Chapter 8
Linear Molecular and Genomic Eigen Selection Index Methods

Abstract The three main linear phenotypic eigen selection index methods are the eigen selection index method (ESIM), the restricted ESIM (RESIM) and the predetermined proportional gain ESIM (PPG-ESIM). The ESIM is an unrestricted index, but the RESIM and PPG-ESIM allow null and predetermined restrictions respectively to be imposed on the expected genetic gains of some traits, whereas the rest remain without any restrictions. These indices are based on the canonical correlation, on the singular value decomposition, and on the linear phenotypic selection indices theory. We extended the ESIM theory to the molecular-assisted and genomic selection context to develop a molecular ESIM (MESIM), a genomic ESIM (GESIM), and a genome-wide ESIM (GW-ESIM). Also, we extend the RESIM and PPG-ESIM theory to the restricted genomic ESIM (RGESIM), and to the predetermined proportional gain genomic ESIM (PPG-GESIM) respectively. The latter five indices use marker and phenotypic information jointly to predict the net genetic merit of the candidates for selection, but although MESIM uses only statistically significant markers linked to quantitative trait loci, the GW-ESIM uses all genome markers and phenotypic information and the GESIM, RGESIM, and PPG-GESIM use the genomic estimated breeding values and the phenotypic values to predict the net genetic merit. Using real and simulated data, we validated the theoretical results of all five indices.

8.1 The Molecular Eigen Selection Index Method

The molecular eigen selection index method (MESIM) is very similar to the linear molecular selection index (LMSI) described in Chap. 4; thus, it uses the same set of information to predict the net genetic merit of individual candidates for selection, and therefore needs the same set of conditions as those of the LMSI. The only difference between the two indices is how the vector of coefficients is obtained and the assumption associated with the vector of economic weights. Thus, although the LMSI obtains the vector of coefficients according to the linear phenotypic selection index (LPSI) described in Chap. 2 and assumes that the economic weights are known

and fixed, the MESIM assumes that the economic weights are unknown and fixed and obtains the vector of coefficients according to the ESIM theory.

8.1.1   The MESIM Parameters

In the MESIM context, the net genetic merit can be written as

$$H = \mathbf{w}_1'\mathbf{g} + \mathbf{w}_2'\mathbf{s} = \begin{bmatrix} \mathbf{w}_1' & \mathbf{w}_2' \end{bmatrix} \begin{bmatrix} \mathbf{g} \\ \mathbf{s} \end{bmatrix} = \mathbf{w}'\mathbf{a}, \tag{8.1}$$

where $\mathbf{g}' = \begin{bmatrix} g_1 & \cdots & g_t \end{bmatrix}$ is the vector of true breeding values, t is the number of traits, $\mathbf{w}_1' = \begin{bmatrix} w_1 & \cdots & w_t \end{bmatrix}$ is a vector of unknown economic weights associated with $\mathbf{g}$, $\mathbf{w}_2' = \begin{bmatrix} 0_1 & \cdots & 0_t \end{bmatrix}$ is a null vector associated with the vector of marker score values $\mathbf{s}' = \begin{bmatrix} s_1 & s_2 & \cdots & s_t \end{bmatrix}$, $\mathbf{w}' = \begin{bmatrix} \mathbf{w}_1' & \mathbf{w}_2' \end{bmatrix}$ and $\mathbf{a}' = \begin{bmatrix} \mathbf{g}' & \mathbf{s}' \end{bmatrix}$ (Chap. 4 for details). The MESIM index can be written as

$$I = \boldsymbol{\beta}_{\mathbf{y}}'\mathbf{y} + \boldsymbol{\beta}_s'\mathbf{s} = \begin{bmatrix} \boldsymbol{\beta}_{\mathbf{y}}' & \boldsymbol{\beta}_s' \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{s} \end{bmatrix} = \boldsymbol{\beta}'\mathbf{t}, \tag{8.2}$$

where $\mathbf{y}' = \begin{bmatrix} y_1 & \cdots & y_t \end{bmatrix}$ is the vector of phenotypic values; $\mathbf{s}' = \begin{bmatrix} s_1 & s_2 & \cdots & s_t \end{bmatrix}$ is the vector of marker scores; $\boldsymbol{\beta}_{\mathbf{y}}'$ and $\boldsymbol{\beta}_s$ are vectors of phenotypic and marker score weight values respectively, $\boldsymbol{\beta}' = \begin{bmatrix} \boldsymbol{\beta}_{\mathbf{y}}' & \boldsymbol{\beta}_G' \end{bmatrix}$ and $\mathbf{t}' = \begin{bmatrix} \mathbf{y}' & \mathbf{s}' \end{bmatrix}$. The objectives of the MESIM are the same as those of the ESIM (see Chap. 7 for details).

Let $Var(H) = \mathbf{w}'\boldsymbol{\Psi}_M\mathbf{w} = \sigma_H^2$ be the variance of H, $Var(I) = \boldsymbol{\beta}'\mathbf{T}_M\boldsymbol{\beta} = \sigma_I^2$ the variance of I, and $Cov(H, I) = \mathbf{w}'\boldsymbol{\Psi}_M\boldsymbol{\beta}$ the covariance between H and I, where $\boldsymbol{\Psi}_M = Var\begin{bmatrix} \mathbf{g} \\ \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{S}_M \\ \mathbf{S}_M & \mathbf{S}_M \end{bmatrix}$ and $\mathbf{T}_M = Var\begin{bmatrix} \mathbf{y} \\ \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{S}_M \\ \mathbf{S}_M & \mathbf{S}_M \end{bmatrix}$ are block matrices of size $2t \times 2t$ (t is the number of traits) of covariance matrices where $\mathbf{P}$, $\mathbf{S}_M$, and $\mathbf{C}$ are covariance matrices $t \times t$ of phenotypic ($\mathbf{y}$), marker score ($\mathbf{s}$), and genetic breeding ($\mathbf{g}$) values respectively. Let $\rho_{HI} = \frac{\mathbf{w}'\boldsymbol{\Psi}_M\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\boldsymbol{\Psi}_M\mathbf{w}}\sqrt{\boldsymbol{\beta}'\mathbf{T}_M\boldsymbol{\beta}}}$ and $h_I^2 = \frac{\boldsymbol{\beta}'\boldsymbol{\Psi}_M\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{T}_M\boldsymbol{\beta}}$ be the correlation between H and I, and the heritability of I respectively; then, the MESIM selection response can be written as

$$R = k_I\sigma_H\rho_{HI} \tag{8.3}$$

and

$$R = k_I\sigma_I h_I^2, \tag{8.4}$$

where k_I is the standardized selection differential (or selection intensity) associated with MESIM; $\sigma_H = \sqrt{\mathbf{w}'\boldsymbol{\Psi}_M\mathbf{w}}$ and $\sigma_I = \sqrt{\boldsymbol{\beta}'\mathbf{T}_M\boldsymbol{\beta}}$ are the standard deviations of the

variance of H and I respectively. It is assumed that k_I is fixed, and that matrices $\mathbf{T}_M$ and $\mathbf{\Psi}_M$ are known; therefore, we can maximize R by maximizing ρ_{HI} (Eq. 8.3) with respect to vectors $\mathbf{w}$ and $\mathbf{\beta}$, or by maximizing h_I^2 (Eq. 8.4) only with respect to vector $\mathbf{\beta}$.

Maximizing h_I^2 only with respect to $\mathbf{\beta}$ is simpler than maximizing ρ_{HI} with respect to $\mathbf{w}$ and $\mathbf{\beta}$; however, in the latter case the maximization process of ρ_{HI} gives more information associated with MESIM parameters than when h_I^2 is maximized only with respect to $\mathbf{\beta}$ (see Chap. 7, Eq. 7.13, for details). In this subsection, we maximize ρ_{HI} with respect to vectors $\mathbf{w}$ and $\mathbf{\beta}$ similar to the ESIM in Chap. 7, Sect. 7.1.1. Thus, we omit the steps and details of the maximization process of ρ_{HI}.

We maximize $\rho_{HI} = \frac{\mathbf{w}'\mathbf{\Psi}_M\mathbf{\beta}}{\sqrt{\mathbf{w}'\mathbf{\Psi}_M\mathbf{w}}\sqrt{\mathbf{\beta}'\mathbf{T}_M\mathbf{\beta}}}$ with respect to vectors $\mathbf{w}$ and $\mathbf{\beta}$ under the restrictions $\sigma_H^2 = \mathbf{w}'\mathbf{\Psi}\mathbf{w}$, $\sigma_I^2 = \mathbf{\beta}'\mathbf{T}\mathbf{\beta}$, and $0 < \sigma_H^2, \sigma_I^2 < \infty$, where σ_H^2 is the variance of $H = \mathbf{w}'\mathbf{a}$ and σ_I^2 is the variance of $I = \mathbf{\beta}'\mathbf{t}$. Thus, it is necessary to maximize the function

$$f(\mathbf{\beta}, \mathbf{w}, \mu, \phi) = \mathbf{w}'\mathbf{\Psi}\mathbf{\beta} - 0.5\mu(\mathbf{\beta}'\mathbf{T}\mathbf{\beta} - \sigma_I^2) - 0.5\phi(\mathbf{w}'\mathbf{\Psi}\mathbf{w} - \sigma_H^2) \quad (8.5)$$

with respect to $\mathbf{\beta}$, $\mathbf{w}$, μ, and ϕ, where μ and ϕ are Lagrange multipliers. The derivatives of Eq. (8.5) with respect to $\mathbf{\beta}$, $\mathbf{w}$, μ, and ϕ are:

$$\mathbf{\Psi}\mathbf{w} - \mu\mathbf{T}\mathbf{\beta} = \mathbf{0}, \quad (8.6)$$
$$\mathbf{\Psi}\mathbf{\beta} - \phi\mathbf{\Psi}\mathbf{w} = \mathbf{0}, \quad (8.7)$$
$$\mathbf{\beta}'\mathbf{T}\mathbf{\beta} = \sigma_I^2 \quad \text{and} \quad \mathbf{w}'\mathbf{\Psi}\mathbf{w} = \sigma_H^2, \quad (8.8)$$

respectively, where Eq. (8.8) denotes the restrictions imposed for maximizing ρ_{HI}. It can be shown (see Chap. 7) that vector $\mathbf{w}$ can be obtained as

$$\mathbf{w}_M = \mathbf{\Psi}_M^{-1}\mathbf{T}_M\mathbf{\beta} \quad (8.9)$$

and the net genetic merit in the MESIM context can be written as $H_M = \mathbf{w}_M'\mathbf{a}$; thus, the correlation between $H_M = \mathbf{w}_M'\mathbf{a}$ and I is $\rho_{H_M I} = \frac{\sqrt{\mathbf{\beta}'\mathbf{T}\mathbf{\beta}}}{\sqrt{\mathbf{\beta}'\mathbf{T}\mathbf{\Psi}^{-1}\mathbf{T}\mathbf{\beta}}}$ and the MESIM vector of coefficients ($\mathbf{\beta}$) that maximizes $\rho_{H_M I}$ can be obtained from equation

$$(\mathbf{T}^{-1}\mathbf{\Psi} - \lambda_M^2\mathbf{I}_{2t})\mathbf{\beta}_M = \mathbf{0}, \quad (8.10)$$

where $\mathbf{I}_{2t}$ is an identity matrix of size $2t \times 2t$ (t is the number of traits), and λ_M^2 and $\mathbf{\beta}_M$ are the *eigenvalue* and *eigenvector* of matrix $\mathbf{T}_M^{-1}\mathbf{\Psi}_M$. The words *eigenvalue* and *eigenvector* are derived from the German word *eigen*, which means *owned by* or *peculiar to*. Eigenvalues and eigenvectors are sometimes called *characteristic values* and *characteristic vectors*, *proper values* and *proper vectors*, or *latent values* and *latent vectors* (Meyer 2000). The square root of λ_M^2 (λ_M) is the canonical correlation between $H_M = \mathbf{w}_M'\mathbf{a}$ and $I_M = \mathbf{\beta}_M'\mathbf{t}$, and the optimized MESIM index can be written as $I_M = \mathbf{\beta}_M'\mathbf{t}$. Using a similar procedure to that described in Chap. 7 (Eq. 7.17), it can

be show that vector $\boldsymbol{\beta}_M$ can be transformed into $\boldsymbol{\beta}_C = \mathbf{F}\boldsymbol{\beta}_M$, where $\mathbf{F}$ is a diagonal matrix with values equal to any real number, except zero values.

The maximized correlation between $H_M = \mathbf{w}'_M\mathbf{a}$ and $I_M = \boldsymbol{\beta}'_M\mathbf{t}$, or MESIM accuracy, is

$$\rho_{H_M I_M} = \frac{\sqrt{\boldsymbol{\beta}'_M \mathbf{T}_M \boldsymbol{\beta}_M}}{\sqrt{\boldsymbol{\beta}'_M \mathbf{T}_M \boldsymbol{\Psi}_M^{-1} \mathbf{T}_M \boldsymbol{\beta}_M}} = \frac{\sigma_{I_M}}{\sigma_{H_M}}, \tag{8.11}$$

where $\sigma_{I_M} = \sqrt{\boldsymbol{\beta}'_M \mathbf{T}_M \boldsymbol{\beta}_M}$ is the standard deviation of $I_M = \boldsymbol{\beta}'_M\mathbf{t}$, and $\sigma_{H_M} = \sqrt{\boldsymbol{\beta}'_M \mathbf{T}_M \boldsymbol{\Psi}_M^{-1} \mathbf{T}_M \boldsymbol{\beta}_M}$ is the standard deviation of $H_M = \mathbf{w}'_M\mathbf{a}$.

The maximized selection response and expected genetic gain per trait of MESIM are

$$R_M = k_I \sqrt{\boldsymbol{\beta}'_{M_1} \mathbf{T}_M \boldsymbol{\beta}_{M_1}} \tag{8.12}$$

and

$$\mathbf{E}_M = k_I \frac{\boldsymbol{\Psi}_M \boldsymbol{\beta}_{M_1}}{\sqrt{\boldsymbol{\beta}'_{M_1} \mathbf{T}_M \boldsymbol{\beta}_{M_1}}}, \tag{8.13}$$

respectively, where $\boldsymbol{\beta}_{M_1}$ is the first eigenvector of matrix $\mathbf{T}_M^{-1}\boldsymbol{\Psi}_M$. If vector $\boldsymbol{\beta}_{M_1}$ is multiplied by matrix $\mathbf{F}$, we obtain $\boldsymbol{\beta}_{C_1} = \mathbf{F}\boldsymbol{\beta}_{M_1}$; in this case, we can replace $\boldsymbol{\beta}_{M_1}$ with $\boldsymbol{\beta}_{C_1} = \mathbf{F}\boldsymbol{\beta}_{M_1}$ in Eqs. (8.12) and (8.13), and the optimized MESIM index should be written as $I_M = \boldsymbol{\beta}'_{C_1}\mathbf{y}$.

8.1.2 Estimating MESIM Parameters

We estimate the MESIM parameters using the same procedure described in Chap. 7 (Sect. 7.1.4) to estimate the ESIM parameters. Let $\widehat{\mathbf{C}}$, $\widehat{\mathbf{P}}$, and $\widehat{\mathbf{S}}_M$ be the estimates of the genotypic, phenotypic, and marker scores covariance matrices, $\widehat{\mathbf{T}}_M = \begin{bmatrix} \widehat{\mathbf{P}} & \widehat{\mathbf{S}}_M \\ \widehat{\mathbf{S}}_M & \widehat{\mathbf{S}}_M \end{bmatrix}$ and $\widehat{\boldsymbol{\Psi}}_M = \begin{bmatrix} \widehat{\mathbf{C}} & \widehat{\mathbf{S}}_M \\ \widehat{\mathbf{S}}_M & \widehat{\mathbf{S}}_M \end{bmatrix}$ the estimated block matrices (Chap. 4) and $\widehat{\mathbf{W}} = \widehat{\mathbf{T}}_M^{-1}\widehat{\boldsymbol{\Psi}}_M$; then, to find the estimators $\widehat{\boldsymbol{\beta}}_{M_1}$ and $\widehat{\lambda}^2_{M_1}$ of the first eigenvector ($\boldsymbol{\beta}_{M_1}$) and the first eigenvalue ($\lambda^2_{M_1}$) respectively, we need to solve the equation

$$\left(\widehat{\mathbf{W}}\widehat{\mathbf{W}}' - \widehat{\mu}_j\mathbf{I}\right)\widehat{\boldsymbol{\beta}}_{M_j} = \mathbf{0}, \tag{8.14}$$

where $\widehat{\mu}_j = \widehat{\lambda}^4_{M_j}$, $j = 1, 2, \ldots, 2t$. For additional details, see Eqs. (7.22) and (7.23), and Sect. 7.1.5 of Chap. 7. The result of Equation (8.14) allow the MESIM index

$(I_M = \boldsymbol{\beta}'_{M_1}\mathbf{t})$ to be estimated as $\widehat{I}_M = \widehat{\boldsymbol{\beta}}'_{M_1}\mathbf{t}$, whereas the estimator of the maximized ESIM selection response and its expected genetic gain per trait can be denoted by

$$\widehat{R}_M = k_I\sqrt{\widehat{\boldsymbol{\beta}}'_{M_1}\widehat{\mathbf{T}}_M\widehat{\boldsymbol{\beta}}_{M_1}} \quad \text{and} \quad \widehat{\mathbf{E}}_M = k_I\frac{\widehat{\boldsymbol{\Psi}}_M\widehat{\boldsymbol{\beta}}_{M_1}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{M_1}\widehat{\mathbf{T}}_M\widehat{\boldsymbol{\beta}}_{M_1}}}, \qquad (8.15)$$

respectively.

8.1.3 Numerical Examples

To validate the MESIM theoretical results, we use a real maize (*Zea mays*) F_2 population with 247 genotypes (each with two repetitions), 195 molecular markers, and two traits—plant height (PHT, cm) and ear height (EHT, cm)—evaluated in one environment. We coded the marker homozygous loci for the allele from the first parental line by 1, whereas the marker homozygous loci for the allele from the second parental line was coded by -1 and the marker heterozygous loci by 0. The estimated phenotypic, genetic, and marker scores covariance matrices were
$$\widehat{\mathbf{P}} = \begin{bmatrix} 191.81 & 106.89 \\ 106.89 & 167.93 \end{bmatrix}, \quad \widehat{\mathbf{C}} = \begin{bmatrix} 83.00 & 57.44 \\ 57.44 & 59.80 \end{bmatrix}, \quad \text{and} \quad \widehat{\mathbf{S}}_M = \begin{bmatrix} 15.750 & 0.983 \\ 0.983 & 28.083 \end{bmatrix}$$
respectively, and the vector of economic weights was $\mathbf{a}' = [\,\mathbf{w}' \quad \mathbf{0}'\,]$, where $\mathbf{w}' = [\,-1 \quad -1\,]$ and $\mathbf{0}' = [\,0 \quad 0\,]$. Details of how to estimate the marker scores and their variance were given in Chap. 4.

We compare LMSI versus MESIM efficiency. The estimated LMSI vector of coefficients was $\widehat{\boldsymbol{\beta}}' = \mathbf{a}'\widehat{\boldsymbol{\Psi}}_M\widehat{\mathbf{T}}_M^{-1} = [\,-0.59 \quad -0.18 \quad -0.41 \quad -0.82\,]$. Using a 10% selection intensity ($k_I = 1.755$), the estimated LMSI selection response and the expected genetic gain per trait were $\widehat{R} = k_I\sqrt{\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{T}}_M\widehat{\boldsymbol{\beta}}} = 20.41$ and $\widehat{\mathbf{E}}' = k_I\dfrac{\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\Psi}}_M}{\sqrt{\widehat{\boldsymbol{\beta}}'\widehat{\mathbf{T}}_M\widehat{\boldsymbol{\beta}}}} = [\,-10.09 \quad -10.31 \quad -2.53 \quad -4.39\,]$ respectively, whereas the estimated LMSI accuracy was $\widehat{\rho}_{H\widehat{I}} = \dfrac{\widehat{\sigma}_I}{\widehat{\sigma}_H} = 0.72$.

Vector $\widehat{\boldsymbol{\beta}}'_{M_1} = [\,0.089 \quad -0.061 \quad -0.536 \quad 0.837\,]$ was the original estimated MESIM vector of coefficients. Using matrix $\mathbf{F} = \begin{bmatrix} -0.1 & 0 & 0 & 0 \\ 0 & -0.1 & 0 & 0 \\ 0 & 0 & 0.75 & 0 \\ 0 & 0 & 0 & -0.75 \end{bmatrix}$, vector $\widehat{\boldsymbol{\beta}}'_{M_1}$ was transformed as $\widehat{\boldsymbol{\beta}}'_{C_1} = \widehat{\boldsymbol{\beta}}'_{M_1}\mathbf{F} = [\,-0.009 \quad 0.006 \quad -0.402 \quad 0.628\,]$ and then the estimated MESIM index was $\widehat{I}_M = -0.009\,\text{PHT} + 0.006\,\text{EHT} - 0.402\,\text{S}_{\text{PHT}} + 0.628\,\text{S}_{\text{EHT}}$, where S_{PHT} and S_{EHT} denote the marker scores associated with PHT and EHT respectively. The estimated MESIM expected

genetic gain, selection response, and accuracy were $\widehat{\mathbf{E}}'_M = k_I \dfrac{\widehat{\boldsymbol{\beta}}'_{C_1} \widehat{\boldsymbol{\Psi}}_M}{\sqrt{\widehat{\boldsymbol{\beta}}'_{C_1} \widehat{\mathbf{T}}_M \widehat{\boldsymbol{\beta}}_{C_1}}} =$

$[-3.438 \quad -8.516 \quad -3.319 \quad -8.372]$, $\quad \widehat{R}_M = k_I \sqrt{\widehat{\boldsymbol{\beta}}'_{C_1} \widehat{\mathbf{T}}_M \widehat{\boldsymbol{\beta}}_{C_1}} = 6.573$ and

$\widehat{\rho}_{H_M \widehat{I}_M} = \dfrac{\widehat{\sigma}_{I_M}}{\widehat{\sigma}_{H_M}} = 0.99$ respectively.

The inner product of the estimated LMSI and MESIM vector of coefficients were 1.221 and 0.556 respectively, whence the estimated LMSI selection response (20.41) divided by 1.221 was 16.716, and the estimated MESIM selection response (6.573) divided by 0.556 was 11.821. That is, the estimated LMSI selection response was higher than the estimated MESIM selection response for this data set. Similar results were found when we compared the estimated LMSI expected genetic gain per trait with the estimated MESIM expected genetic gain per trait. Finally, Fig. 8.1 presents the frequency distribution of the 247 estimated MESIM values for the real data set described earlier, which approaches normal distribution, as we would expect.

Now with a selection intensity of 10% ($k_I = 1.755$), we compare the LMSI and MESIM efficiency using the simulated data set described in Sect. 2.8.1 of Chap. 2 for four phenotypic selection cycles, each with four traits (T_1, T_2, T_3 and T_4), 500 genotypes, and four replicates of each genotype. The economic weights for T_1, T_2, T_3, and T_4 were 1, −1, 1, and 1 respectively. For this data set, we did not use the linear transformation $\widehat{\boldsymbol{\beta}}_{C_1} = \mathbf{F}\widehat{\boldsymbol{\beta}}_{M_1}$.

The estimated selection responses of the linear marker, combined genomic and genome-wide selection indices (LMSI, CLGSI, and GW-LMSI respectively; see
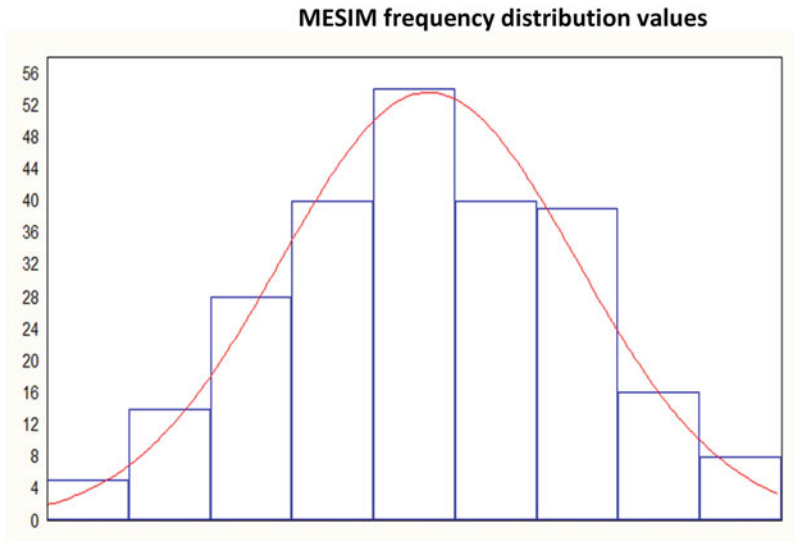

Fig. 8.1 Frequency distribution of 247 estimated molecular eigen selection index method (MESIM) values for one selection cycle in an environment for a real maize (*Zea mays*) F$_2$ population with 195 molecular markers and two traits, plant height (PHT, cm) and ear height (EHT, cm), and their associated marker scores S$_{PHT}$ and S$_{EHT}$ respectively

Chaps. 4 and 5 for details) for four simulated selection cycles when their vectors of coefficients were normalized, are presented in Table 8.1. Also, in this table the selection responses of the estimated linear molecular, genomic, and genome-wide eigen selection index methods (MESIM, GESIM, and GW-ESIM respectively; details in Sect. 8.2) are shown for four simulated selection cycles. The average of the estimated LMSI selection response was 2.22, whereas the average of the estimated MESIM selection response was 1.69. The estimated LMSI selection response was higher than that of the MESIM.

Table 8.2 presents the estimated LMSI and MESIM expected genetic gains for four traits (T1, T2, T3, and T4) and their associated marker scores (S1, S2, S3, and S4) for four simulated selection cycles. The averages of the estimated LMSI

Table 8.1 Estimated linear molecular, combined genomic, and genome-wide selection index (LMSI, CLGSI and GW-LMSI respectively) selection responses when their vectors of coefficients are normalized for four simulated selection cycles

Cycle	Estimated selection response					
	LMSI	CLGSI	GW-LMSI	MESIM	GESIM	GW-ESIM
1	0.02	1.24	0.93	0.50	3.95	0.73
2	4.94	0.80	0.80	1.21	3.07	1.06
3	3.69	0.34	0.93	3.91	2.05	0.77
4	0.23	0.35	0.83	1.15	1.90	1.14
Average	2.22	0.68	0.87	1.69	2.74	0.93

Estimated linear molecular, genomic, and genome-wide eigen selection index method (MESIM, GESIM, and GW-ESIM respectively) selection responses for four simulated selection cycles. The selection intensity was 10% ($k_I = 1.755$)

Table 8.2 Estimated linear molecular selection index (LMSI) and estimated linear molecular eigen selection index method (MESIM) expected genetic gains for four traits (T1, T2, T3, and T4) and their associated marker scores (S1, S2, S3, and S4) for four simulated selection cycles. The selection intensity was 10% ($k_I = 1.755$)

	Estimated LMSI expected genetic gain							
	Traits				Marker scores			
Cycle	T1	T2	T3	T4	S1	S2	S3	S4
1	24.48	−0.01	0.74	−0.87	4.18	−1.14	0.72	0.79
2	7.14	−3.39	2.62	1.55	3.78	−2.30	1.02	1.37
3	9.17	−3.04	1.87	1.21	6.22	−1.51	1.02	0.26
4	10.16	−1.95	1.17	1.88	8.63	−3.83	0.09	0.13
Average	12.74	−2.10	1.60	0.94	5.70	−2.19	0.71	0.64
	Estimated MESIM expected genetic gain							
	Traits				Marker scores			
Cycle	T1	T2	T3	T4	S1	S2	S3	S4
1	27.48	2.60	−1.03	−2.64	3.85	0.00	−0.04	−0.43
2	8.82	−4.75	0.37	2.11	14.06	4.09	0.38	−2.76
3	9.83	1.74	0.72	0.37	8.03	1.76	0.31	0.34
4	11.47	−1.13	−1.64	1.53	8.66	−3.96	−1.47	0.04
Average	14.40	−0.38	−0.39	0.34	8.65	0.47	−0.21	−0.70

expected genetic gains for the four traits and their associated marker scores were 12.74, -2.10, 1.60, 0.94, 5.70, -2.19, 0.71, and 0.64 respectively, whereas the averages of the estimated MESIM expected genetic gains for the four traits and their associated marker scores were 14.40, -0.38, -0.39, 0.34, 8.65, 0.47, -0.21, and -0.70 respectively. Except for trait T1 and its associated molecular scores, the estimated LMSI expected genetic gains per trait were higher than the estimated MESIM expected genetic gains. Thus, for this data set, LMSI efficiency was greater than MESIM efficiency.

Chapter 11 presents RIndSel, a user-friendly graphical unit interface in JAVA that is useful for estimating the LMSI and ESIM parameters and selecting parents for the next selection cycle.

8.2   The Linear Genomic Eigen Selection Index Method

The linear genomic eigen selection index method (GESIM) is based on the standard CLGSI described in Chap. 5, and uses genomic estimated breeding values (GEBVs) and phenotypic values jointly to predict the net genetic merit. Thus, conditions for constructing a valid GESIM are the same as those for constructing the CLGSI. Also, the MESIM theory described in Sect. 8.1 is directly applied to the GESIM and only minor changes are necessary in GESIM theory. For example, instead of marker scores, the GESIM uses GEBVs to predict the net genetic merit; thus, the details of the estimation process are the same as for the MESIM.

8.2.1   The GESIM Parameters

In the GESIM context, the net genetic merit can be written as

$$H = \mathbf{w}_1' \mathbf{g} + \mathbf{w}_2' \boldsymbol{\gamma} = \begin{bmatrix} \mathbf{w}_1' & \mathbf{w}_2' \end{bmatrix} \begin{bmatrix} \mathbf{g} \\ \boldsymbol{\gamma} \end{bmatrix} = \mathbf{w}' \boldsymbol{\alpha}, \tag{8.16}$$

where $\mathbf{g}' = \begin{bmatrix} g_1 & \ldots & g_t \end{bmatrix}$ is the vector of true breeding values, t is the number of traits, $\mathbf{w}_1' = \begin{bmatrix} w_1 & \cdots & w_t \end{bmatrix}$ is a vector of unknown economic weights associated with $\mathbf{g}$, $\mathbf{w}_2' = \begin{bmatrix} 0_1 & \cdots & 0_t \end{bmatrix}$ is a null vector associated with the vector of genomic breeding values $\boldsymbol{\gamma}' = \begin{bmatrix} \gamma_1 & \gamma_2 & \ldots & \gamma_t \end{bmatrix}$, $\mathbf{w}' = \begin{bmatrix} \mathbf{w}_1' & \mathbf{w}_2' \end{bmatrix}$, and $\boldsymbol{\alpha}' = \begin{bmatrix} \mathbf{g}' & \boldsymbol{\gamma}' \end{bmatrix}$. The estimator of $\boldsymbol{\gamma}$ is the GEBV (see Chap. 5 for additional details). The GESIM index can be written as

$$I = \boldsymbol{\beta}_y' \mathbf{y} + \boldsymbol{\beta}_\gamma' \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\beta}_y' & \boldsymbol{\beta}_\gamma' \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \boldsymbol{\gamma} \end{bmatrix} = \boldsymbol{\beta}' \mathbf{f}, \tag{8.17}$$

where $\mathbf{y}' = \begin{bmatrix} y_1 & \cdots & y_t \end{bmatrix}$ is the vector of phenotypic values; $\boldsymbol{\beta}'_{\mathbf{y}}$ and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ are vectors of weights of phenotypic and genomic breeding values weights respectively; $\boldsymbol{\beta}' = \begin{bmatrix} \boldsymbol{\beta}'_{\mathbf{y}} & \boldsymbol{\beta}'_{\boldsymbol{\gamma}} \end{bmatrix}$ and $\mathbf{f}' = \begin{bmatrix} \mathbf{y}' & \boldsymbol{\gamma}' \end{bmatrix}$.

Let $Var(H) = \mathbf{w}'\mathbf{A}\mathbf{w} = \sigma_H^2$ be the variance of $H = \mathbf{w}'\boldsymbol{\alpha}$, $Var(I) = \boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta} = \sigma_I^2$ the variance of $I = \boldsymbol{\beta}'\mathbf{f}$, and $Cov(H, I) = \mathbf{w}'\mathbf{A}\boldsymbol{\beta} = \sigma_{HI}$ the covariance between H and I, where $\mathbf{A} = Var\begin{bmatrix} \mathbf{g} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ and $\boldsymbol{\Phi} = Var\begin{bmatrix} \mathbf{y} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ are block matrices $2t \times 2t$ (t is the number of traits) of covariance matrices and $\mathbf{P}$, $\boldsymbol{\Gamma}$, and $\mathbf{C}$ are covariance matrices of phenotypic ($\mathbf{y}$), genomic ($\boldsymbol{\gamma}$), and genetic ($\mathbf{g}$) values respectively. Then, $\rho_{HI} = \dfrac{\mathbf{w}'\mathbf{A}\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{A}\mathbf{w}}\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}}}$ is the correlation between $H = \mathbf{w}'\boldsymbol{\alpha}$ and $I = \boldsymbol{\beta}'\mathbf{f}$ and the GESIM selection response can be written as

$$R = k_I \sigma_H \rho_{HI}, \tag{8.18}$$

where k_I is the standardized selection differential (or selection intensity) associated with the GESIM and $\sigma_H = \sqrt{\mathbf{w}'\mathbf{A}\mathbf{w}}$ is the standard deviation of the variance of H. It is assumed that k_I is fixed, and that matrices $\boldsymbol{\Phi}$ and $\mathbf{A}$ are known; then, we can maximize R by maximizing ρ_{HI} with respect to vectors $\mathbf{w}$ and $\boldsymbol{\beta}$ under the restrictions $\sigma_H^2 = \mathbf{w}'\mathbf{A}\mathbf{w}$, $\sigma_I^2 = \boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}$, and $0 < \sigma_H^2, \sigma_I^2 < \infty$; similar to the MESIM.

It can be shown that the vector $\mathbf{w}$ in the GESIM context is

$$\mathbf{w}_G = \mathbf{A}^{-1}\boldsymbol{\Phi}\boldsymbol{\beta} \tag{8.19}$$

and that the net genetic merit can be written as $H_G = \mathbf{w}'_G\boldsymbol{\alpha}$. The correlation between $H_G = \mathbf{w}'_G\boldsymbol{\alpha}$ and $I = \boldsymbol{\beta}'\mathbf{f}$ is $\rho_{H_GI} = \dfrac{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}}}{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}\boldsymbol{\beta}}}$ and the GESIM index vector of coefficients that maximizes ρ_{H_GI} can be obtained from the equation

$$\left(\boldsymbol{\Phi}^{-1}\mathbf{A} - \lambda_G^2\mathbf{I}_{2t}\right)\boldsymbol{\beta}_G = \mathbf{0}, \tag{8.20}$$

where $\mathbf{I}_{2t}$ is an identity matrix of size $2t \times 2t$ (t is the number of traits); the optimized GESIM index can be written as $I_G = \boldsymbol{\beta}'_G\mathbf{f}$. By Eqs. (8.19) and (8.20), GESIM accuracy can be written as

$$\rho_{H_GI_G} = \frac{\sigma_{I_G}}{\sigma_{H_G}}, \tag{8.21}$$

where $\sigma_{I_G} = \sqrt{\boldsymbol{\beta}'_G\boldsymbol{\Phi}\boldsymbol{\beta}_G}$ is the standard deviation of $I_G = \boldsymbol{\beta}'_G\mathbf{f}$, and $\sigma_{H_G} = \sqrt{\boldsymbol{\beta}'_G\boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}\boldsymbol{\beta}_G}$ is the standard deviation of $H_G = \mathbf{w}'_G\boldsymbol{\alpha}$. In Eq. (8.20), $\lambda_G^2 = \rho_{H_GI_G}^2$ is the square of the canonical correlation between H_G and I_G, and $\boldsymbol{\beta}_G$ is the canonical vector associated with $\lambda_G^2 = \rho_{H_GI_G}^2$.

The maximized GESIM selection response and expected genetic gain per trait are

$$R_G = k_I \sqrt{\boldsymbol{\beta}'_G \boldsymbol{\Phi} \boldsymbol{\beta}_G} \tag{8.22}$$

and

$$\mathbf{E}_G = k_I \frac{\mathbf{A}\boldsymbol{\beta}_G}{\sqrt{\boldsymbol{\beta}'_G \boldsymbol{\Phi} \boldsymbol{\beta}_G}}, \tag{8.23}$$

respectively, where $\boldsymbol{\beta}_G$ is the first eigenvector of matrix $\boldsymbol{\Phi}^{-1}\mathbf{A}$. Vector $\boldsymbol{\beta}_G$ can be transformed as $\boldsymbol{\beta}_{CG} = \mathbf{F}\boldsymbol{\beta}_G$, where $\mathbf{F}$ is a diagonal matrix defined earlier.

8.2.2 Numerical Examples

To compare the CLGSI versus GESIM theoretical results, we use a real maize (*Zea mays*) F_2 population with 244 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). We estimated matrices $\mathbf{P}$ and $\mathbf{C}$ using Eqs. (2.22) to (2.24) described in Chap. 2, whence 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 $\boldsymbol{\Gamma}$ by applying Eqs. (5.21) to (5.23) described in Chap. 5 using phenotypic and marker information jointly; the estimated matrix was $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}$. The selection intensity for making a selection cycle was 10% ($k_I = 1.755$) and the vector of economic weights was $\mathbf{w}' = \begin{bmatrix} 5 & -0.1 & -0.1 & 0 & 0 & 0 \end{bmatrix}$. To obtain the estimated vector of coefficient of CLGSI ($\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Phi}}^{-1}\widehat{\mathbf{A}}\mathbf{w}$) and GESIM (Eq. 8.20), it is necessary to construct matrices $\widehat{\mathbf{A}} = \begin{bmatrix} \widehat{\mathbf{C}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$ and $\widehat{\boldsymbol{\Phi}} = \begin{bmatrix} \widehat{\mathbf{P}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$.

The estimated CLGSI vector of coefficients for the traits GY, EHT, and PHT and their associated GEBVs (GEBV$_{GY}$, GEBV$_{EHT}$, and GEBV$_{PHT}$ respectively) was $\widehat{\boldsymbol{\beta}}' = \begin{bmatrix} 0.08 & -0.02 & -0.01 & 4.92 & -0.08 & -0.09 \end{bmatrix}$, whereas the estimated CLGSI selection response, accuracy, and expected genetic gain per trait were $\widehat{R} = k_I \sqrt{\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}} = 1.54$, $\widehat{\rho}_{HI} = \dfrac{\widehat{\sigma}_I}{\widehat{\sigma}_H} = 0.814$, and $\widehat{\mathbf{E}}' = k_I \dfrac{\widehat{\boldsymbol{\beta}}' \widehat{\mathbf{A}}}{\sqrt{\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}}} =$ $\begin{bmatrix} 0.36 & 1.04 & 1.70 & 0.36 & 1.53 & 2.38 \end{bmatrix}$ respectively. Finally, $\widehat{I} = 0.08\text{GY} - 0.02$ EHT $- 0.01\text{PHT} + 4.92\text{GEBV}_{GY} - 0.08\text{GEBV}_{EHT} - 0.09\text{GEBV}_{PHT}$ was the estimated CLGSI.

The estimated GESIM vector of coefficients, selection response, accuracy, and expected genetic gain per trait were $\widehat{\boldsymbol{\beta}}'_{G_1} = [-0.207 \quad 0.029 \quad 0.041 \quad 0.820 \quad 0.337 \quad 0.411]$, $\widehat{R}_G = k_I \sqrt{\widehat{\boldsymbol{\beta}}'_{G_1} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{G_1}} = 6.288$,

$$\widehat{\rho}_{\widehat{H}_G \widehat{I}_G} = \frac{\sqrt{\widehat{\boldsymbol{\beta}}'_{G_1} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{G_1}}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{G_1} \widehat{\boldsymbol{\Phi}} \widehat{\mathbf{A}}^{-1} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{G_1}}} = 0.9056, \qquad \text{and} \qquad \widehat{\mathbf{E}}'_G = k_1 \frac{\widehat{\boldsymbol{\beta}}'_{G_1} \widehat{\mathbf{A}}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{G_1} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{G_1}}} =$$

$[0.369 \quad 5.528 \quad 9.186 \quad 0.370 \quad 5.250 \quad 8.702]$ respectively.

Fig. 8.2 presents the frequency distribution of the 244 estimated GESIM index values for one (Fig. 8.2a) and three traits (Fig. 8.2b) using the real data set described earlier. The frequency distribution of the estimated GESIM index values approaches the normal distribution for both indices.

Now, we compare the estimated CLGSI and GESIM selection response and expected genetic gain per trait using the simulated data set described in Sect. 2.8.1 of Chap. 2 for four phenotypic selection cycles, each with four traits (T_1, T_2, T_3 and T_4), 500 genotypes, and four replicates per genotype. The economic weights of T_1, T_2, T_3, and T_4 were 1, −1, 1, and 1 respectively and the selection intensity for both
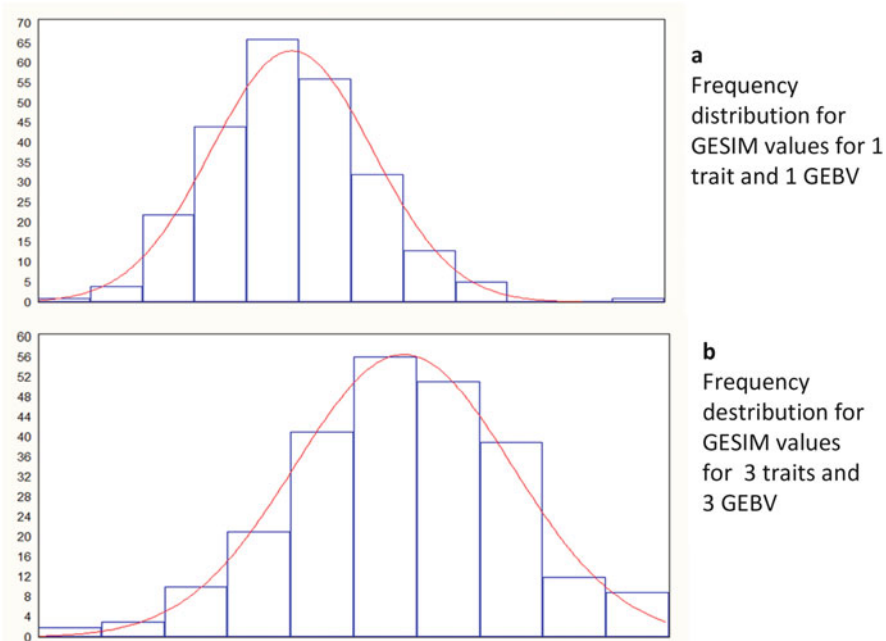

Fig. 8.2 Frequency distribution of the 244 estimated genomic eigen selection index method (GESIM) values for the one-trait case (**a**) and for the three-trait case (**b**) for one selection cycle in an environment for a real maize (*Zea mays*) F$_2$ population with 233 molecular markers. Note that the frequency distribution of the estimated GESIM index values approaches normal distribution for both indices

Table 8.3 Estimated combined linear genomic selection index (CLGSI) and estimated GESIM expected genetic gains for four traits (T1, T2, T3, and T4) and their associated genomic estimated breeding values (GEBV1, GEBV2, GEBV3, and GEBV4) for four simulated selection cycles. The selection intensity was 10% ($k_I = 1.755$)

| | Estimated CLGSI expected genetic gain | | | | | | | |
| | Traits | | | | Genomic estimated breeding value | | | |
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	7.46	−3.69	3.26	1.60	7.28	−4.38	3.72	3.29
2	7.08	−3.45	2.91	1.17	7.08	−3.63	3.66	2.67
3	7.81	−3.51	2.06	0.76	7.30	−3.92	2.35	2.40
4	7.46	−2.76	2.48	0.81	6.84	−2.79	2.79	2.40
Average	7.45	−3.35	2.68	1.09	7.13	−3.68	3.13	2.69
	Estimated GESIM expected genetic gain							
	Traits				Genomic estimated breeding value			
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	8.28	−3.51	2.93	0.92	7.77	−4.27	3.52	2.64
2	7.89	−3.09	2.42	0.82	7.40	−3.41	3.29	2.38
3	8.47	−3.26	1.69	0.46	7.55	−3.78	2.11	2.16
4	8.08	−2.46	2.04	0.66	7.15	−2.67	2.53	2.39
Average	8.18	−3.08	2.27	0.71	7.46	−3.53	2.86	2.39

indices was 10% ($k_I = 1.755$). For this data set, matrix **F** was an identity matrix of size 8×8 in all four selection cycles.

For this data set, the averages of the estimated CLGSI and GESIM selection responses were 0.68 and 2.74 (Table 8.1) respectively. The estimated CLGSI selection response was lower than the estimated GESIM selection response. Table 8.3 presents the estimated CLGSI and GESIM expected genetic gain for four traits (T1, T2, T3, and T4) and their associated genomic estimated breeding values (GEBV1, GEBV2, GEBV3, and GEBV4) for four simulated selection cycles. The averages of the estimated CLGSI expected genetic gains for the four traits and their associated GEBVs were 7.45, −3.35, 2.68, 1.09, 7.13, −3.68, 3.13, and 2.69 respectively, whereas the averages of the estimated GESIM expected genetic gains for the four traits and their associated GEBVs were 8.18, −3.08, 2.27, 0.71, 7.46, −3.53, 2.86, and 2.39 respectively. The estimated CLGSI and GESIM expected genetic gains per trait were very similar.

8.3 The Genome-Wide Linear Eigen Selection Index Method

The MESIM requires regressing phenotypic values on marker coded values to predict the marker score values for each individual candidate for selection, and then combining the marker scores with phenotypic information using the MESIM

to obtain a final prediction of the net genetic merit. In addition, the GESIM requires fitting of a statistical model to estimate all available marker effects in the training population; these estimates are then used to obtain GEBVs, which are predictors of breeding values. Crossa and Cerón-Rojas (2011) extended the ESIM theory to a genome-wide linear molecular ESIM (GW-ESIM) similar to the GW-LMSI described in Chap. 4. The GW-LMSI and GW-ESIM are very similar and only minor changes are necessary in GW-ESIM; for example, instead of estimating the GW-LMSI vector of coefficients according to the LPSI method (Chap. 2), the GW-ESIM vector of coefficients is estimated according to the singular value decomposition (SVD) described in Chap. 7.

8.3.1   The GW-ESIM Parameters

In the GW-ESIM context, the net genetic merit can be written as

$$H = \mathbf{w}'_1\mathbf{g} + \mathbf{w}'_2\mathbf{m} = \begin{bmatrix} \mathbf{w}'_1 & \mathbf{w}'_2 \end{bmatrix} \begin{bmatrix} \mathbf{g} \\ \mathbf{m} \end{bmatrix} = \mathbf{w}'\mathbf{x}, \qquad (8.24)$$

where $\mathbf{g}' = \begin{bmatrix} g_1 & \cdots & g_t \end{bmatrix}$ is the vector of true breeding values, t is the number of traits, $\mathbf{w}'_1 = \begin{bmatrix} w_1 & \cdots & w_t \end{bmatrix}$ is the vector of unknown economic weights associated with the breeding values; $\mathbf{w}'_2 = \begin{bmatrix} 0_1 & \cdots & 0_N \end{bmatrix}$ is a null vector associated with the vector of marker code values $\mathbf{m}' = \begin{bmatrix} m_1 & \cdots & m_N \end{bmatrix}$, where m_j ($j = 1, 2, \ldots,$ $N =$ number of markers) is the jth marker in the training population; $\mathbf{w}' = \begin{bmatrix} \mathbf{w}'_1 & \mathbf{w}'_2 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \mathbf{g}' & \mathbf{m}' \end{bmatrix}$. The GW-ESIM ($I$) index combines the phenotypic value and all the marker information of individuals to predict Eq. (8.24) values in each selection cycle and can be written as

$$I = \boldsymbol{\beta}'_y\mathbf{y} + \boldsymbol{\beta}'_m\mathbf{m} = \begin{bmatrix} \boldsymbol{\beta}'_y & \boldsymbol{\beta}'_m \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{m} \end{bmatrix} = \boldsymbol{\beta}'\mathbf{q}, \qquad (8.25)$$

where $\boldsymbol{\beta}'_y$ and $\boldsymbol{\beta}_m$ are vectors of phenotypic and marker weights respectively; $\mathbf{y}' = \begin{bmatrix} y_1 & \cdots & y_t \end{bmatrix}$ is the vector of phenotypic values; $\mathbf{m}$ was defined in Eq. (8.24); $\boldsymbol{\beta}' = \begin{bmatrix} \boldsymbol{\beta}'_y & \boldsymbol{\beta}'_m \end{bmatrix}$ and $\mathbf{q}' = \begin{bmatrix} \mathbf{y}' & \mathbf{m}' \end{bmatrix}$.

Let $\sigma_I^2 = \boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}$ and $\sigma_H^2 = \mathbf{w}'\mathbf{Z}\mathbf{w}$ be the variance of $I = \boldsymbol{\beta}'\mathbf{q}$ and $H = \mathbf{w}'\mathbf{z}$ respectively, and $\sigma_{HI} = \mathbf{w}'\mathbf{Z}\boldsymbol{\beta}$ the covariance between I and H, where $\mathbf{Q} = Var\begin{bmatrix} \mathbf{y} \\ \mathbf{m} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{G}'_M \\ \mathbf{G}_M & \mathbf{M} \end{bmatrix}$ and $\mathbf{X} = Var\begin{bmatrix} \mathbf{g} \\ \mathbf{m} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{G}'_M \\ \mathbf{G}_M & \mathbf{M} \end{bmatrix}$ are block matrices of size $(t + N) \times (t + N)$ (t is the number of traits and N is the number of markers) where $\mathbf{P} = Var(\mathbf{y})$, $\mathbf{M} = Var(\mathbf{m})$, $\mathbf{C} = Var(\mathbf{g})$, and $\mathbf{G}_M = cov\,(\mathbf{y}, \mathbf{m}) = cov\,(\mathbf{g}, \mathbf{m})$ are covariance matrices of phenotypic ($\mathbf{y}$), coded marker ($\mathbf{m}$), and genetic ($\mathbf{g}$) values respectively, whereas $\mathbf{G}_M$ is the covariance matrix between $\mathbf{y}$ and $\mathbf{m}$, and between $\mathbf{g}$ and $\mathbf{m}$ (for details see Chap. 4); $\mathbf{w}$ and $\boldsymbol{\beta}$ were defined earlier. Note that although the

size of matrices $\mathbf{P}$ and $\mathbf{C}$ are $t \times t$, the sizes of matrices $\mathbf{M}$ and $\mathbf{G}_M$ are $N \times N$ and $N \times t$ respectively. Thus, if the number of markers is very high, the size of matrices $\mathbf{M}$ and $\mathbf{G}_M$ could also be very high.

In Chap. 4 we described matrix $\mathbf{M}$ as

$$\mathbf{M} = \begin{bmatrix} 1 & (1 - 2\theta_{11}) & \dots & (1 - 2\theta_{1N}) \\ (1 - 2\theta_{21}) & 1 & \dots & (1 - 2\theta_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ (1 - 2\theta_{N1}) & (1 - 2\theta_{N2}) & \dots & 1 \end{bmatrix}, \tag{8.26}$$

where $(1 - 2\theta_{ij})$ and θ_{ij} ($i, j = 1, 2, \dots, N$= number of markers) are the covariance (or correlation) and the recombination frequency between the ith and jth marker respectively, whereas matrix $\mathbf{G}_M$ can be written as

$$\mathbf{G}_M = \begin{bmatrix} (1 - 2r_{11})\alpha_{11} & (1 - 2r_{11})\alpha_{12} & \dots & (1 - 2r_{1N})\alpha_{1N_Q} \\ (1 - 2r_{21})\alpha_{21} & (1 - 2r_{22})\alpha_{22} & \dots & (1 - 2r_{2N})\alpha_{2N_Q} \\ \vdots & \vdots & \ddots & \vdots \\ (1 - 2r_{t1})\alpha_{t1} & (1 - 2r_{N2})\alpha_{t2} & \dots & (1 - 2r_{NN})\alpha_{tN_Q} \end{bmatrix}, \tag{8.27}$$

where $(1 - 2r_{ik})\alpha_{qk}$ ($i = 1, 2, \dots, N$, $k = 1, 2, \dots, N_Q$ = number of quantitative trait loci (QTL), $q = 1, 2, \dots, t$) is the covariance between the qth trait and the ith marker; r_{ik} is the recombination frequency between the ith and kth QTL, and α_{qk} is the effect of the kth QTL over the qth trait.

Let $\rho_{HI} = \frac{\mathbf{w}'\mathbf{X}\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{X}\mathbf{w}}\sqrt{\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}}}$ be the correlation between $I = \boldsymbol{\beta}'\mathbf{q}$ and $H = \mathbf{w}'\mathbf{x}$; then, the GW-ESIM selection response can be written as

$$R = k_I \sigma_H \rho_{HI}, \tag{8.28}$$

where k_I is the standardized selection differential (or selection intensity) associated with GW-ESIM and $\sigma_H = \sqrt{\mathbf{w}'\mathbf{X}\mathbf{w}}$ is the standard deviation of the variance of H.

Assuming that k_I is fixed, and that matrices $\mathbf{Q}$ and $\mathbf{X}$ are known, we can maximize R (Eq. 8.28) by maximizing ρ_{HI} with respect to vectors $\mathbf{w}'$ and $\boldsymbol{\beta}$ under the restrictions $\sigma_H^2 = \mathbf{w}'\mathbf{X}\mathbf{w}$, $\sigma_I^2 = \boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}$, and $0 < \sigma_H^2, \sigma_I^2 < \infty$, similar to the MESIM and GESIM. It can be shown that vector $\mathbf{w}$ can be written as

$$\mathbf{w}_W = \mathbf{X}^{-1}\mathbf{Q}\boldsymbol{\beta} \tag{8.29}$$

and that $H_W = \mathbf{w}'_W\mathbf{x}$ is the net genetic merit in the GW-ESIM context. The correlation between $H_W = \mathbf{w}'_W\mathbf{x}$ and $I = \boldsymbol{\beta}'\mathbf{q}$ is $\rho_{H_WI} = \frac{\sqrt{\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}}}{\sqrt{\boldsymbol{\beta}'\mathbf{Q}\mathbf{X}^{-1}\mathbf{Q}\boldsymbol{\beta}}}$ and the GW-ESIM vector of coefficients ($\boldsymbol{\beta}$) that maximizes ρ_{H_WI} can be obtained from equation

$$\left(\mathbf{Q}^{-1}\mathbf{Z} - \lambda_W^2 \mathbf{I}_{(t+N)}\right)\boldsymbol{\beta}_W = \mathbf{0}, \tag{8.30}$$

where $\mathbf{I}_{(t + N)}$ is an identity matrix of size $(t + N) \times (t + N)$ and $I_W = \boldsymbol{\beta}'_W \mathbf{q}$ is the optimized GW-ESIM. The accuracy of the GW-ESIM can be written as

$$\rho_{H_W I_W} = \frac{\sqrt{\boldsymbol{\beta}'_W \mathbf{Q} \boldsymbol{\beta}_W}}{\sqrt{\boldsymbol{\beta}'_W \mathbf{Q} \mathbf{X}^{-1} \mathbf{Q} \boldsymbol{\beta}_W}} = \frac{\sigma_{I_W}}{\sigma_{H_W}}, \tag{8.31}$$

where $\sigma_{I_W} = \sqrt{\boldsymbol{\beta}'_W \mathbf{Q} \boldsymbol{\beta}_W}$ is the standard deviation of $I_W = \boldsymbol{\beta}'_W \mathbf{q}$, and $\sigma_{H_W} = \sqrt{\boldsymbol{\beta}'_W \mathbf{Q} \mathbf{X}^{-1} \mathbf{Q} \boldsymbol{\beta}_W}$ is the standard deviation of $H_W = \mathbf{w}'_W \mathbf{x}$. In Eq. (8.30) $\lambda_W^2 = \rho_{H_W I_W}^2$ is the square of the canonical correlation between H_W and I_W.

The maximized GW-ESIM selection response and expected genetic gain per trait are

$$R_W = k_I \sqrt{\boldsymbol{\beta}'_W \mathbf{Q} \boldsymbol{\beta}_W} \tag{8.32}$$

and

$$\mathbf{E}_W = k_1 \frac{\mathbf{X} \boldsymbol{\beta}_W}{\sqrt{\boldsymbol{\beta}'_W \mathbf{Q} \boldsymbol{\beta}_W}}, \tag{8.33}$$

respectively, where $\boldsymbol{\beta}_W$ is the first eigenvector of Eq. (8.30).

8.3.2   Estimating GW-ESIM Parameters

In Chap. 2, Eqs. (2.22) to (2.24), we described the restricted maximum likelihood methods to estimate matrices $\mathbf{C}$ and $\mathbf{P}$, which can be denoted by $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{P}}$. In Chap. 4, we described how to estimate matrices $\mathbf{M}$ and $\mathbf{G}_M$, which can be denoted by $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{G}}_M$. With these estimates, we constructed the block estimated matrices as $\widehat{\mathbf{Q}} = \begin{bmatrix} \widehat{\mathbf{P}} & \widehat{\mathbf{G}}'_M \\ \widehat{\mathbf{G}}_M & \widehat{\mathbf{M}} \end{bmatrix}$ and $\widehat{\mathbf{X}} = \begin{bmatrix} \widehat{\mathbf{C}} & \widehat{\mathbf{G}}'_M \\ \widehat{\mathbf{G}}_M & \widehat{\mathbf{M}} \end{bmatrix}$, whence we obtained the equation

$$\left(\widehat{\mathbf{Q}}^- \widehat{\mathbf{X}} - \widehat{\lambda}_{Wj}^2 \mathbf{I} \right) \widehat{\boldsymbol{\beta}}_{Wj} = \mathbf{0}, \tag{8.34}$$

$j = 1, 2, \ldots, (t + N)$, where $(t + N)$ is the number of traits and markers in the GW-ESIM index. Similar to the MESIM, we obtained estimators $\widehat{\boldsymbol{\beta}}_{W_1}$ and $\widehat{\lambda}_{W_1}^2$ of the first eigenvector $\boldsymbol{\beta}_{W_1}$ and the first eigenvalue $\widehat{\lambda}_{W_1}^2$ respectively, from equation

$$\left(\widehat{\mathbf{E}} \widehat{\mathbf{E}}' - \widehat{\mu}_j \mathbf{I} \right) \widehat{\boldsymbol{\beta}}_{W_j} = \mathbf{0}, \tag{8.35}$$

where $\widehat{\mathbf{E}} = \widehat{\mathbf{Q}}^{-}\widehat{\mathbf{X}}$ and $\widehat{\mu}_j = \widehat{\lambda}^4_{W_j}$. These results allow the GW-ESIM index selection response and its expected genetic gain per trait to be estimated as $\widehat{I}_W = \widehat{\boldsymbol{\beta}}'_{W_1}\widehat{\mathbf{q}}$, $\widehat{R}_W = k_I\sqrt{\widehat{\boldsymbol{\beta}}'_{W_1}\widehat{\mathbf{Q}}\boldsymbol{\beta}'_{W_1}}$ and $\widehat{\mathbf{E}}_w = k_I \dfrac{\widehat{\mathbf{X}\boldsymbol{\beta}}'_{W_1}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{W_1}\widehat{\mathbf{Q}}\boldsymbol{\beta}'_{W_1}}}$ respectively, whereas the estimator of GW-ESIM accuracy is $\widehat{\lambda}_{W_1}$.

8.3.3 Numerical Examples

We compare the estimated GW-LMSI and GW-ESIM selection responses using the simulated data set described in Sect. 2.8.1 of Chap. 2, with a selection intensity of 10% ($k_I = 1.755$). Table 8.1 presents the estimated GW-LMSI selection response for four simulated selection cycles when their vectors of coefficients are normalized, whence it can be seen that the average estimated GW-LMSI selection response was 0.87. Table 8.1 also presents the estimated GW-ESIM selection response for four simulated selection cycles; the average of the estimated GW-ESIM selection responses was 0.93. Thus, for this data set, the estimated GW-LMSI and selection responses were very similar.

8.4 The Restricted Linear Genomic Eigen Selection Index Method

The restricted linear genomic eigen selection index method (RGESIM) is based on the restricted linear phenotypic ESIM (RESIM) theory described in Chap. 7. In the RESIM, the breeder's objective is to improve only $(t - r)$ of t ($r < t$) traits, leaving r of them fixed. The same is true for RGESIM, but in this case, we should impose $2r$ restrictions, i.e., we need to fix r traits and their associated r GEBV to obtain results similar to those obtained with the RESIM (see Chap. 7 for details). This is the main difference between the RGESIM and the RESIM.

It can be shown that $Cov(I, \boldsymbol{\alpha}) = \mathbf{A}\boldsymbol{\beta}$ is the covariance between the breeding value vector ($\boldsymbol{\alpha}' = [\mathbf{g}' \quad \boldsymbol{\gamma}']$) and the GESIM index ($I = \boldsymbol{\beta}'\mathbf{f}$). In the RGESIM, we want some covariances between the linear combinations of $\boldsymbol{\alpha}$ ($\mathbf{U}'_G\boldsymbol{\alpha}$) and $I = \boldsymbol{\beta}'\mathbf{f}$ to be zero, i.e., $Cov(\mathbf{I}_G, \mathbf{U}'_G\boldsymbol{\alpha}) = \mathbf{U}'_G\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{U}'_G$ is a matrix $2(t - 1) \times 2t$ of 1s and 0s (1 indicates that the trait and its associated GEBV are restricted, and 0 indicates that the trait and its GEBV have no restrictions). We can solve this problem by maximizing $\dfrac{\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}}}$ with respect to vector $\boldsymbol{\beta}$ under the restriction $\mathbf{U}'_G\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ and $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ similar to the RESIM, or by maximizing the correlation between $H = \mathbf{w}'\boldsymbol{\alpha}$ and

$I = \boldsymbol{\beta}'\mathbf{f}$, $\rho_{HI} = \dfrac{\mathbf{w}'\mathbf{A}\boldsymbol{\beta}}{\sqrt{\mathbf{w}'\mathbf{A}\mathbf{w}}\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}}}$, with respect to vectors $\mathbf{w}'$ and $\boldsymbol{\beta}$ under the restrictions $\mathbf{U}'_G\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$, $\sigma_H^2 = \mathbf{w}'\mathbf{A}\mathbf{w}$, $\sigma_I^2 = \boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}$ and $0 < \sigma_H^2, \sigma_I^2 < \infty$, as we did for the GESIM.

8.4.1   The RGESIM Parameters

To obtain the RGESIM vector of coefficients, we maximize the function

$$f(\boldsymbol{\beta}, \mathbf{v}') = \frac{\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}}} - \mathbf{v}'\mathbf{U}'_G\mathbf{A}\boldsymbol{\beta} \tag{8.36}$$

with respect to $\boldsymbol{\beta}$ and $\mathbf{v}'$, where $\mathbf{v}' = [v_1 \ v_2 \ \cdots \ v_{2(r-1)}]$ is a vector of Lagrange multipliers. The derivatives of function $f(\boldsymbol{\beta}, \mathbf{v}')$ with respect to $\boldsymbol{\beta}$ and $\mathbf{v}'$ can be written as

$$2(\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta})^{1/2}\mathbf{A}\boldsymbol{\beta} - (\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta})^{-1/2}(\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta})\boldsymbol{\Phi}\boldsymbol{\beta} - \mathbf{A}\mathbf{U}_G\mathbf{v} = \mathbf{0}, \tag{8.37}$$
$$\mathbf{U}'_G\mathbf{A}\boldsymbol{\beta} = \mathbf{0}, \tag{8.38}$$

respectively, where Eq. (8.38) denotes the restriction imposed for maximizing Eq. (8.36). Using algebraic methods on Eq. (8.37), we get

$$\left(\mathbf{K}_{RG}\boldsymbol{\Phi}^{-1}\mathbf{A} - \lambda_{RG}^2\mathbf{I}_{2t}\right)\boldsymbol{\beta}_{RG} = \mathbf{0}, \tag{8.39}$$

where $\lambda_{RG}^2 = h_{I_{RG}}^2$, $h_{I_{RG}}^2$ is the RGESIM heritability obtained under the restriction $\mathbf{U}'_G$ $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$; $\mathbf{K}_{RG} = [\mathbf{I}_{2t} - \mathbf{Q}_{RG}]$, $\mathbf{I}_{2t}$ is an identity matrix of size $2t \times 2t$, and $\mathbf{Q}_{RG} = \boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\left(\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\right)^{-1}\mathbf{U}'_G\mathbf{A}$. When $\mathbf{U}'_G$ is a null matrix, $\boldsymbol{\beta}'_{RG} = \boldsymbol{\beta}'_G$ (the vector of the GESIM coefficients); thus, the RGESIM is more general than the GESIM and includes the GESIM as a particular case. The RGESIM index $I_{GR} = \boldsymbol{\beta}'_{RG}\mathbf{y}$ and its selection response and expected genetic gain per trait use the first eigenvector of matrix $\mathbf{K}_G\boldsymbol{\Phi}^{-1}\mathbf{A}$. It can be shown that the vector of coefficients of $H = \mathbf{w}'_{RG}\boldsymbol{\alpha}$ in the RGESIM can be written as

$$\mathbf{w}_{RG} = \mathbf{A}^{-1}\left[\boldsymbol{\Phi} + \mathbf{Q}'_{RG}\mathbf{A}\right]\boldsymbol{\beta}_{RG}, \tag{8.40}$$

where $\mathbf{Q}'_{RG} = \mathbf{A}\mathbf{U}_G\left(\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\right)^{-1}\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}$.

Note that the restriction $\mathbf{U}'_G\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ can be written as $\boldsymbol{\beta}'\mathbf{A}\mathbf{U}_G = \mathbf{0}$; this means that $\boldsymbol{\beta}'\mathbf{Q}'_{RG} = \mathbf{0}$ and that the covariance between $H_{RG} = \mathbf{w}'_{RG}\boldsymbol{\alpha}$ and $I_{RG} = \boldsymbol{\beta}'_{RG}\mathbf{f}$ $(\sigma_{H_{RG}I_{RG}})$ can be written as

$$\sigma_{H_{RG}I_{RG}} = \mathbf{w}'_{RG}\mathbf{A}\boldsymbol{\beta}'_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\Phi}\boldsymbol{\beta}_{RG} + \boldsymbol{\beta}'_{RG}\mathbf{Q}'_{RG}\mathbf{C}\boldsymbol{\beta}_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\Phi}\boldsymbol{\beta}_{RG}. \tag{8.41}$$

Equation (8.41) indicates that $\sigma_{H_{RG}I_{RG}}$ is equal to the variance of $I_{RG} = \boldsymbol{\beta}'_{RG}\mathbf{f}$ ($\sigma^2_{I_{RG}} = \boldsymbol{\beta}'_{RG}\boldsymbol{\Phi}\boldsymbol{\beta}_{RG}$); therefore, the maximized correlation between I_{RG} and H_{RG} or RGESIM accuracy can be written as

$$\rho_{H_{RG}I_{RG}} = \frac{\sqrt{\boldsymbol{\beta}'_{RG}\boldsymbol{\Phi}\boldsymbol{\beta}_{RG}}}{\sqrt{\mathbf{w}'_{RG}\mathbf{A}\mathbf{w}_{RG}}}, \tag{8.42}$$

where $\mathbf{w}'_{RG}\mathbf{A}\mathbf{w}_{RG}$ is the variance of H_{RG}. Hereafter, to simplify the notation, we write Eq. (8.42) as λ_{RG}.

The maximized selection response and the expected genetic gain per trait of the RGESIM are

$$R_{RG} = k_I\sqrt{\boldsymbol{\beta}'_{RG}\boldsymbol{\Phi}\boldsymbol{\beta}_{RG}} \tag{8.43}$$

and

$$\mathbf{E}_{RG} = k_I\frac{\mathbf{A}\boldsymbol{\beta}_{RG}}{\sqrt{\boldsymbol{\beta}'_{RG}\boldsymbol{\Phi}\boldsymbol{\beta}_{RG}}}, \tag{8.44}$$

respectively, where $\boldsymbol{\beta}_{RG}$ is the first eigenvector of matrix $\mathbf{K}_{RG}\boldsymbol{\Phi}^{-1}\mathbf{A}$.

8.4.2   Estimating RGESIM Parameters

In Sect. 8.2, we indicated how to estimate matrices $\mathbf{P}$, $\boldsymbol{\Gamma}$, and $\mathbf{C}$ using phenotypic and genomic information, whence we can estimate matrices $\mathbf{A} = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ and $\boldsymbol{\Phi} = \begin{bmatrix} \mathbf{P} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$. Those methods are also valid for the RGESIM. This means that the SVD methods described for estimating MESIM parameters are also valid for estimating RGESIM parameters.

8.4.3   Numerical Examples

With a selection intensity of 10% ($k_I = 1.755$), we compare the CRLGSI (for details see Chap. 6) versus the RGESIM theoretical results using a real maize (*Zea mays*) F_2 population with 244 genotypes (each with two repetitions), 233 molecular markers, and three traits—GY (ton ha^{-1}), EHT (cm), and PHT (cm)—described in Sect. 8.2.2, where $\widehat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}$, $\widehat{\mathbf{C}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$,

and $\widehat{\mathbf{\Gamma}} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}$ were the estimated matrices of $\mathbf{P}$, $\mathbf{C}$, and $\mathbf{\Gamma}$ respectively.

We have indicated that the main difference between the RLPSI and the CRLGSI is the matrix $\mathbf{U}'_C$, on which we now need to impose two restrictions: one for the trait and another for its associated GEBV. Consider the data set described earlier and suppose that we restrict the trait GY (ton ha^{-1}) and its associated GEBV$_{\text{GY}}$; then, matrix $\mathbf{U}'_C$ should be constructed as $\mathbf{U}'_{C1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$. If we restrict traits GY and EHT (cm) and their associated GEBV$_{\text{GY}}$ and GEBV$_{\text{EHT}}$, matrix $\mathbf{U}'_C$ should be constructed as $\mathbf{U}'_{C2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$, etc. The procedure for obtaining matrices $\widehat{\mathbf{K}}_{RG} = \left[\mathbf{I}_{2t} - \widehat{\mathbf{Q}}_{RG}\right]$ and $\widehat{\mathbf{Q}}_{RG} = \widehat{\mathbf{\Phi}}^{-1}\widehat{\mathbf{A}}\mathbf{U}_G\left(\mathbf{U}'_G\widehat{\mathbf{A}}\widehat{\mathbf{\Phi}}^{-1}\widehat{\mathbf{A}}\mathbf{U}_G\right)^{-1}\mathbf{U}'_G$ $\widehat{\mathbf{A}}$ was described in Chap. 6, and is also valid for estimating RGESIM parameters.

The estimated CRLGSI vector of coefficients is $\widehat{\mathbf{\beta}}_{CR} = \widehat{\mathbf{K}}_{RG}\widehat{\mathbf{\beta}}$, where $\widehat{\mathbf{\beta}} = \widehat{\mathbf{\Phi}}^{-1}\widehat{\mathbf{A}}$ $\mathbf{w}$ is the estimated CLGSI vector of coefficients (Chap. 6). Let $\mathbf{w}' = [5 \ -0.1 \ -0.1 \ 0 \ 0 \ 0]$ be the vector of economic weights and suppose that we restrict trait GY and its associated GEBV$_{\text{GY}}$; in this case, $\mathbf{U}'_{C1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$, and according to matrices $\widehat{\mathbf{P}}$, $\widehat{\mathbf{C}}$, and $\widehat{\mathbf{\Gamma}}$ described earlier, $\widehat{\mathbf{\beta}}'_{CR} = [0.076 \ -0.004 \ -0.018 \ 2.353 \ -0.096 \ -0.082]$ was the estimated CRLGSI vector of coefficients and the estimated CRLGSI was

$$\widehat{I}_{CR} = 0.076\text{GY} - 0.004\text{EHT} - 0.018\text{PHT} + 2.353\text{GEBV}_{\text{GY}} - 0.096\text{GEBV}_{\text{EHT}} \\ - 0.082\text{GEBV}_{\text{PHT}}$$

where GEBV$_{\text{GY}}$, GEBV$_{\text{EHT}}$, and GEBV$_{\text{PHT}}$ are the GEBVs associated with the traits GY, EHT, and PHT respectively. The same procedure is valid for two or more restrictions.

The estimated CRLGSI selection response and expected genetic gain per trait were $\widehat{R}_{CR} = k_I\sqrt{\widehat{\mathbf{\beta}}'_{CR}\widehat{\mathbf{\Phi}}\widehat{\mathbf{\beta}}_{CR}} = 0.96$ and $\widehat{\mathbf{E}}'_{CR} = k_I = \dfrac{\widehat{\mathbf{\beta}}'_{CR}\widehat{\mathbf{A}}}{\sqrt{\widehat{\mathbf{\beta}}'_{CR}\widehat{\mathbf{\Phi}}\widehat{\mathbf{\beta}}_{CR}}}$ $[0 \ -3.53 \ -6.03 \ 0 \ -2.93 \ -4.87]$ respectively, whereas the estimated CRLGSI accuracy was $\widehat{\rho}_{HI_{CR}} = \dfrac{\widehat{\sigma}_{I_{CR}}}{\widehat{\sigma}_H} = 0.51$. Note that in $\widehat{\mathbf{E}}'_{CR}$, the trait GY and its associated GEBV$_{\text{GY}}$ have null values, as we would expect.

The estimated RGESIM vector of coefficients was $\widehat{\mathbf{\beta}}'_{CR} = [0.015 \ -0.001 \ -0.004 \ 0.998 \ -0.029 \ -0.045]$, and the estimated RGESIM index was $\widehat{I}_{RG} = 0.015\text{GY} - 0.001\text{EHT} - 0.004\text{PHT} + 0.998\text{GEBV}_{\text{GY}}$ $-0.029\text{GEBV}_{\text{EHT}} - 0.045\text{GEBV}_{\text{PHT}}$ where GEBV$_{\text{GY}}$, GEBV$_{\text{EHT}}$, and GEBV$_{\text{PHT}}$

are the GEBVs associated with traits GY, EHT, and PHT respectively. The same procedure is valid for two or more restrictions.

The estimated RGESIM selection response and expected genetic gain per

trait were $\quad \widehat{R}_{RG} = k_I \sqrt{\widehat{\boldsymbol{\beta}}'_{RG} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{RG}} = 0.37 \quad$ and $\quad \widehat{\mathbf{E}}'_{RG} = k_I = \dfrac{\widehat{\boldsymbol{\beta}}'_{RG} \widehat{\mathbf{A}}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{RG} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{RG}}}$

$[0 \quad -3.28 \quad -6.03 \quad 0 \quad -2.93 \quad -5.40]$ respectively, whereas the estimated

RGESIM accuracy was $\widehat{\rho}_{\widehat{H}_{RG}\widehat{I}_{RG}} = \dfrac{\widehat{\sigma}_{\widehat{I}_{RG}}}{\widehat{\sigma}_{\widehat{H}_{RG}}} = 0.86.$

Fig. 8.3 presents the frequency distribution of the 244 estimated RGESIM index values for two null restrictions on traits GY and EHT and their associated GEBV$_{GY}$ and GEBV$_{EHT}$, for one selection cycle in an environment for a real maize (*Zea mays*) F$_2$ population with 233 molecular markers. Note that the frequency distribution of the estimated RGESIM index values approaches the normal distribution.

Now we compare the estimated CRLGSI and RGESIM selection responses and expected genetic gains per trait using the simulated data set described in Sect. 2.8.1 of Chap. 2. We used that data set for four phenotypic selection cycles (C2, C3, C4, and C5), each with four traits (T_1, T_2, T_3, and T_4), 500 genotypes, and four replicates per genotype. The economic weights for T_1, T_2, T_3, and T_4 were 1, -1, 1, and
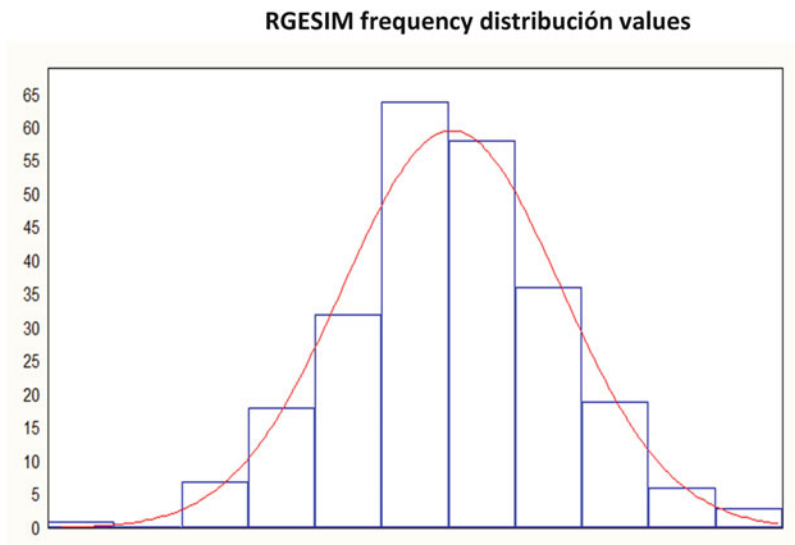

RGESIM frequency distribución values

Fig. 8.3 Frequency distribution of the 244 estimated restricted genomic eigen selection index method (RGESIM) values for two null restrictions on traits grain yield (GY) and EHT and their associated genomic estimated breeding values (GEBVs), GEBV$_{GY}$ and GEBV$_{EHT}$ respectively, for one selection cycle in an environment for a real maize (*Zea mays*) F$_2$ population with 233 molecular markers. Note that the frequency distribution of the estimated RGESIM index values approaches normal distribution

1 respectively. For this data set, matrix $\mathbf{F}$ was an identity matrix of size 8×8 for all four selection cycles.

Columns 2, 3, and 4 (from left to right) of Table 8.4 present the estimated CRLGSI selection responses when their vectors of coefficients are normalized and the estimated RGESIM and selection responses for one, two, and three restrictions for four simulated selection cycles. The averages of the estimated CRLGSI selection responses of the traits and their associated GEBVs for each of the three null restrictions were 3.24 for one restriction, 4.08 for two restrictions, and 5.06 for three restrictions, whereas the averages of the estimated RGESIM selection responses were 3.08 for one restriction, 2.79 for two restrictions, and 3.23 for three restrictions. Note that although for one restriction the selection response was similar for both indices, for two and three restrictions the CRLGSI selection responses were greater than the RGESIM selection responses.

Table 8.5 presents the estimated CRLGSI and RGESIM expected genetic gains per trait for four traits (T1, T2, T3, and T4) and their associated GEBVs (in this case denoted by G1, G2, G3, and G4 to simplify the notation) in four simulated selection cycles and for one, two, and three null restrictions in four simulated selection cycles. Note that the null values of the traits and their restricted GEBVs are not shown in Table 8.5 with the aim of simplifying the table. The averages of the estimated CRLGSI expected genetic gains for the three traits and their associated GEBVs were -2.60, 2.16, 2.84, -1.21, 0.67, and 1.02 for one restriction; 2.74, 3.23, 0.78,

Table 8.4 Estimated combined null restricted linear genomic selection index (CRLGSI) and estimated combined predetermined proportional gain linear genomic selection index (CPPG-LGSI) selection responses for one, two, and three restrictions when their vectors of coefficients are normalized for four simulated selection cycles

	CRLGSI response for one, two and three null restrictions			CPPG-LGSI response for one, two and three predetermined restrictions		
Cycle	1	2	3	1	2	3
1	3.25	4.09	4.89	5.36	2.80	1.81
2	3.28	4.19	5.21	5.07	3.64	1.99
3	2.91	3.89	4.97	5.37	3.86	1.42
4	3.53	4.17	5.15	4.52	3.38	1.20
Average	3.24	4.08	5.06	5.08	3.42	1.60
	RGESIM response for one, two, and three null restrictions			PPG-GESIM response for one, two, and three predetermined restrictions		
Cycle	1	2	3	1	2	3
1	3.21	2.78	3.47	1.95	4.07	4.26
2	3.11	2.86	3.06	1.85	4.12	5.49
3	2.93	2.76	3.20	2.04	4.18	6.30
4	3.07	2.76	3.21	2.02	4.17	5.82
Average	3.08	2.79	3.23	1.96	4.14	5.47

Estimated null restricted genomic eigen selection index method (RGESIM) and predetermined proportional gain genomic eigen selection index method (PPG-GESIM) selection responses for one, two, and three restrictions for four simulated selection cycles. The selection intensity was 10% ($k_I = 1.755$)

Table 8.5 Estimated CRLGSI and estimated null RGESIM expected genetic gains per trait for four traits (T1, T2, T3, and T4) and their associated genomic estimated breeding values (G1, G2, G3, and G4) for four simulated selection cycles and for one, two, and three null restrictions for four simulated selection cycles. The selection intensity was 10% ($k_I = 1.755$)

| | CRLGSI expected genetic gains for one, two and three null restrictions | | | | | | | | | | Three restrictions[c] | |
| | One restriction[a] | | | | | | Two restrictions[b] | | | | | |
Cycle	T2	T3	T4	G2	G3	G4	T3	T4	G3	G4	T4	G4
1	−2.32	2.17	2.87	−1.48	0.73	1.24	2.60	3.38	0.86	1.15	4.08	1.50
2	−2.76	2.14	2.89	−1.19	0.76	0.96	2.81	3.30	0.87	0.98	3.95	1.25
3	−2.22	2.27	2.98	−1.15	0.62	0.97	2.77	3.14	0.69	0.90	3.93	1.33
4	−3.09	2.08	2.64	−1.05	0.58	0.92	2.80	3.08	0.70	0.93	4.13	1.24
Mean	−2.60	2.16	2.84	−1.21	0.67	1.02	2.74	3.23	0.78	0.99	4.02	1.33
	RGESIM expected genetic gains for one, two and three null restrictions										Three restrictions[c]	
	One restriction[a]						Two restrictions[b]					
Cycle	T2	T3	T4	G2	G3	G4	T3	T4	G3	G4	T4	G4
1	3.27	−1.52	−1.24	2.48	−0.88	−1.00	3.18	0.93	1.88	0.43	3.66	2.21
2	3.30	−1.79	−1.41	2.10	−1.09	−0.82	3.26	1.34	1.82	0.66	3.41	2.00
3	2.98	−1.62	−1.44	2.13	−0.83	−0.75	3.31	0.86	1.70	0.21	3.45	2.05
4	3.56	−1.73	−1.23	1.92	−0.89	−0.78	3.40	0.96	1.62	0.53	3.58	2.02
Mean	3.27	−1.67	−1.33	2.16	−0.92	−0.84	3.29	1.02	1.76	0.46	3.53	2.07

[a]All T1 and G1 expected genetic gains were null
[b]All T1, T2, G1, and G2 expected genetic gains were null
[c]All T1, T2, T3, G1, G2, and G3 expected genetic gains were null

and 0.99 for two restrictions; and 4.02 and 1.33 for three restrictions. On the other hand, the averages of the estimated RGESIM expected genetic gains for the three traits and their associated GEBVs were 3.27, −1.67, −1.33, 2.16, −0.92, and −0.84 for one restriction; 3.29, 1.02, 1.76, and 0.46 for two restrictions; and 3.53 and 2.07 for three restrictions. These results indicate that in terms of absolute values, the estimated expected genetic gains for the traits and their associated GEBVs were similar for both indices.

8.5   The Predetermined Proportional Gain Linear Genomic Eigen Selection Index Method

The predetermined proportional gain linear genomic eigen selection index method (PPG-GESIM) theory is based on the predetermined proportional gain linear phenotypic ESIM (PPG-ESIM) described in Chap. 7. In the PPG-ESIM, the vector of PPG (predetermined proportional gain) imposed by the breeder was $\mathbf{d}' = \begin{bmatrix} d_1 & d_2 & \cdots & d_r \end{bmatrix}$. However, because the PPG-GESIM uses phenotypic and GEBV information jointly to predict the net genetic merit, the vector of PPG

imposed by the breeder ($\mathbf{d}_{PG}$) should be twice the standard vector $\mathbf{d}'$, that is, $\mathbf{d}'_{PG} = \begin{bmatrix} d_1 & d_2 & \cdots & d_r & d_{r+1} & d_{r+2} & \cdots & d_{2r} \end{bmatrix}$, where we would expect that if d_1 is the PPG imposed on trait 1, then d_{r+1} should be the PPG imposed on the GEBV associated with trait 1, etc. Thus, in the PPG-GESIM we have three possible options for determining (for each trait and GEBV) the PPG: e.g., for trait 1, $d_1 = d_{r+1}$, $d_1 > d_{r+1}$ or $d_1 < d_{r+1}$. This is the main difference between the standard PPG-ESIM described in Chap. 7 and the PPG-GESIM.

8.5.1 The PPG-GESIM Parameters

Using the same procedure described for RGESIM and PPG-ESIM, the PPG-GESIM vector of coefficients ($\boldsymbol{\beta}_{PG}$), which maximizes the PPG-GESIM selection response and the expected genetic gain per trait, is the first eigenvector of the following equation

$$\left(\mathbf{T}_{PG} - \lambda_{PG}^2 \mathbf{I}_{2t}\right)\boldsymbol{\beta}_{PG} = \mathbf{0}, \tag{8.45}$$

where $\mathbf{T}_{PG} = \mathbf{K}_{RG}\boldsymbol{\Phi}^{-1}\mathbf{A} + \mathbf{B}$, $\mathbf{K}_{PG} = [\mathbf{I}_{2t} - \mathbf{Q}_{RG}]$, $\mathbf{I}_{2t}$ is an identity matrix of size $2t \times 2t$, $\mathbf{Q}_{RG} = \boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\left(\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\right)^{-1}\mathbf{U}'_G\mathbf{A}$, $\mathbf{B} = \boldsymbol{\delta}\varphi'$, $\boldsymbol{\delta} = \boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\left(\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\right)^{-1}\mathbf{d}_{PG}$, and $\varphi' = \dfrac{\mathbf{d}'_{PG}\left(\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\right)^{-1}\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}}{\mathbf{d}'_{PG}\left(\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\right)^{-1}\mathbf{d}_{PG}}$.

When $\mathbf{B}$ is a null matrix, $\mathbf{T}_{PG} = \mathbf{K}_{RG}\boldsymbol{\Phi}^{-1}\mathbf{A}$ (matrix of the RGESIM), and when $\mathbf{U}'_G$ is a null matrix, $\mathbf{T}_{PG} = \boldsymbol{\Phi}^{-1}\mathbf{A}$ (matrix of the GESIM); this means that the PPG-GESIM includes the RGESIM and GESIM as particular cases. The optimized PPG-GESIM index can be written as $I_{PG} = \boldsymbol{\beta}'_{PG}\mathbf{f}$.

The vector of coefficients of $H = \mathbf{w}'_{PG}\boldsymbol{\alpha}$ in the PPG-GESIM can be written as

$$\mathbf{w}_{PG} = \mathbf{A}^{-1}\left[\boldsymbol{\Phi} + \mathbf{Q}'_{PG}\mathbf{A}\right]\boldsymbol{\beta}_{PG}, \tag{8.46}$$

where $\mathbf{Q}'_{PG} = \mathbf{A}\mathbf{U}_G\mathbf{D}_G\left(\mathbf{D}'_G\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}\mathbf{A}\mathbf{U}_G\mathbf{D}_G\right)^{-1}\mathbf{D}'_G\mathbf{U}'_G\mathbf{A}\boldsymbol{\Phi}^{-1}$, and $\mathbf{D}'_G = \begin{bmatrix} d_{2r} & 0 & \cdots & 0 & -d_1 \\ 0 & d_{2r} & \cdots & 0 & -d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_{2r} & -d_{2r-1} \end{bmatrix}$. Similar to RGESIM, it can be shown that

the covariance between $H_{RG} = \mathbf{w}'_{PG}\boldsymbol{\alpha}$ and $I_{PG} = \boldsymbol{\beta}'_{PG}\mathbf{f}$ ($\sigma_{H_{PG}I_{PG}}$) is equal to the variance of $I_{PG} = \boldsymbol{\beta}'_{PG}\mathbf{f}$ ($\sigma^2_{I_{PG}} = \boldsymbol{\beta}'_{PG}\boldsymbol{\Phi}\boldsymbol{\beta}_{PG}$), that is, $\sigma_{H_{PG}I_{PG}} = \mathbf{w}'_{PG}\mathbf{A}\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}'_{PG}\boldsymbol{\Phi}\boldsymbol{\beta}_{PG} = \sigma^2_{I_{PG}}$.

The maximized correlation between I_{PG} and H_{PG}, or PPG-GESIM accuracy, is

$$\rho_{H_{PG}I_{PG}} = \frac{\sqrt{\boldsymbol{\beta}'_{PG}\boldsymbol{\Phi}\boldsymbol{\beta}_{PG}}}{\sqrt{\mathbf{w}'_{PG}\mathbf{A}\mathbf{w}_{PG}}} \tag{8.47}$$

where $\mathbf{w}'_{PG}\mathbf{A}\mathbf{w}_{PG}$ is the variance of H_{PG}. Hereafter, to simplify the notation, we write Eq. (8.47) as λ_{PG}.

The maximized selection response and the expected genetic gain per trait of the PPG-GESIM are

$$R_{PG} = k_I \sqrt{\boldsymbol{\beta}'_{PG}\boldsymbol{\Phi}\boldsymbol{\beta}_{PG}} \tag{8.48}$$

and

$$\mathbf{E}_{PG} = k_I \frac{\mathbf{A}\boldsymbol{\beta}_{PG}}{\sqrt{\boldsymbol{\beta}'_{PG}\boldsymbol{\Phi}\boldsymbol{\beta}_{PG}}}, \tag{8.49}$$

respectively, where $\boldsymbol{\beta}_{PG}$ is the first eigenvector of Eq. (8.45).

8.5.2 Numerical Examples

The process for estimating PPG-ESIM parameters is similar to the method described for estimating RGESIM parameters. With a selection intensity of 10% ($k_I = 1.755$), we compare the combined predetermined proportional gain linear genomic selection index (CPPG-LGSI) and PPG-GESIM results using the real maize (*Zea mays*) F_2 population with 244 genotypes, 233 molecular markers, and three traits—GY (ton ha^{-1}), EHT (cm), and PHT (cm)—where $\widehat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}$, $\widehat{\mathbf{G}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$ and $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}$ are the estimated matrices of $\mathbf{P}$, $\mathbf{G}$, and $\boldsymbol{\Gamma}$ respectively, whereas $\mathbf{w}' = \begin{bmatrix} 5 & -0.1 & -0.1 & 0 & 0 & 0 \end{bmatrix}$ was the vector of economic weights.

The estimated CPPG-LGSI vector of coefficients was $\widehat{\boldsymbol{\beta}}_{CP} = \widehat{\boldsymbol{\beta}}_{CG} + \widehat{\theta}_{CP}\widehat{\boldsymbol{\delta}}$ (see Chap. 6 for additional details). Let $\widehat{\mathbf{A}} = \begin{bmatrix} \widehat{\mathbf{G}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$ and $\widehat{\boldsymbol{\Phi}} = \begin{bmatrix} \widehat{\mathbf{P}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$ be the estimated block matrices and $\mathbf{d}'_{PG} = \begin{bmatrix} 7 & -3 & 3.5 & -1.5 \end{bmatrix}$ the vector of PPG imposed by the breeder on the traits GY and EHT, and their associated genomic estimated breeding values (GEBV$_{GY}$ and GEBV$_{EHT}$), and let

$$\mathbf{U}'_C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$ be the matrix of null restrictions on the CPPG-LGSI

and $\mathbf{w}' = [5 \;\; -0.1 \;\; -0.1 \;\; 0 \;\; 0 \;\; 0]$ the vector of economic weights. It can be shown that $\widehat{\theta}_{CP} = -0.00009$ is the estimated value of the proportionality constant, $\widehat{\boldsymbol{\delta}}' = [-112.92 \;\; -72.16 \;\; 61.35 \;\; 231.79 \;\; 64.75 \;\; -61.35]$, $\widehat{\boldsymbol{\beta}}'_{CP} = [-0.01 \;\; 0.01 \;\; -0.01 \;\; 0.59 \;\; 0.09 \;\; -0.09]$ is the estimated CPPG-LGSI vector of coefficients, and the estimated CPPG-LGSI can be written as

$$\widehat{I}_{CP} = -0.01\text{GY} + 0.01\text{EHT} - 0.01\text{PHT} + 0.59\text{GEBV}_{\text{GY}} + 0.09\text{GEBV}_{\text{EHT}}$$
$$- 0.09\text{GEBV}_{\text{PHT}}$$

where GEBV_{GY}, GEBV_{EHT}, and GEBV_{PHT} are the GEBVs associated with traits GY, EHT, and PHT respectively. The same procedure is valid for more than two predetermined restrictions. The estimated CPPG-LGSI selection response and expected genetic gain per trait were $\widehat{R}_{CP} = k_I \sqrt{\widehat{\boldsymbol{\beta}}'_{CP} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{CP}} = 0.443$ and

$$\widehat{\mathbf{E}}'_{CP} = k_I \frac{\widehat{\boldsymbol{\beta}}'_{CP} \widehat{\mathbf{A}}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{CP} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{CP}}} = [-0.004 \;\; 0.002 \;\; -4.639 \;\; -0.002 \;\; 0.001 \;\; -4.326]$$

respectively, whereas the estimated CPPG-LGSI accuracy is $\widehat{\rho}_{HI_{CP}} = \dfrac{\widehat{\sigma}_{I_{CP}}}{\widehat{\sigma}_H} = 0.234$.

Because the estimated value of the proportionality constant was negative ($\widehat{\theta}_{CP} = -0.00009$), the expected genetic gains of the traits GY and EHT, and their associated genomic estimated breeding values (GEBV_{GY} and GEBV_{EHT}), which appeared in the $\widehat{\mathbf{E}}'_{CP}$ values, were not in accordance with the values of the vector of PPG imposed by the breeder, $\mathbf{d}'_{PG} = [7 \;\; -3 \;\; 3.5 \;\; -1.5]$, as we would expect, and CPPG-LGSI accuracy (0.234) was low. These results indicate that in the CPPG-LGSI, it is very important for the estimated values of $\widehat{\theta}_{CP}$ to be positive (see Chaps. 3 and 6 for details).

In the PPG-GESIM, we need to find the solutions to equation $\left(\widehat{\mathbf{T}}_{PG} - \widetilde{\lambda}^2_{PG_j} \mathbf{I}_{2t}\right) \widehat{\boldsymbol{\beta}}_{PG_j} = \mathbf{0}$, for $\widetilde{\lambda}^2_{PG_j}$ and $\widehat{\boldsymbol{\beta}}_{PG_j}$ (see Eq. 8.45). The estimated PPG-GESIM vector of coefficients was $\widehat{\boldsymbol{\beta}}'_{PG} = [0.001 \;\; -0.050 \;\; 0.029 \;\; 0.975 \;\; 0.154 \;\; -0.157]$, which was transformed using matrix $\mathbf{F} = \begin{bmatrix} -0.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$, that is, we

changed the direction of the original vector. With the $\widehat{\boldsymbol{\beta}}'_{PG}$ values, we can estimate the PPG-GESIM index as

$$\widehat{I}_{PG} = 0.001\text{GY} - 0.05\text{EHT} + 0.029\text{PHT} + 0.975\text{GEBV}_{\text{GY}} + 0.154\text{GEBV}_{\text{EHT}}$$
$$- 0.157\text{GEBV}_{\text{PHT}}$$

where GEBV_{GY}, GEBV_{EHT}, and GEBV_{PHT} are the GEBVs associated with the traits GY, EHT, and PHT respectively. The estimated PPG-GESIM selection response, accuracy, and expected genetic gain per trait were $\widehat{R}_{PG} = k_I \sqrt{\widehat{\boldsymbol{\beta}}'_{PG} \widehat{\boldsymbol{\Phi}} \widehat{\boldsymbol{\beta}}_{PG}} = 0.696$, $\widehat{\rho}_{\widehat{H}_{PG}\widehat{I}_{PG}} = \dfrac{\widehat{\sigma}_{\widehat{I}_{PG}}}{\widehat{\sigma}_{\widehat{H}_{PG}}} = 0.843$, and $\widehat{\mathbf{E}}'_{PG} = k_I \dfrac{\widehat{\boldsymbol{\beta}}'_{PG}\widehat{\mathbf{A}}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{PG}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\beta}}_{PG}}} =$

$[0.01 \ -1.00 \ -3.56 \ 0 \ -0.46 \ -3.98]$ respectively.

Fig. 8.4 presents the frequency distribution of the 244 estimated PPG-GESIM index values for two predetermined restrictions on the traits GY and EHT and their associated GEBVs (GEBV_{GY} and GEBV_{EHT}), for one selection cycle in an environment for a real maize (*Zea mays*) F_2 population with 233 molecular markers. Note that the frequency distribution of the estimated PPG-GESIM index values approaches normal distribution.

Now, with a selection intensity of 10% ($k_I = 1.755$) and a vector of predetermined restrictions $\mathbf{d}'_{PG} = [7 \ -3 \ 5 \ 3.5 \ -1.5 \ 2.5]$, we compare the estimated CPPG-LGSI and PPG-GESIM selection responses and expected genetic gains per
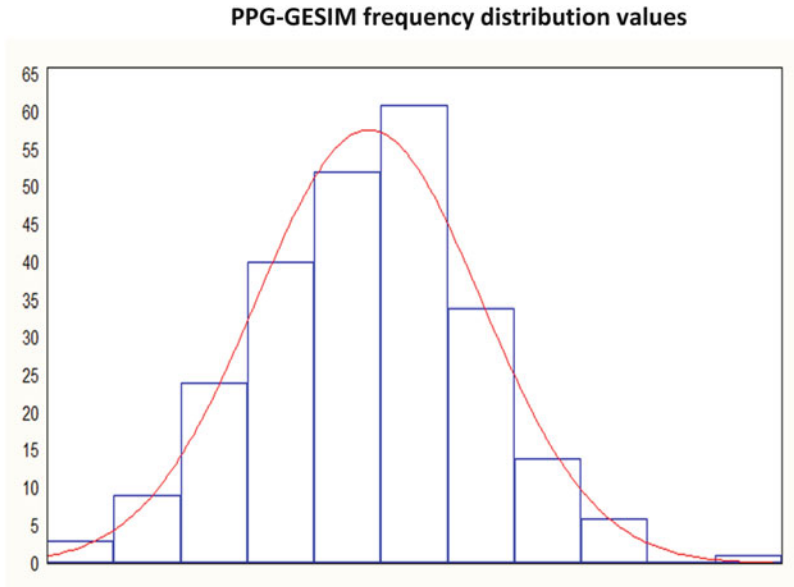

Fig. 8.4 Frequency distribution of the 244 estimated predetermined proportional gain genomic eigen selection index method (PPG-GESIM) values for two predetermined restrictions on the traits GY and EHT and their associated GEBVs, GEBV_{GY} and GEBV_{EHT}, for one selection cycle in an environment for a real maize (*Zea mays*) F_2 population with 233 molecular markers

trait using the simulated data set described in Sect. 2.8.1 of Chap. 2. Traits T1, T2, and T3 and their associated GEBVs (GEBV1, GEBV2, and GEBV3 respectively) were restricted, but trait T4 and its associated GEBV4 were not restricted. For this data set, matrix **F** was an identity matrix of size 8×8 for all four selection cycles.

Table 8.6 presents the estimated CPPG-LGSI selection responses when their vectors of coefficients are normalized, and the estimated PPG-GESIM selection responses for one, two, and three predetermined restrictions for four simulated selection cycles. The averages of the estimated CPPG-LGSI selection responses were 5.08 for one restriction, 3.42 for two restrictions, and 1.60 for three restrictions, whereas the averages of the estimated PPG-GESIM selection responses were 1.96 for one restriction, 4.14 for two restrictions, and 5.46 for three restrictions. For this data set, when the number of restrictions increases, the estimated CPPG-LGSI

Table 8.6 Estimated CPPG-LGSI expected genetic gains for one, two, and three restricted predetermined traits (T1, T2, and T3) and for one, two, and three restricted predetermined GEBVs (GEBV1, GEBV2, and GEBV3) for four simulated selection cycles

| | CPPG-LGSI expected genetic gain for one predetermined restriction | | | | | | | |
| | Traits | | | | Genomic estimated breeding values | | | |
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	8.24	−3.62	3.32	2.26	4.12	−2.33	1.75	1.09
2	7.98	−4.06	3.03	2.68	3.99	−2.24	1.79	1.04
3	8.61	−4.48	3.24	1.96	4.30	−2.32	1.70	0.98
4	8.30	−4.34	3.32	2.04	4.15	−2.16	1.62	0.92
Average	8.28	−4.12	3.23	2.23	4.14	−2.26	1.71	1.01
	CPPG-LGSI expected genetic gain for two predetermined restrictions							
	Traits				Genomic estimated breeding values			
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	8.06	−3.46	3.30	2.06	4.03	−1.73	1.72	0.98
2	8.17	−3.50	3.08	2.65	4.09	−1.75	1.79	0.98
3	8.88	−3.81	3.31	1.83	4.44	−1.90	1.72	0.90
4	8.61	−3.69	3.43	1.99	4.30	−1.84	1.65	0.87
Average	8.43	−3.61	3.28	2.13	4.22	−1.81	1.72	0.93
	CPPG-LGSI expected genetic gain for three predetermined restrictions							
	Traits				Genomic estimated breeding values			
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	5.77	−2.47	4.12	2.28	2.88	−1.24	2.06	0.98
2	5.68	−2.43	4.06	2.76	2.84	−1.22	2.03	0.97
3	5.87	−2.52	4.20	1.98	2.94	−1.26	2.10	0.79
4	5.91	−2.53	4.22	2.00	2.95	−1.27	2.11	0.83
Average	5.81	−2.49	4.15	2.26	2.90	−1.24	2.07	0.89

The selection intensity was 10% ($k_I = 1.755$) and the vector of predetermined restrictions was $\mathbf{d}'_{PG} = \begin{bmatrix} 7 & -3 & 5 & 3.5 & -1.5 & 2.5 \end{bmatrix}$. Trait T4 and its associated GEBV4 were not restricted

selection response tends to decrease, whereas the estimated PPG-GESIM selection response increases.

Tables 8.7 presents the estimated CPPG-LGSI and PPG-GESIM expected genetic gains for one, two, and three predetermined restrictions respectively, for four simulated selection cycles. The averages of the estimated CPPG-LGSI expected genetic gains for the four traits and their four associated GEBVs were 8.28, −4.12, 3.23, 2.23, 4.14, −2.26, 1.71, and 1.01 for one restriction; 8.43, −3.61, 3.28, 2.13, 4.22, −1.81, 1.72, and 0.93 for two restrictions; and 5.81, −2.49, 4.15, 2.26, 2.90, −1.24, 2.07, and 0.89 for three restrictions. On the other hand, the averages of the estimated PPG-GESIM expected genetic gains for the four traits and their four associated GEBVs were 6.97, −1.31, 1.78, 0.52, 5.64, −1.74, 1.75, and 0.58 for one restriction; 6.93, −2.73, 1.29, 0.85, 5.75, −2.55, 1.49, and 0.79 for two restrictions, and 8.12, −3.27, 2.99, 1.13, 2.19, −1.15, 1.30, and 0.45 for three

Table 8.7 Estimated PPG-GESIM expected genetic gains for one, two, and three restricted traits (T1, T2, and T3) and for one, two, and three restricted GEBVs (GEBV1, GEBV2, and GEBV3) for four simulated selection cycles

| | PPG-GESIM expected genetic gain for one predetermined restriction | | | | | | | |
| | Traits | | | | Genomic estimated breeding values | | | |
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	6.89	−1.44	1.94	0.63	6.36	−1.89	2.04	0.62
2	6.71	−1.33	1.90	0.65	6.06	−2.00	1.97	0.75
3	7.09	−1.69	1.67	0.40	5.40	−1.72	1.63	0.55
4	7.18	−0.78	1.58	0.39	4.73	−1.34	1.35	0.39
Average	6.97	−1.31	1.78	0.52	5.64	−1.74	1.75	0.58
	PPG-GESIM expected genetic gain for two predetermined restrictions							
	Traits				Genomic estimated breeding values			
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	6.61	−2.55	1.40	0.94	6.49	−2.80	1.75	0.87
2	5.67	−2.48	1.24	0.87	6.16	−2.84	1.70	0.91
3	7.35	−3.08	1.21	0.85	5.54	−2.49	1.37	0.82
4	8.10	−2.80	1.29	0.76	4.80	−2.08	1.16	0.56
Average	6.93	−2.73	1.29	0.85	5.75	−2.55	1.49	0.79
	PPG-GESIM expected genetic gain for three predetermined restrictions							
	Traits				Genomic estimated breeding values			
Cycle	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
1	7.21	−2.94	2.64	1.02	1.69	−1.10	1.07	0.45
2	7.71	−2.97	2.41	1.46	2.22	−1.15	1.21	0.45
3	8.72	−3.43	3.17	0.93	2.21	−1.06	1.34	0.42
4	8.85	−3.73	3.72	1.09	2.63	−1.29	1.60	0.48
Average	8.12	−3.27	2.99	1.13	2.19	−1.15	1.30	0.45

The selection intensity was 10% ($k_I = 1.755$) and the vector of predetermined restrictions was $\mathbf{d}'_{PG} = [7 \quad −3 \quad 5 \quad 3.5 \quad −1.5 \quad 2.5]$. Trait T4 and its associated GEBV4 were not restricted

restrictions. These results indicate that the estimated CPPG-LGSI expected genetic gains for the four traits and their four associated GEBVs were generally higher than the estimated PPG-GESIM expected genetic gains for the four traits and their four associated GEBVs.

References

Crossa J, Cerón-Rojas JJ (2011) Multi-trait multi-environment genome-wide molecular marker selection indices. J Indian Soc Agric Stat 62(2):125–142

Meyer CD (2000) Matrix analysis and applied linear algebra. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA