}$, i.e.,

$$\frac{\partial}{\partial \boldsymbol{\beta}} \ln \rho_{HI_g} = \frac{\partial}{\partial \boldsymbol{\beta}} \ln \left(\frac{\mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}\sqrt{\boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta}}} \right) = \mathbf{0}. \tag{5.13}$$

The result is $\boldsymbol{\beta} = s\mathbf{w}$, where $s = \boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta}/\mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta}$ is a proportional constant that does not affect the maximum value of ρ_{HI_G}, because this is invariant to the scale change; then, assuming that $\boldsymbol{\beta} = \mathbf{w}$, the maximized LGSI selection response can be written as

$$R_{I_G} = \frac{k_I}{L_G} \sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w}}. \tag{5.14}$$

Hereafter, we refer to the LGSI genomic selection response as that of Eq. (5.14). Also, because $\boldsymbol{\beta} = \mathbf{w}$, Eq. (5.12) can be written as

$$\rho_{HI_G} = \frac{\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w}}}{\sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}} = \frac{\sigma_{I_G}}{\sigma_H}, \tag{5.15}$$

which is the maximized correlation between $H = \mathbf{w}'\mathbf{g}$ and $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$, or LGSI accuracy; $\sigma_H = \sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}$ is the standard deviation of the variance of H, and $\sigma_{I_G} = \sqrt{\boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta}}$ is the standard deviation of the variance of I_G.

The LGSI expected genetic gain per trait ($\mathbf{E}_{I_G}$) can be written as

$$\mathbf{E}_{I_G} = \frac{k_I}{L_G}\frac{\boldsymbol{\Gamma}\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w}}}. \tag{5.16}$$

All the terms in Eq. (5.16) were previously defined.

Let $\lambda_G = \frac{\rho_{HI_G}}{\rho_{HI}}$ be LGSI efficiency versus LPSI efficiency to predict the net genetic merit, where ρ_{HI_G} is the LGSI accuracy and ρ_{HI} the LPSI accuracy; in percentage terms, LGSI efficiency versus LPSI efficiency for each selection cycle can be written as

$$p_G = 100(\lambda_G - 1). \tag{5.17}$$

According to Eq. (5.17), if $p_G > 0$, LGSI efficiency is greater than LPSI efficiency; if $p_G = 0$, the efficiency of both selection indices is equal, and if $p_G < 0$, the LPSI is more efficient than the LGSI at predicting $H = \mathbf{w}'\mathbf{g}$.

Equation (5.17) is useful for measuring LGSI efficiency in terms of accuracy when predicting the net genetic merit ($H = \mathbf{w}'\mathbf{g}$), whereas the Technow et al. (2013) inequality measures LGSI efficiency in terms of the time needed to complete one selection cycle. In the context of the LGSI and the LPSI, the Technow inequality can be written as

$$L_G < \frac{\rho_{HI_G}}{h_I} L_P, \tag{5.18}$$

where L_G and L_P denote the time required to complete one selection cycle for the LGSI and the LPSI respectively, ρ_{HI_G} is the LGSI accuracy, and h_I is the square root of the heritability (Lin and Allaire 1977; Nordskog 1978) of the LPSI, which can be denoted as $h_I = \sqrt{\frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{b}'\mathbf{P}\mathbf{b}}}$ (see Chap. 2 for details). Then, assuming that the selection intensity is the same for both selection indices, if Eq. (5.18) is true, the LGSI is more efficient than the LPSI per unit of time.

5.1.5 *Relationship Between the LGSI and LPSI Selection Responses*

To obtain the relationship between R_{I_G} and R_I in the asymptotic context, we omitted the intervals between selection cycles (L_G and L_I respectively) to simplify the

algebra. Consider a population where the number of genotypes and markers tends to infinity; in this case, markers explain most of the true additive genetic variances and covariances. Thus, we can assume that matrices $\boldsymbol{\Gamma}$ and $\mathbf{C}$ are very similar, and at the limit, $\boldsymbol{\Gamma} = \mathbf{C}$. Now suppose that in this population the phenotypic variance–covariance matrix ($\mathbf{P}$) is known and comprises matrix $\boldsymbol{\Gamma}$ and the variance–covariance residual matrix ($\mathbf{R}$). In this case, the inverse of $\mathbf{P}$ can be written as $\mathbf{P}^{-1} = (\boldsymbol{\Gamma} + \mathbf{R})^{-1} = \boldsymbol{\Gamma}^{-1} - \boldsymbol{\Gamma}^{-1}(\boldsymbol{\Gamma}^{-1} + \mathbf{R}^{-1})^{-1}\boldsymbol{\Gamma}^{-1}$, where $\boldsymbol{\Gamma}^{-1}$ and $\mathbf{R}^{-1}$ are the inverses of matrices $\boldsymbol{\Gamma}$ and $\mathbf{R}$ respectively. Thus, the LPSI selection response is given by

$$R_I = k_I\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}} = k_I\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{P}^{-1}\boldsymbol{\Gamma}\mathbf{w}} = k_I\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w} - \mathbf{w}'\left(\boldsymbol{\Gamma}^{-1} + \mathbf{R}^{-1}\right)^{-1}\mathbf{w}}, \quad (5.19)$$

where $\mathbf{b} = \mathbf{P}^{-1}\boldsymbol{\Gamma}\mathbf{w}$ is the vector of coefficients of the LPSI in the asymptotic context. Note that $\mathbf{b}'\mathbf{P}\mathbf{b} \geq 0$ and $\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w} \geq 0$, i.e., $\mathbf{b}'\mathbf{P}\mathbf{b}$ and $\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w}$ are positive semi-definite, meaning that $\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w} \geq \mathbf{w}'(\boldsymbol{\Gamma}^{-1} + \mathbf{R}^{-1})^{-1}\mathbf{w} \geq 0$; then, in the asymptotic context, $R_{I_G} \geq R_I$. This result is not common when the number of genotypes and markers is small; however, it gives an idea of the theoretical behavior of R_{I_G} with respect to R_I when the number of markers and genotypes is very large.

Because $\mathbf{g}_q$ can be written as $\mathbf{g}_q = \boldsymbol{\gamma}_q + \eta_q$, where $\eta_q = \mathbf{g}_q - \boldsymbol{\gamma}_q\,(q = 1, 2, \cdots, t)$, for low numbers of markers and genotypes, the covariance genotypic matrix $\mathbf{C}$ can be written as $\mathbf{C} = \boldsymbol{\Gamma} + \mathbf{E}$, where $\mathbf{E} = \mathbf{C} - \boldsymbol{\Gamma}$; then, the inverse of matrix $\mathbf{P}$ can be written as $\mathbf{P}^{-1} = [(\boldsymbol{\Gamma} + \mathbf{E}) + \mathbf{R}]^{-1} = (\boldsymbol{\Gamma} + \mathbf{E})^{-1} - (\boldsymbol{\Gamma} + \mathbf{E})^{-1}[(\boldsymbol{\Gamma} + \mathbf{E})^{-1} + \mathbf{R}^{-1}]^{-1}$ $(\boldsymbol{\Gamma} + \mathbf{E})^{-1}$. In the latter case, the LPSI selection response R_I can be written as

$$\begin{aligned} R_I &= k_I\sqrt{\mathbf{w}'(\boldsymbol{\Gamma} + \mathbf{E})\mathbf{P}^{-1}(\boldsymbol{\Gamma} + \mathbf{E})\mathbf{w}} \\ &= k_I\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w} + \mathbf{w}'\mathbf{E}\mathbf{w} - \mathbf{w}'\left[(\boldsymbol{\Gamma} + \mathbf{E})^{-1} + \mathbf{R}^{-1}\right]^{-1}\mathbf{w}}. \end{aligned} \quad (5.20)$$

Equation (5.20) indicates that in the non-asymptotic context (low numbers of markers and genotypes), R_{I_G} and R_I are related in three possible ways:

1. $R_I > R_{I_G}$ if $\mathbf{w}'\mathbf{E}\mathbf{w} > \mathbf{w}'[(\boldsymbol{\Gamma} + \mathbf{E})^{-1} + \mathbf{R}^{-1}]^{-1}\mathbf{w}$
2. $R_I = R_{I_G}$ if $\mathbf{w}'\mathbf{E}\mathbf{w} = \mathbf{w}'[(\boldsymbol{\Gamma} + \mathbf{E})^{-1} + \mathbf{R}^{-1}]^{-1}\mathbf{w}$
3. $R_{I_G} > R_I$ if $\mathbf{w}'\mathbf{E}\mathbf{w} < \mathbf{w}'[(\boldsymbol{\Gamma} + \mathbf{E})^{-1} + \mathbf{R}^{-1}]^{-1}\mathbf{w}$

The second and third points indicate that R_{I_G} may be equal to or larger than R_I, even under a small number of markers, depending on the size of $\mathbf{w}'\mathbf{E}\mathbf{w}$ and $\mathbf{w}'$ $[(\boldsymbol{\Gamma} + \mathbf{E})^{-1} + \mathbf{R}^{-1}]^{-1}\mathbf{w}$. These three points explain the theoretical relationship between R_I and R_{I_G} for a low number of markers and genotypes. When $\boldsymbol{\Gamma} = \mathbf{C}$, $\mathbf{E} = \mathbf{0}$, and $R_I = k_I\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w} - \mathbf{w}'\left(\boldsymbol{\Gamma}^{-1} + \mathbf{R}^{-1}\right)^{-1}\mathbf{w}}$, then $R_{I_G} \geq R_I$.

5.1.6 Statistical LGSI Properties

Assuming that H and I_G have joint bivariate normal distribution and that $\mathbf{\Gamma}$, $\mathbf{C}$, and $\mathbf{w}$ are known, the LGSI has the following properties:

1. The variance of I_G ($\sigma^2_{I_G}$) and the covariance between H and I_G (σ_{HI_G}) are equal, i.e., $\sigma^2_{I_G} = \sigma_{HI_G}$.
2. The maximized correlation between H and I_G (or LGSI accuracy) is equal to $\rho_{HI_G} = \frac{\sigma_{I_G}}{\sigma_H}$, where σ_{I_G} is the standard deviation of $\sigma^2_{I_G}$ and σ_H is the standard deviation of the variance of H (σ^2_H).
3. The variance of the predicted error, $Var(H - I_G) = \left(1 - \rho^2_{HI_G}\right)\sigma^2_H$, is minimal. Note that $Var(H - I_G) = \sigma^2_{I_G} + \sigma^2_H - 2\sigma_{HI_G}$, and when $\mathbf{\beta} = \mathbf{w}$, $\sigma^2_{I_G} = \sigma_{HI_G}$, whence $Var(H - I_G) = \sigma^2_H - \sigma^2_{I_G} = \left(1 - \rho^2_{HI_G}\right)\sigma^2_H$ is minimal.
4. The total variance of H explained by I_G is $\sigma^2_{I_G} = \rho^2_{HI_G}\sigma^2_H$. It is evident that if $\rho_{HI_G} = 1, \sigma^2_{I_G} = \sigma^2_H$, and if $\rho_{HI_G} = 0, \sigma^2_{I_G} = 0$. That is, the variance of H explained by I_G is proportional to ρ_{HI_G}, and when ρ_{HI_G} is close to 1, $\sigma^2_{I_G}$ is close to σ^2_H; if ρ_{HI_G} is close to 0, $\sigma^2_{I_G}$ is close to 0.

The LGSI properties described in points 1–4 of this subsection are the same as the LPSI properties described in Chap. 2. This corroborates the LGSI as an application of the LPSI theory to the genomic selection context.

5.1.7 Genomic Covariance Matrix in the Training and Testing Population

To derive the LGSI theory, we assumed that the true genomic additive variance–covariance matrix $\mathbf{\Gamma}$ was known. However, in practice, we need to estimate it. In the training population, matrix $\mathbf{\Gamma}$ can be estimated by restricted maximum likelihood (REML) using phenotypic and genomic information, as described by Vattikuti et al. (2012) and Su et al. (2012). In Eqs. (2.22) to (2.24) of Chap. 2, we presented the formulas for estimating the genotypic and residual variance and covariance based on the formulas described by Lynch and Walsh (1998). Here, we present a brief description of how we can estimate the qth component ($\sigma_{\gamma qq}$) of $\mathbf{\Gamma}$ in the training population using the REML method.

We estimated $\sigma_{\gamma qq} = \sigma^2_{\gamma q}$ (q, $q' = t$ = number of traits) in the absence of dominance and epistatic effects, using the model $\mathbf{y}_q = \mathbf{1}\mu_q + \mathbf{Z}\boldsymbol{\gamma}_q + \boldsymbol{\varepsilon}_q$, where the vector $\mathbf{y}_q \sim \text{NMV}(\mathbf{1}\mu_q, \mathbf{V}_q)$ $g \times 1$ (g = number of genotypes in the population) had a multivariate normal distribution; $\mathbf{1}$ was a $g \times 1$ vector of 1s, μ_q was the mean of the qth trait, $\mathbf{Z}$ was an identity matrix $g \times g$; $\boldsymbol{\gamma}_q \sim \text{NMV}(\mathbf{0}, \mathbf{G}\sigma^2_{\gamma q})$ was a vector of genomic

breeding values, and $\boldsymbol{\varepsilon}_q \sim \text{NMV}(\mathbf{0}, \mathbf{I}\sigma^2_{\varepsilon_q})$ was a $g \times 1$ vector of residuals. Matrix $\mathbf{G} = \mathbf{XX}'/c$ was the genomic relationship matrix, and in an F_2 population, $c = \sum_{j=1}^{N} 2p_j q_j$; $\mathbf{X}$ was a $g \times m$ matrix (m = number of markers) of the coded marker values ($2 - 2p$ for AA, $1 - 2p$ for Aa, and $-2p$ for aa) for the additive effects of the markers; p and q denote the frequency of allele A and the frequency of allele a in the jth marker ($j = 1, 2, \ldots, m$), and $\mathbf{V}_q = \mathbf{G}\sigma^2_{\gamma q} + \mathbf{I}\sigma^2_{\varepsilon_q}$.

The expectation–maximization algorithm allowed the REML for the variance components $\sigma^2_{\gamma q}$ and $\sigma^2_{\varepsilon_q}$ to be computed by iterating the following equations:

$$\sigma^{2(n+1)}_{\gamma q} = \sigma^{2(n)}_{\gamma q} + \frac{\left(\sigma^{2(n)}_{\gamma q}\right)^2}{g}\left[\mathbf{y}'_q\left(\mathbf{T}^{(n)}\mathbf{G}\mathbf{T}^{(n)}\right)\mathbf{y}_q - tr\left(\mathbf{T}^{(n)}\mathbf{G}\right)\right] \tag{5.21}$$

and

$$\sigma^{2(n+1)}_{\varepsilon_q} = \sigma^{2(n)}_{\varepsilon_q} + \frac{\left(\sigma^{2(n)}_{\varepsilon_q}\right)^2}{g}\left[\mathbf{y}'_q\left(\mathbf{T}^{(n)}\mathbf{T}^{(n)}\right)\mathbf{y}_q - tr\left(\mathbf{T}^{(n)}\right)\right], \tag{5.22}$$

where g is the number of genotypes. After n iterations, when $\sigma^{2(n+1)}_{\gamma q}$ was very similar to $\sigma^{2(n)}_{\gamma q}$ and $\sigma^{2(n+1)}_{\varepsilon_q}$ was very similar to $\sigma^{2(n)}_{\varepsilon_q}$, $\sigma^{2(n+1)}_{\gamma q}$ and $\sigma^{2(n+1)}_{\varepsilon_q}$ were the estimated variance components of $\sigma^2_{\gamma q}$ and $\sigma^2_{\varepsilon_q}$ respectively. In Eqs. (5.21) and (5.22) $tr(.)$ denoted the trace of the matrices within brackets; $\mathbf{T} = \mathbf{V}^{-1}_q - \mathbf{V}^{-1}_q\mathbf{1}\left(\mathbf{1}'\mathbf{V}^{-1}_q\mathbf{1}\right)^{-1}\mathbf{1}'\mathbf{V}^{-1}_q$, and $\mathbf{V}^{-1}_q$ was the inverse of $\mathbf{V}_q = \mathbf{G}\sigma^2_{\gamma q} + \mathbf{I}\sigma^2_{\varepsilon_q}$. In matrix $\mathbf{T}^{(n)}$, $\mathbf{V}^{-1(n)}_q$ was the inverse of matrix $\mathbf{V}^{(n)}_q = \mathbf{G}\sigma^{2(n)}_{\gamma q} + \mathbf{I}\sigma^{2(n)}_{\varepsilon_q}$.

The genomic additive genetic covariance between the observations of the qth and ith traits, $\mathbf{y}_q$ and $\mathbf{y}_i$ ($\sigma_{\gamma_{qi}}$, q, $i = 1, 2, \ldots, t$), can be estimated by REML. Here, we adapted Eqs. (5.21) and (5.22) using the variance of the sum of $\mathbf{y}_q$ and $\mathbf{y}_i$, i.e., $Var(\mathbf{y}_i + \mathbf{y}_q) = \mathbf{V}_i + \mathbf{V}_q + 2\mathbf{C}_{iq}$, where $\mathbf{V}_i = \mathbf{G}\sigma^2_{\boldsymbol{\gamma}_i} + \mathbf{I}\sigma^2_{\varepsilon_i} = Var(\mathbf{y}_i)$ is the variance of $\mathbf{y}_i$ and $\mathbf{V}_q = \mathbf{G}\sigma^2_{\boldsymbol{\gamma}_q} + \mathbf{I}\sigma^2_{\varepsilon_q} = Var(\mathbf{y}_q)$ is the variance of $\mathbf{y}_q$; $2\mathbf{C}_{iq} = 2\mathbf{G}\sigma_{\boldsymbol{\gamma}_{iq}} + 2\mathbf{I}\sigma_{\varepsilon_{iq}} = 2Cov(\mathbf{y}_i, \mathbf{y}_q)$ is the covariance of $\mathbf{y}_q$ and $\mathbf{y}_i$, and $\sigma_{\boldsymbol{\gamma}_{iq}}$ and $\sigma_{\varepsilon_{iq}}$ are the genomic and residual covariance respectively, associated with $\mathbf{y}_i$ and $\mathbf{y}_q$. Thus, one way of estimating $\sigma_{\boldsymbol{\gamma}_{iq}}$ and $\sigma_{\varepsilon_{iq}}$ is by using the following equation:

$$0.5Var(\mathbf{y}_i + \mathbf{y}_q) - 0.5Var(\mathbf{y}_i) - 0.5Var(\mathbf{y}_q), \tag{5.23}$$

for which Eqs. (5.21) and (5.22) can be adapted.

If there is only marker information on the testing population, then it is not possible to estimate $\boldsymbol{\Gamma}$ using Eqs. (5.21) to (5.23). Another way of estimating $\boldsymbol{\Gamma}$ is to use the method proposed by Ceron-Rojas et al. (2015), which requires the estimated values of $\boldsymbol{\gamma}_q$ ($\widehat{\boldsymbol{\gamma}}_q$) in the cycle of interest. Let $\widehat{\mathbf{u}}$ be the estimator of the

vector of marker effects $\mathbf{u}' = [\mathbf{u}_1' \quad \mathbf{u}_2' \quad \cdots \quad \mathbf{u}_t']$ for t traits obtained in the training population. We obtained the qth GEBVs ($q = 1, 2, \ldots, t$) in the lth selection cycle ($l = 1, 2, \ldots$, number of cycles) as

$$\widehat{\boldsymbol{\gamma}}_{ql} = \mathbf{X}_l \widehat{\mathbf{u}}_q \tag{5.24}$$

where $\widehat{\mathbf{u}}_q$ is the vector of size $m \times 1$ of the estimated marker effects of the qth trait in the training population and $\mathbf{X}_l$ is a matrix of size $n \times m$ of the coded values of marker genotypes in the lth selection cycle of the testing population.

Now suppose that $\boldsymbol{\gamma}_q$ and $\boldsymbol{\gamma}_{q'}$ have multivariate normal distribution jointly, with mean $\mathbf{1}\mu_{\gamma_q}$ and $\mathbf{1}\mu_{\gamma_{q'}}$ respectively, and covariance matrix $\mathbf{G}\sigma_{\gamma_{qq'}}$, where $\mathbf{1}$ is an $n \times 1$ vector of 1s and $\mathbf{G} = \mathbf{X}\mathbf{X}'/c$ is the additive genomic relationship matrix. Then, $\boldsymbol{\Gamma} = \left\{\sigma_{\gamma_{qq'}}\right\}$ can be estimated as

$$\widehat{\boldsymbol{\Gamma}}_l = \left\{\widehat{\sigma}_{\gamma_{qq'}}\right\}, \tag{5.25}$$

where $\widehat{\sigma}_{\gamma_{qq'}} = \frac{1}{g}\left(\widehat{\boldsymbol{\gamma}}_{ql} - \mathbf{1}\widehat{\mu}_{\gamma_{ql}}\right)' \mathbf{G}_l^{-1} \left(\widehat{\boldsymbol{\gamma}}_{q'l} - \mathbf{1}\widehat{\mu}_{\gamma_{q'l}}\right)$ is the estimated covariance between $\boldsymbol{\gamma}_q$ and $\boldsymbol{\gamma}_{q'}$ in the lth selection cycle of the testing population; g is the number of genotypes; $\widehat{\boldsymbol{\gamma}}_{ql}$ was defined in Eq. (5.24); $\widehat{\mu}_{\gamma_{ql}}$ and $\widehat{\mu}_{\gamma_{q'l}}$ are the estimated arithmetic means of the values of $\widehat{\boldsymbol{\gamma}}_{ql}$ and $\widehat{\boldsymbol{\gamma}}_{q'l}$; $\mathbf{1}$ is a $g \times 1$ vector of 1s and $\mathbf{G}_l = c^{-1}\mathbf{X}_l\mathbf{X}_l'$ is the additive genomic relationship matrix in the lth selection cycle ($l = 1, 2, \ldots$, number of cycles) in the testing population.

From Eq. (5.25) we can estimate the LGSI response and expected genetic gain per trait in the testing population as

$$\widehat{R}_{I_G} = \frac{k_I}{L_G}\sqrt{\mathbf{w}'\widehat{\boldsymbol{\Gamma}}\mathbf{w}} \quad \text{and} \quad \widehat{\mathbf{E}}_{I_G} = \frac{k_I}{L_G}\frac{\widehat{\boldsymbol{\Gamma}}\mathbf{w}}{\sqrt{\mathbf{w}'\widehat{\boldsymbol{\Gamma}}\mathbf{w}}}, \tag{5.26}$$

respectively. The estimated LGSI ($\widehat{I}_{\mathrm{G}}$) values in the lth selection cycle can be obtained as

$$\widehat{I}_{\mathrm{G}} = \sum_{q=1}^{\mathrm{t}} w_q \widehat{\boldsymbol{\gamma}}_{ql}, \tag{5.27}$$

where w_q is the qth economic weight and $\widehat{\boldsymbol{\gamma}}_{ql}$ was defined in Eq. (5.24). Equation (5.27) is a vector of size $g \times 1$ (g= number of genotypes). In practice, $\widehat{I}_{\mathrm{G}}$ values are ranked to select individual genotypes with optimal GEBVs.

5.1.8 Numerical Examples

To estimate matrices **C** and **R** and the marker effects in the training population, we used a real maize (*Zea mays*) F_2 population with 248 genotypes (each with two repetitions), 233 molecular markers, and three traits—grain yield (GY, ton ha^{-1}), ear height (EHT, cm), and plant height (PHT, cm)—evaluated in one environment. The estimated matrices were $\widehat{\mathbf{C}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$ and $\widehat{\mathbf{R}} = \begin{bmatrix} 0.38 & 0.72 & 1.27 \\ 0.72 & 47.14 & 60.96 \\ 1.27 & 60.96 & 121.46 \end{bmatrix}$, which were estimated by Eqs. (5.21) to (5.23) using the numerical relationship **A** instead of the genomic relationship matrix ($\mathbf{G} = \mathbf{XX}'/c$).

Table 5.1 presents the first 20 BLUPs of the estimated marker effects (Eq. 5.8) in the training population and the first 20 marker coded values and GEBVs (Eq. 5.9) obtained in the testing population associated with trait GY. In the

Table 5.1 The 20 best linear unbiased predictors (BLUPs) of the estimated marker effects in the training population and the first 20 marker coded values and genomic estimated breeding values (GEBVs) obtained in the testing population associated with grain yield

Training population	Testing population					
	Marker coded values					
BLUPs	M1	M2	M3	...	M233	GEBVs
−0.0003	1	1	0	...	−1	0.195
−0.0038	0	0	0	...	−1	0.221
−0.0085	−1	1	0	...	−1	−0.643
0.0069	0	1	0	...	1	0.525
−0.0042	0	0	0	...	0	−0.603
0.0038	−1	0	0	...	0	0.062
0.0008	0	1	1	...	0	−0.226
0.0012	0	1	1	...	1	0.023
−0.0004	0	−1	0	...	0	0.444
0.0062	0	0	1	...	−1	−0.286
0.0121	−1	1	0	...	1	−0.196
0.0077	−1	−1	−1	...	0	−0.566
0.0033	−1	0	0	...	0	0.073
0.0102	−1	1	0	...	1	0.058
0.0054	0	1	0	...	0	0.874
0.0002	0	0	0	...	0	0.102
0.0171	0	1	0	...	−1	−0.342
0.0159	−1	0	1	...	−1	−0.428
0.0117	−1	0	0	...	−1	0.072
0.0121	0	−1	0	...	−1	−0.428

testing population, there were 380 genotypes and 233 molecular markers. In this population, the estimated genomic covariance matrix $\mathbf{\Gamma} = \left\{\sigma_{\gamma_{qq'}}\right\}$ was $\widehat{\mathbf{\Gamma}} = \begin{bmatrix} 0.21 & 2.95 & 5.00 \\ 2.95 & 42.41 & 71.11 \\ 5.00 & 71.11 & 121.53 \end{bmatrix}$. The first GEBV (0.195) related to GY in Table 5.1 was obtained as $0.195 = \quad - 0.0003(1) - 0.0038(1) - 0.0085(0) + \cdots - 0.03(-1)$. The other GEBVs can be obtained in a similar manner.

Suppose a selection intensity of 10% ($k_I = 1.755$) and a vector of economic weights of $\mathbf{w}' = [5 \quad -0.1 \quad -0.1]$; then, the estimated LGSI selection response and the expected genetic gain per trait without including the interval between selection cycle is $\widehat{R}_{I_G} = (1.755)\sqrt{\mathbf{w}'\widehat{\mathbf{\Gamma}}\mathbf{w}} = 0.92$ and $\widehat{\mathbf{E}}'_{I_G} = (1.755)\dfrac{\mathbf{w}'\widehat{\mathbf{\Gamma}}}{\sqrt{\mathbf{w}'\widehat{\mathbf{\Gamma}}\mathbf{w}}} = [0.80 \quad 11.41 \quad 19.28]$ respectively, whereas the estimated LGSI accuracy was $\widehat{\rho}_{HI_G} = 0.48$.

Chapter 11 presents RIndSel, a graphical unit interface that uses selection index theory to select individual candidates as parents for the next selection cycle, which can be used to obtain the results of the real numerical example described in this subsection.

To compare LGSI efficiency versus LPSI efficiency we used the simulated data described in Chap. 2, Sect. 2.8.1. According to Beyene et al. (2015), at least 4 years are required to complete one phenotypic selection cycle in maize, whereas genomic selection requires only 1.5 years. Thus, to compare LGSI efficiency versus LPSI efficiency in terms of time, we can use the Technow et al. (2013) inequality described in Eq. (5.18).

Table 5.2 presents the estimated value of Eq. (5.18) for five simulated selection cycles. The LGSI efficiency was higher than LPSI efficiency in terms of time, because the Technow et al. (2013) inequality was true in the five selection cycles. An additional result obtained by Ceron-Rojas et al. (2015) is presented in Fig. 5.2, which shows the correlation among the LGSI, the LPSI, and the true net genetic

Table 5.2 Five simulated selection cycles

Cycle	L_G	L_P	$\widehat{\rho}_{HI_G}$	$\widehat{h}_I$	$\frac{\widehat{\rho}_{HI_G}}{\widehat{h}_I} L_P$
1	1.5	4.0	0.73	0.92	3.17
2	1.5	4.0	0.78	0.89	3.50
3	1.5	4.0	0.83	0.88	3.77
4	1.5	4.0	0.74	0.87	3.40
5	1.5	4.0	0.71	0.87	3.30

Time required for the linear genomic selection index (L_G) and linear phenotypic selection index (L_P) to complete one selection cycle; estimated accuracy ($\widehat{\rho}_{HI_G}$) of the linear genomic selection index and the square root of the estimated heritability of the linear phenotypic selection index ($\widehat{h}_I$); estimated right-hand side ($\frac{\widehat{\rho}_{HI_G}}{\widehat{h}_I} L_P$) of the inequality formula ($L_G < \frac{\rho_{H,I_G}}{h_I} L_P$)

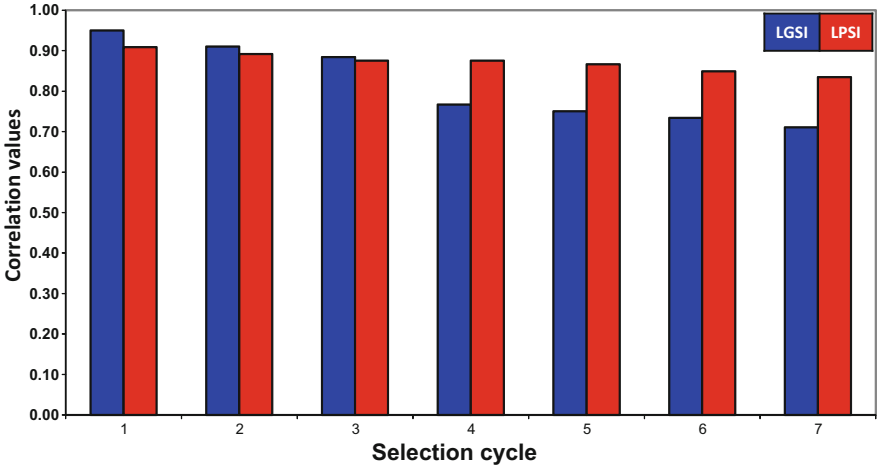

Fig. 5.2 Correlation between the linear genomic selection index (LGSI), the linear phenotypic selection index (LPSI), and true net genetic merit (H) values in seven selection cycles. For each selection cycle, the first column indicates the correlation between the LGSI estimated values and the *H* true values, whereas the second column shows the correlation between the LPSI estimated values and the *H* true values

merit values in seven selection cycles. According to Fig. 5.2, the correlation between the LGSI and the true net genetic merit values was higher than the correlation between the LPSI and the true net genetic merit values for the first three selection cycles; after this cycle, the correlation between LGSI and the true net genetic merit values tended to decrease.

5.2 The Combined Linear Genomic Selection Index

The combined LGSI (CLGSI) developed by Dekkers (2007) is a slightly modified version of the LMSI (see Chap. 4 for details), which, instead of using the marker scores, uses the GEBVs and the phenotypic information jointly to predict the net genetic merit. The main difference between the CLGSI and the LGSI is that the CLGSI can only be used in training populations, whereas the LGSI is used in testing populations. The basic conditions for constructing a valid CLGSI include conditions for constructing the LPSI, the LMSI, and the LGSI, because the CLGSI uses GEBVs and phenotypic information jointly to predict the net genetic merit.

5.2.1 The CLGSI Parameters

The net genetic merit can be written in a similar manner to that in the LMSI context, that is, as

$$H = \mathbf{w}'\mathbf{g} + \mathbf{w}_2'\boldsymbol{\gamma} = [\,\mathbf{w}' \quad \mathbf{w}_2'\,]\begin{bmatrix}\mathbf{g}\\ \boldsymbol{\gamma}\end{bmatrix} = \mathbf{a}_G'\mathbf{z}_G, \tag{5.28}$$

where $\mathbf{g}' = [\,g_1 \quad \dots \quad g_t\,]$ is the vector of breeding values, $\mathbf{w}' = [\,w_1 \quad \cdots \quad w_t\,]$ is the vector of economic weights associated with breeding values, $\mathbf{w}_2' = [\,0_1 \quad \cdots \quad 0_t\,]$ is a null vector associated with the vector of genomic breeding values $\boldsymbol{\gamma}' = [\,\gamma_1 \quad \gamma_2 \quad \dots \quad \gamma_t\,]$, $\mathbf{a}_G' = [\,\mathbf{w}' \quad \mathbf{w}_2'\,]$ and $\mathbf{z}_G = [\,\mathbf{g}' \quad \boldsymbol{\gamma}'\,]$.

The CLGSI can be written as

$$I_C = \boldsymbol{\beta}_{\mathbf{y}}'\mathbf{y} + \boldsymbol{\beta}_G'\boldsymbol{\gamma} = \begin{bmatrix}\boldsymbol{\beta}_{\mathbf{y}}' & \boldsymbol{\beta}_G'\end{bmatrix}\begin{bmatrix}\mathbf{y}\\ \boldsymbol{\gamma}\end{bmatrix} = \boldsymbol{\beta}_C'\mathbf{t}_C, \tag{5.29}$$

where $\mathbf{y}' = [\,y_1 \quad \cdots \quad y_t\,]$ (t = number of traits) is the vector of phenotypic values; $\boldsymbol{\gamma}$ was defined earlier; $\boldsymbol{\beta}_{\mathbf{y}}'$ and $\boldsymbol{\beta}_G$ are vectors of coefficients of phenotypic and genomic weight values respectively; $\boldsymbol{\beta}_C' = \begin{bmatrix}\boldsymbol{\beta}_{\mathbf{y}}' & \boldsymbol{\beta}_G'\end{bmatrix}$ and $\mathbf{t}_G' = [\,\mathbf{y}' \quad \boldsymbol{\gamma}'\,]$.

The CLGSI selection response can be written as

$$R_C = k_I\sigma_H\rho_{HI_C} = k_I\sigma_H\frac{\mathbf{a}_C'\boldsymbol{\Psi}_C\boldsymbol{\beta}_C}{\sqrt{\mathbf{a}_C'\boldsymbol{\Psi}_C\mathbf{a}_C}\sqrt{\boldsymbol{\beta}_C'\mathbf{T}_C\boldsymbol{\beta}_C}}, \tag{5.30}$$

where k_I is the standardized selection differential of the CLGSI, $\sigma_H^2 = \mathbf{a}_C'\boldsymbol{\Psi}_C\mathbf{a}_C$ and $Var(I_C) = \boldsymbol{\beta}_C'\mathbf{T}_C\boldsymbol{\beta}_C$ are the variances of H and I_C, whereas $\mathbf{a}_C'\boldsymbol{\Psi}_C\boldsymbol{\beta}_C$ and ρ_{HI_C} are the covariance and the correlation between H and I_C respectively; $\mathbf{T}_C = Var\begin{bmatrix}\mathbf{y}\\ \boldsymbol{\gamma}\end{bmatrix} = \begin{bmatrix}\mathbf{P} & \boldsymbol{\Gamma}\\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\end{bmatrix}$ and $\boldsymbol{\Psi}_C = Var\begin{bmatrix}\mathbf{g}\\ \boldsymbol{\gamma}\end{bmatrix} = \begin{bmatrix}\mathbf{C} & \boldsymbol{\Gamma}\\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\end{bmatrix}$ are block matrices of the phenotypic covariance matrix, $\mathbf{P} = Var(\mathbf{y})$, the genomic covariance matrix, $\boldsymbol{\Gamma} = Var(\boldsymbol{\gamma})$, and the genetic breeding values covariance matrix, $\mathbf{C} = Var(\mathbf{g})$.

Suppose that matrices $\boldsymbol{\Psi}_C$ and $\mathbf{T}_C$ are known; then the CLGSI vector of coefficients that simultaneously maximizes ρ_{HI_C} and R_C can be written as

$$\boldsymbol{\beta}_C = \mathbf{T}_C^{-1}\boldsymbol{\Psi}_C\mathbf{a}_C, \tag{5.31}$$

whence the optimized CLGSI is

$$I_C = \boldsymbol{\beta}_C'\mathbf{t}_C, \tag{5.32}$$

Equations (5.31) and (5.32) indicate that the CLGSI is an application of the LPSI to the genomic selection context.

From Eq. (5.31), the maximized CLGSI selection response, expected genetic gain per trait and accuracy can be written as

$$R_C = k_I \sqrt{\boldsymbol{\beta}_C' \mathbf{T}_C \boldsymbol{\beta}_C}, \tag{5.33}$$

$$\mathbf{E}_C = k_I \frac{\boldsymbol{\Psi}_C \boldsymbol{\beta}_C}{\sqrt{\boldsymbol{\beta}_C' \mathbf{T}_C \boldsymbol{\beta}_C}} \tag{5.34}$$

and

$$\rho_{HI_C} = \frac{\sqrt{\boldsymbol{\beta}_C' \mathbf{T}_C \boldsymbol{\beta}_C}}{\sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}}, \tag{5.35}$$

respectively. Note that the maximized LPSI accuracy is $\rho_{HI} = \frac{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}}{\sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}}$ (see Chap. 2). The denominator of the accuracy of the CLGSI and $\rho_{HI} = \frac{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}}{\sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}}$ is the same; however, the numerator of the two indices accuracy is different. We would expect that $\sqrt{\boldsymbol{\beta}_C' \mathbf{T}_C \boldsymbol{\beta}_C} \geq \sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}$, and then $\rho_{HI_C} \geq \rho_{HI}$. Similar results can be observed when we compared the maximized LPSI selection response and expected genetic gain per trait with the maximized CLGSI selection response and expected genetic gain per trait.

5.2.2 *Relationship Between the CLGSI and the LGSI*

As we have indicated, the CLGSI is mathematically equivalent to the LMSI; thus, it has similar statistical properties to those of the LMSI, some of which are described in this section. The rest can be seen in Chap. 4. Let $\mathbf{Q}_C = \mathbf{T}_C^{-1} \boldsymbol{\Psi}_C$, then matrix $\mathbf{Q}_C$ can be written as

$$\mathbf{Q}_C = \begin{bmatrix} (\mathbf{P} - \boldsymbol{\Gamma})^{-1} (\mathbf{C} - \boldsymbol{\Gamma}) & \mathbf{0} \\ \mathbf{I} - (\mathbf{P} - \boldsymbol{\Gamma})^{-1} (\mathbf{C} - \boldsymbol{\Gamma}) & \mathbf{I} \end{bmatrix}, \tag{5.36}$$

whence as $\mathbf{w}_2' = [0_1 \quad \cdots \quad 0_t]$, the two sub-vectors that conform vector $\boldsymbol{\beta}_C = \mathbf{Q}_C \mathbf{a}_C$ or $\boldsymbol{\beta}_C' = \begin{bmatrix} \boldsymbol{\beta}_{\mathbf{y}}' & \boldsymbol{\beta}_G' \end{bmatrix}$ can be written as

$$\boldsymbol{\beta}_{\mathbf{y}} = (\mathbf{P} - \boldsymbol{\Gamma})^{-1} (\mathbf{C} - \boldsymbol{\Gamma}) \mathbf{w}, \tag{5.37}$$

and

$$\boldsymbol{\beta}_G = \left[\mathbf{I} - (\mathbf{P} - \boldsymbol{\Gamma})^{-1} (\mathbf{C} - \boldsymbol{\Gamma}) \right] \mathbf{w} = \mathbf{w} - \boldsymbol{\beta}_{\mathbf{y}}. \tag{5.38}$$

When $\boldsymbol{\Gamma}$ is equal to the null matrix (no genomic information), Eq. (5.37) is equal to $\boldsymbol{\beta}_{\mathbf{y}} = \mathbf{P}^{-1}\mathbf{C}\mathbf{w} = \mathbf{b}$ and $R_C = k_I\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}} = R_I$, which are the LPSI vector of coefficients and the selection response.

By Eqs. (5.37) and (5.38), the maximized CLGSI selection response and the optimized CLGSI can be written as

$$R_C = k_I\sqrt{\mathbf{w}'\mathbf{C}(\mathbf{P}-\boldsymbol{\Gamma})^{-1}(\mathbf{C}-\boldsymbol{\Gamma})\mathbf{w} + \mathbf{w}'\boldsymbol{\Gamma}\left[\mathbf{I}-(\mathbf{P}-\boldsymbol{\Gamma})^{-1}(\mathbf{C}-\boldsymbol{\Gamma})\right]\mathbf{w}} \tag{5.39}$$

and

$$I_C = \boldsymbol{\beta}_{\mathbf{y}}\mathbf{y} + \boldsymbol{\beta}_G\boldsymbol{\gamma} = \mathbf{w}'\boldsymbol{\gamma} + \boldsymbol{\beta}_{\mathbf{y}}(\mathbf{y}-\boldsymbol{\gamma}), \tag{5.40}$$

respectively.

Assume that when the number of markers and genotypes increases, matrix $\boldsymbol{\Gamma}$ tends to matrix $\mathbf{C}$ and that, at the limit, $\boldsymbol{\Gamma} = \mathbf{C}$; then, Eq. (5.39) can be written as $R_C = k_I\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w}} = R_G$ (except by L_G); in addition, $\boldsymbol{\beta}_{\mathbf{y}} = \mathbf{0}$ and $\boldsymbol{\beta}_G = \mathbf{w}$, the weights of the LGSI, and, in this latter case, the CLGSI is equal to the LGSI, as we would expect. Thus, in the asymptotic context, the LGSI and the CLGSI are the same.

An additional interesting result of the relationship between the CLGSI and the LGSI is as follows. The maximized correlation between H and I_C (or CLGSI accuracy) can be written as

$$\rho_{HI_C} = \frac{\mathbf{a}_C'\boldsymbol{\Psi}_C\boldsymbol{\beta}_C}{\sqrt{\mathbf{a}_C'\boldsymbol{\Psi}_C\mathbf{a}_C}\sqrt{\boldsymbol{\beta}_C'\mathbf{T}_C\boldsymbol{\beta}_C}}; \tag{5.41}$$

However, when $\boldsymbol{\Gamma} = \mathbf{C}$, $\boldsymbol{\Psi}_C = \begin{bmatrix}\boldsymbol{\Gamma} & \boldsymbol{\Gamma}\\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\end{bmatrix}$, $\boldsymbol{\beta}_{\mathbf{y}} = \mathbf{0}$, $\boldsymbol{\beta}_G = \mathbf{w}$, and $\boldsymbol{\beta}_C' = \begin{bmatrix}\boldsymbol{\beta}_{\mathbf{y}}' & \boldsymbol{\beta}_G'\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{w}'\end{bmatrix}$, whence $\mathbf{a}_C'\boldsymbol{\Psi}_C\boldsymbol{\beta}_C = \mathbf{a}_C'\boldsymbol{\Psi}_C\mathbf{a}_C = \boldsymbol{\beta}_C'\mathbf{T}_C\boldsymbol{\beta}_C = \mathbf{w}'\boldsymbol{\Gamma}\mathbf{w}$, and Eq. (5.41) is equal to 1. That is, the maximum correlation between H and I_C in the asymptotic context is equal to the maximum correlation between H and the LGSI, and that value will be equal to 1.

The asymptotic relationship between the CLGSI expected genetic gain per trait, $\mathbf{E}_C$ (Eq. 5.34), and the LGSI expected genetic gain per trait, $\mathbf{E}_{I_G}$ (Eq. 5.16), is as follows. When $\boldsymbol{\Gamma} = \mathbf{C}$, $\boldsymbol{\Psi}_C = \begin{bmatrix}\boldsymbol{\Gamma} & \boldsymbol{\Gamma}\\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\end{bmatrix}$ and $\boldsymbol{\beta}_C' = \begin{bmatrix}\mathbf{0} & \mathbf{w}'\end{bmatrix}$, whence

$$\mathbf{E}_C = k_I\frac{\boldsymbol{\Psi}_C\boldsymbol{\beta}_C}{\sqrt{\boldsymbol{\beta}_C'\mathbf{T}_C\boldsymbol{\beta}_C}} = k_I\frac{2\boldsymbol{\Gamma}\mathbf{w}}{\sqrt{\mathbf{w}'\boldsymbol{\Gamma}\mathbf{w}}} = 2\mathbf{E}_{I_G}. \tag{5.42}$$

This means that in the asymptotic context, the CLGSI expected genetic gain per trait is twice the LGSI expected genetic gain per trait. Of course, 2 is only a proportionality constant; thus, in reality, $\mathbf{E}_C = \mathbf{E}_{I_G}$.

5.2.3 Statistical Properties of the CLGSI

Assume that H and I_C have bivariate joint normal distribution; $\mathbf{P}$, $\mathbf{C}$, $\boldsymbol{\Gamma}$, and $\mathbf{w}$ are known, and $\boldsymbol{\beta}_C = \mathbf{T}_C^{-1}\boldsymbol{\Psi}_C\mathbf{a}_C$; then, the CLGSI properties are as follow:

1. $\sigma^2_{I_C} = \sigma_{HI_C}$, i.e., the variance of I_C ($\sigma^2_{I_C}$) and the covariance between H and I_C (σ_{HI_C}) are the same.
2. The maximized correlation between H and I_C is $\rho_{HI_C} = \frac{\sigma_{I_C}}{\sigma_H}$, where σ_{I_C} is the standard deviation of the variance of I_C ($\sigma^2_{I_C}$) and σ_H is the standard deviation of the variance of $H(\sigma^2_H)$.
3. The variance of the predicted error, $Var(H - I_C) = \left(1 - \rho^2_{HI_C}\right)\sigma^2_H$, is minimal.
4. The total variance of H explained by I_C is $\sigma^2_{I_C} = \rho^2_{HI_C}\sigma^2_H$.

Note that CLGSI properties 1 to 4 are the same as LMSI properties 1 to 4 and that both indices jointly incorporate phenotypic and marker information to predict the net genetic merit; however, the LMSI incorporates the marker information by the marker score values, whereas the CLGSI uses the GEBVs.

5.2.4 Estimating the CLGSI Parameters

Using the real maize (*Zea mays*) F_2 population with 248 genotypes (each with two repetitions), 233 molecular markers and three traits—GY (ton ha^{-1}), EHT (cm), and PHT (cm)—described in Sect. 5.1.8 of this chapter, we estimated matrices $\mathbf{P}$ and $\mathbf{C}$ using Eqs. (2.22) to (2.24) described in Chap. 2 of this book. The estimated matrices were $\widehat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}$ and $\widehat{\mathbf{C}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$.

In a similar manner, we estimated matrix $\boldsymbol{\Gamma}$ using Eqs. (5.21) to (5.23). The estimated matrix was $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}$. Note that matrices $\widehat{\mathbf{C}}$ and $\widehat{\boldsymbol{\Gamma}}$ have similar values. This means that, in the asymptotic context, we can assume that matrix $\boldsymbol{\Gamma}$ tends to matrix $\mathbf{C}$.

To estimate the CLMSI and its associated parameters (selection response, expected genetic gain per trait, etc.), we need to estimate the vector of coefficients $\boldsymbol{\beta}_C = \mathbf{T}_C^{-1}\boldsymbol{\Psi}_C\mathbf{a}_C$ as $\widehat{\boldsymbol{\beta}}_C = \widehat{\mathbf{T}}_C^{-1}\widehat{\boldsymbol{\Psi}}_C\mathbf{a}_C$, where $\widehat{\mathbf{T}}_C = \begin{bmatrix} \widehat{\mathbf{P}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$ and $\widehat{\boldsymbol{\Psi}}_C = \begin{bmatrix} \widehat{\mathbf{C}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$ are estimates of matrices $\mathbf{T}_C = \begin{bmatrix} \mathbf{P} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ and $\boldsymbol{\Psi}_C = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ respectively. The estimated CLGSI vector of coefficients $\widehat{\boldsymbol{\beta}}_C = \widehat{\mathbf{T}}_C^{-1}\widehat{\boldsymbol{\Psi}}_C\mathbf{a}_C$ is conformed by the vector of

phenotypic weights, $\widehat{\boldsymbol{\beta}}_{\mathbf{y}} = \left(\widehat{\mathbf{P}} - \widehat{\boldsymbol{\Gamma}}\right)^{-1}\left(\widehat{\mathbf{C}} - \widehat{\boldsymbol{\Gamma}}\right)\mathbf{w}$, and by the vector of genomic weights, $\widehat{\boldsymbol{\beta}}_G = \left[\mathbf{I} - \left(\widehat{\mathbf{P}} - \widehat{\boldsymbol{\Gamma}}\right)^{-1}\left(\widehat{\mathbf{C}} - \widehat{\boldsymbol{\Gamma}}\right)\right]\mathbf{w}$.

Let $\mathbf{w}' = [5 \quad -0.1 \quad -0.1]$ be the vector of economic weights; then, according to the estimated matrices $\widehat{\mathbf{P}}$, $\widehat{\mathbf{C}}$, and $\widehat{\boldsymbol{\Gamma}}$, $\widehat{\boldsymbol{\beta}}'_{\mathbf{y}} = [0.08 \quad -0.02 \quad -0.01]$ and $\widehat{\boldsymbol{\beta}}'_G = [4.92 \quad -0.08 \quad -0.09]$, whence the estimated CLGSI in the training population can be written as

$$\widehat{I}_C = \widehat{\boldsymbol{\beta}}_{\mathbf{y}}\mathbf{y} + \widehat{\boldsymbol{\beta}}_G\widehat{\boldsymbol{\gamma}}. \tag{5.43}$$

Suppose a selection intensity of 10% ($k_I = 1.755$); then, the estimated CLGSI selection response and expected genetic gain per trait were $\widehat{R}_C = k_I\sqrt{\widehat{\boldsymbol{\beta}}'_C\widehat{\mathbf{T}}_C\widehat{\boldsymbol{\beta}}_C} = 1.54$ and $\widehat{\mathbf{E}}'_C = k_I\dfrac{\widehat{\boldsymbol{\beta}}'_C\widehat{\boldsymbol{\Psi}}_C}{\sqrt{\widehat{\boldsymbol{\beta}}'_C\widehat{\mathbf{T}}_C\widehat{\boldsymbol{\beta}}_C}} = [0.36 \quad 1.04 \quad 1.70 \quad 0.36 \quad 1.53 \quad 2.38]$ respectively, whereas the estimated CLGSI accuracy was $\widehat{\rho}_{HI_C} = \dfrac{\widehat{\sigma}_{I_C}}{\widehat{\sigma}_H} = 0.814$.

The estimated LPSI selection response, expected genetic gain per trait, and accuracy were 0.601, $[0.09 \quad -0.81 \quad -0.89]$, and 0.32 respectively; thus, the CLGSI was more efficient to predict the net genetic merit than the LPSI because the CLGSI accuracy and selection response were 0.814 and 1.54 respectively.

5.2.5 *LGSI and CLGSI Efficiency Vs LMSI, GW-LMSI and LPSI Efficiency*

In this subsection, we compare the accuracy, selection response, and efficiency of the LGSI and CLGSI with the LMSI, the GW-LMSI, and the LPSI using the simulated data for a maize (*Zea mays*) population described in Chap. 2, Sect. 2.8.1.

Figure 5.3 presents the estimated accuracy values of the LMSI, the LGSI, the CLGSI, the LPSI, and the GW-LMSI for five simulated selection cycles. According to these results, for the first three selection cycles, the estimated accuracies of the indices, in decreasing order, were LMSI > LGSI > CLGSI > LPSI > GW-LMSI. That is, the highest estimated accuracy was obtained with the LMSI, whereas the lowest was obtained with the GW-LMSI. For the fourth and fifth selection cycles, the estimated accuracies, in decreasing order, were LMSI > LPSI > CLGSI > LGSI > GW-LMSI. This means that in all five selection cycles, the LMSI had the highest accuracy and the GW-LMSI had the lowest accuracy, whereas the estimated LGSI accuracy was reduced to fourth place. Thus, the accuracy of the LGSI tended to decrease after the first three selection cycles whereas LPSI accuracy was a constant.

To compare LGSI efficiency versus the efficiency of the other selection indices, we assumed that the interval between selection cycles in the LGSI is 1.5 years, whereas for CLGSI, LMSI, GW-LMSI, and LPSI, the interval was 4.0 years. Table 5.3 presents the estimated selection response of the LPSI, the LMSI, the

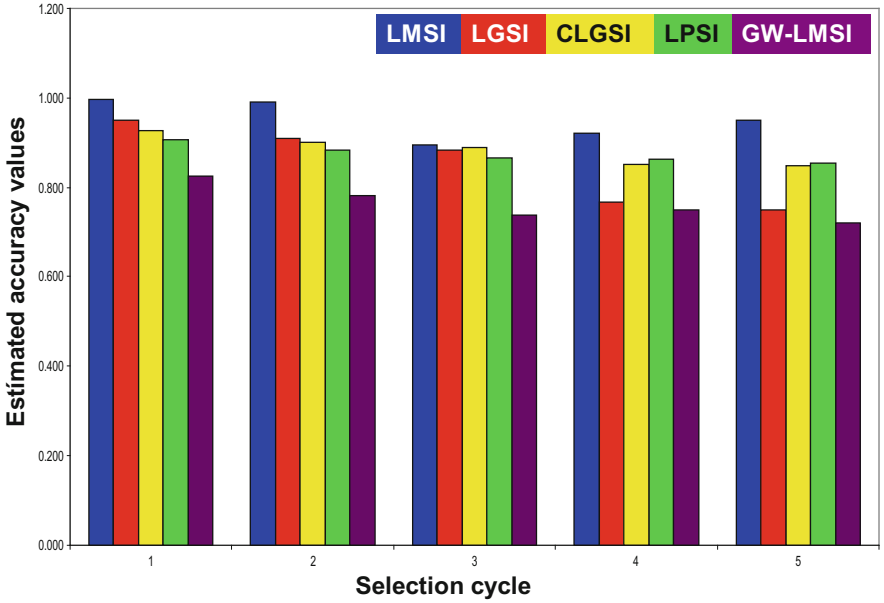

Fig. 5.3 Estimated accuracy values of the linear molecular selection index (LMSI), the LGSI, the combined LGSI (CLGSI), the LPSI, and the genome-wide LMSI (GW-LMSI) with the net genetic merit for four traits, 2500 markers, and 500 genotypes (each with four repetitions) in one environment for five simulated selection cycles

Table 5.3 Estimated selection response of the linear phenotypic selection index (LPSI), the linear molecular selection index (LMSI), the genome-wide LMSI (GW-LMSI), the linear genomic selection index (LGSI), and the combined LGSI (CLGSI), not including (first part of the Table) and including (second part of the Table) the interval length between selection cycles, obtained using five simulated selection cycles

Cycle	LPSI	LMSI	GW-LMSI	LGSI	C-LGSI
Estimated selection response not including the interval length					
1	17.84	19.60	16.24	14.36	18.24
2	15.66	24.36	13.88	13.90	16.02
3	14.44	14.70	12.13	13.59	14.61
4	14.29	15.29	12.48	12.30	14.14
5	13.86	15.15	11.49	11.38	13.51
Average	15.22	17.82	13.24	13.11	15.30
Estimated selection response including the interval length[a]					
1	4.46	4.90	4.06	9.58	4.56
2	3.92	6.09	3.47	9.27	4.00
3	3.61	3.68	3.03	9.06	3.65
4	3.57	3.82	3.12	8.20	3.53
5	3.47	3.79	2.87	7.59	3.38
Average	3.80	4.46	3.31	8.74	3.83

[a]The interval length for the LPSI, LMSI, GW-LMSI, and C-LGSI was 4 years, whereas the interval length for the LGSI was 1.5 years

Table 5.4 Estimated accuracy of the LMSI, the LGSI, the CLGSI, the LPSI, and the GW-LMSI; LMSI efficiency compared with LGSI, CLGSI, LPSI, and GW-LMSI efficiencies, expressed in percentages, for five simulated selection cycles

	Estimated accuracy					LMSI efficiency compared with			
Cycle	LMSI	LGSI	CLGSI	LPSI	GW-LMSI	LGSI	CLGSI	LPSI	GW-LMSI
1	1.00	0.95	0.93	0.91	0.83	4.93	7.45	10.07	20.67
2	0.99	0.91	0.90	0.88	0.78	8.78	9.88	12.14	26.81
3	0.90	0.88	0.89	0.87	0.74	1.26	0.64	3.43	21.27
4	0.92	0.77	0.85	0.86	0.75	19.99	8.12	6.57	22.5
5	0.95	0.75	0.85	0.86	0.72	26.71	12.2	11.11	31.88
Average	0.95	0.85	0.88	0.87	0.76	12.33	7.66	8.66	24.63

GW-LMSI, the LGSI, and the CLGSI, including and not including the interval between selection cycles (first and second parts of Table 5.3 respectively), obtained using five simulated selection cycles. According to the first part of Table 5.3, the average estimated selection responses, in decreasing order, of the LMSI, CLGSI, LPSI, GW-LMSI, and LGSI for the five simulated selection cycles were 17.82, 15.30, 15.22, 13.24, and 13.11 respectively, when the length of the interval between selection was not included. If the length of the interval between selection cycles is included when comparing the selection response of the indices in terms of time, the estimated selection response of LMSI, CLGSI, LPSI, GW-LMSI must be divided by 4 in each selection cycle, and the estimated LGSI selection response should be divided by 1.5. Thus, according to the second part of Table 5.3, if we include the length of the interval between selection cycles, the average estimated selection responses, in decreasing order, of LGSI, LMSI, CLGSI, LPSI, and GW-LMSI for the five simulated selection cycles were 8.74, 4.46, 3.83, 3.80, and 3.31. This means that in terms of time, the efficiency of the LGSI was higher than the efficiency of the other four selection indices.

Table 5.4 presents the estimated accuracy of the LMSI, LGSI, CLGSI, LPSI, and the GW-LMSI. In addition, Table 5.4 presents the efficiency when predicting the net genetic merit of the LMSI with respect to the LGSI, CLGSI, LPSI, and GW-LMSI as percentages, for five simulated selection cycles. Note that in this case, LMSI efficiency was higher than the efficiency of the other four selection indices, because the LMSI had the highest correlation with the net genetic merit.

References

Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R et al (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci 55:154–163

Ceron-Rojas JJ, Crossa J, Arief VN, 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 (Bethesda) 5:2155–2164

Cerón-Rojas JJ, Sahagún-Castellanos J (2016) Multivariate empirical Bayes to predict plant breeding values. Agrociencia 50:633–648

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331–341

Gianola D, Perez-Enciso M, Toro MA (2003) On marker-assisted prediction of genetic value: beyond the ridge. Genetics 163:347–365

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49 (1):12

Lin CY, Allaire FR (1977) Heritability of a linear combination of traits. Theor Appl Genet 51:1–3

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T et al (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77–123

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, ME

Mrode RA (2005) Linear models for the prediction of animal breeding values, 2nd edn. CABI Publishing, Cambridge, MA

Nordskog AW (1978) Some statistical properties of an index of multiple traits. Theor Appl Genet 52:91–94

Schott JR (2005) Matrix analysis for statistics, 2nd edn. Wiley, Hoboken, NJ

Searle S, Casella G, McCulloch CE (2006) Variance components. Wiley, Hoboken, NJ

Su G, Christensen OF, Ostersen T, Henryon M, Lund MS (2012) Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One 7(9):e45293

Technow F, Bürger A, Melchinger AE (2013) Genomic prediction of northern corn leaf blight resistance in maize with combined or separate training sets for heterotic groups. G3 (Bethesda) 3:197–203

Togashi K, Lin CY, Yamazaki T (2011) The efficiency of genome-wide selection for genetic improvement of net merit. J Anim Sci 89:2972–2980

Van Raden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Vattikuti S, Guo J, Chow CC (2012) Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. PLoS Genet 8(3):e1002637