

# HANDBOOK OF PLANT BREEDING



NATALIE SAWYER

# Handbook of Plant Breeding



# Handbook of Plant Breeding

Natalie Sawyer

Handbook of Plant Breeding  
Natalie Sawyer  
ISBN: 978-1-9790-0173-2

© 2022 World Technologies

Published by World Technologies,  
5 Penn Plaza,  
19th Floor,  
New York, NY 10001, USA

This book contains information obtained from authentic and highly regarded sources. All chapters are published with permission under the Creative Commons Attribution Share Alike License or equivalent. A wide variety of references are listed. Permissions and sources are indicated; for detailed attributions, please refer to the permissions page. Reasonable efforts have been made to publish reliable data and information, but the authors, editors and publisher cannot assume any responsibility for the validity of all materials or the consequences of their use.

Trademark Notice: All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy. Furthermore, the publisher ensures that the text paper and cover boards used have met acceptable environmental accreditation standards.

# Table of Contents

<b>Chapter 1</b>	Constrained Linear Genomic Selection Indices	<b>1</b>
<b>Chapter 2</b>	Linear Phenotypic Eigen Selection Index Methods	<b>28</b>
<b>Chapter 3</b>	Linear Molecular and Genomic Eigen Selection Index Methods	<b>56</b>
<b>Chapter 4</b>	Multistage Linear Selection Indices	<b>85</b>
<b>Chapter 5</b>	Stochastic Simulation of Four Linear Phenotypic Selection Indices	<b>108</b>
<b>Chapter 6</b>	RIndSel: Selection Indices with R	<b>119</b>

# WT

WORLD TECHNOLOGIES

# Efficiency of a Constrained Linear Genomic Selection Index: Statistical Properties and Examples



**Abstract** The constrained linear genomic selection indices are null restricted and predetermined proportional gain linear genomic selection indices (RLGSI and PPG-LGSI respectively), which are a linear combination of genomic estimated breeding values (GEBVs) to predict the net genetic merit. They are the results of a direct application of the restricted and the predetermined proportional gain linear phenotypic selection index theory to the genomic selection context. The RLGSI can be extended to a combined RLGSI (CRLGSI) and the PPG-LGSI can be extended to a combined PPG-LGSI (CPPG-LGSI); the latter indices use phenotypic and GEBV information jointly in the prediction of net genetic merit. The main difference between the RLGSI and PPG-LGSI with respect to the CRLGSI and the CPPG-LGSI is that although the RLGSI and PPG-LGSI are useful in a testing population where there is only marker information, the CRLGSI and CPPG-LGSI can be used only in training populations when there are joint phenotypic and marker information. The RLGSI and CRLGSI allow restrictions equal to zero to be imposed on the expected genetic advance of some traits, whereas the PPG-LGSI and CPPG-LGSI allow predetermined proportional restriction values to be imposed on the expected trait genetic gains to make some traits change their mean values based on a predetermined level. We describe the foregoing four indices and we validated their theoretical results using real and simulated data.

## 1.1 The Restricted Linear Genomic Selection Index

Let  $H = \mathbf{w}'\mathbf{g}$  be the net genetic merit and  $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$  the linear genomic selection index (LGSI, see Chap. 5 for details), where  $\mathbf{g}$ ,  $\boldsymbol{\gamma}$ ,  $\mathbf{w}$ , and  $\boldsymbol{\beta}$  are vectors  $t \times 1$  ( $t$ = number of traits) of breeding values, genomic breeding values, economic weights, and LGSI coefficients respectively. It can be shown that  $Cov(I_G, \mathbf{g}) = \boldsymbol{\Gamma}\boldsymbol{\beta}$  is the covariance between  $\mathbf{g}$  and  $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$ , and that  $Var(\boldsymbol{\gamma}) = \boldsymbol{\Gamma}$  is the genomic covariance matrix of size  $t \times t$  (see Chap. for details). The objective of the restricted linear genomic selection index (RLGSI) is to improve only  $(t - r)$  of  $t$  ( $r < t$ ) traits (leaving  $r$  of them fixed) in a testing population using only genomic estimated breeding values

(GEBVs). The RLGSI minimizes the mean squared difference between  $I_G$  and  $H$ ,  $E[(H - I_G)^2]$ , with respect to  $\beta$  under the restriction  $Cov(I_G, \mathbf{U}'\mathbf{g}) = \mathbf{U}'\boldsymbol{\Gamma}\beta = \mathbf{0}$ , where  $\mathbf{U}'$  is a matrix  $(t - 1) \times t$  of 1s and 0s, in a similar manner to the restricted linear phenotypic selection index (RLPSI) described in Chap. 3 in the phenotypic selection context.

### 1.1.1 The Maximized RLGSI Parameters

Let  $Var(I_G) = \beta'\boldsymbol{\Gamma}\beta$  be the variance of  $I_G = \beta'\boldsymbol{\gamma}$ ,  $\mathbf{w}'\mathbf{C}\mathbf{w}$  the variance of  $H = \mathbf{w}'\mathbf{g}$ , and  $Cov(I_G, H) = \mathbf{w}'\boldsymbol{\Gamma}\beta$  the covariance between  $H = \mathbf{w}'\mathbf{g}$  and  $I_G = \beta'\boldsymbol{\gamma}$ . The mean squared difference between  $H$  and  $I_G$  can be written as  $E[(H - I_G)^2]$ , which should be minimized under the restriction  $\mathbf{U}'\boldsymbol{\Gamma}\beta = \mathbf{0}$  assuming that  $\boldsymbol{\Gamma}$ ,  $\mathbf{C}$ ,  $\mathbf{U}'$ , and  $\mathbf{w}$  are known, i.e., it is necessary to minimize the function

$$f_R(\beta, \mathbf{v}) = \mathbf{w}'\mathbf{C}\mathbf{w} + \beta'\boldsymbol{\Gamma}\beta - 2\mathbf{w}'\boldsymbol{\Gamma}\beta + 2\mathbf{v}'\mathbf{U}'\boldsymbol{\Gamma}\beta \quad (1.1)$$

with respect to vectors  $\beta$  and  $\mathbf{v}' = [v_1 \ v_2 \ \cdots \ v_{t-1}]$ , where  $\mathbf{v}$  is a vector of Lagrange multipliers. In matrix notation, the derivative results of Eq. (1.1) are

$$\begin{bmatrix} \beta \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\mathbf{U} \\ \mathbf{U}'\boldsymbol{\Gamma} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\Gamma}\mathbf{w} \\ \mathbf{0} \end{bmatrix}. \quad (1.2)$$

Following the procedure described in Chap. (Eqs. 1.2 to 1.5), it can be shown that the RLGSI vector of coefficients that minimizes  $E[(H - I_G)^2]$  under the restriction  $\mathbf{U}'\boldsymbol{\Gamma}\beta = \mathbf{0}$  is

$$\beta_{RG} = \mathbf{K}_G \mathbf{w}, \quad (1.3)$$

where  $\mathbf{K}_G = [\mathbf{I}_t - \mathbf{Q}_G]$ ,  $\mathbf{Q}_G = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\Gamma}$ ,  $\mathbf{w}$  is a vector of economic weights, and  $\mathbf{I}_t$  is an identity matrix  $t \times t$ . When no restrictions are imposed on any of the traits,  $\mathbf{U}'$  is a null matrix and  $\beta_{RG} = \mathbf{w}$ , the optimized LGSI vector of coefficients (see Chap. for details).

By Eq. (1.3), the RLGSI, and the maximized RLGSI selection response and expected genetic gain per trait can be written as

$$I_{RG} = \beta'_{RG}\boldsymbol{\gamma}, \quad (1.4)$$

$$R_{RG} = \frac{k_I}{L_G} \sqrt{\beta'_{RG}\boldsymbol{\Gamma}\beta_{RG}} \quad (1.5)$$

and

$$\mathbf{E}_{RG} = \frac{k_I}{L_G} \frac{\boldsymbol{\Gamma} \boldsymbol{\beta}_{RG}}{\sqrt{\boldsymbol{\beta}'_{RG} \boldsymbol{\Gamma} \boldsymbol{\beta}_{RG}}}, \quad (1.6)$$

respectively, where  $k_I$  is the standardized selection differential (or selection intensity) associated with the RLGSI, and  $L_G$  is the interval between selection cycles or the time required to complete a selection cycle using the RLGSI. Equations (1.4) to (1.6) depend only on GEBV information; thus, they are useful in testing populations.

### 1.1.2 Statistical Properties of RLGSI

Assuming that  $H = \mathbf{w}'\mathbf{g}$  and  $I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}$  have bivariate joint normal distribution,  $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$ , and  $\boldsymbol{\Gamma}$ ,  $\mathbf{C}$ , and  $\mathbf{w}$  are known, it can be shown that the RLGSI has the following properties:

1. Matrices  $\mathbf{K}_G$  and  $\mathbf{Q}_G$  are idempotent ( $\mathbf{K}_G = \mathbf{K}_G^2$  and  $\mathbf{Q}_G = \mathbf{Q}_G^2$ ) and orthogonal ( $\mathbf{K}_G\mathbf{Q}_G = \mathbf{Q}_G\mathbf{K}_G = \mathbf{0}$ ), that is, they are projectors. Matrix  $\mathbf{Q}_G$  projects vector  $\boldsymbol{\beta} = \mathbf{w}$  into a space generated by the columns of matrix  $\mathbf{U}'\boldsymbol{\Gamma}$  due to the restriction  $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$  used when  $f_R(\boldsymbol{\beta}, \mathbf{v})$  (Eq. 1.1) is minimized with respect to vectors  $\boldsymbol{\beta}$  and  $\mathbf{v}$ , whereas matrix  $\mathbf{K}_G$  projects  $\mathbf{w}$  into a space perpendicular to that generated by the  $\mathbf{U}'\boldsymbol{\Gamma}$  matrix columns.
2. Because of the restriction  $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$ , matrix  $\mathbf{K}_G$  projects vector  $\mathbf{w}$  into a space smaller than the original space of  $\mathbf{w}$ . The space reduction into which matrix  $\mathbf{K}_G$  projects  $\mathbf{w}$  is equal to the number of zeros that appears in Eq. (1.6).
3. Vector  $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$  minimizes the mean square error under the restriction  $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$ .
4. The variance of  $I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}$  ( $\sigma_{I_{RG}}^2 = \boldsymbol{\beta}'_{RG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}$ ) is equal to the covariance between  $I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}$  and  $H = \mathbf{w}'\mathbf{g}$  ( $\sigma_{HI_{RG}} = \mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}$ ).
5. The maximized correlation between  $H$  and  $I_{RG}$  is equal to  $\rho_{HI_{RG}} = \frac{\sigma_{I_{RG}}}{\sigma_H}$ , where  $\sigma_{I_{RG}} = \sqrt{\boldsymbol{\beta}'_{RG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}}$  and  $\sigma_H = \sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}$  are the standard deviations of  $I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}$  and  $H = \mathbf{w}'\mathbf{g}$  respectively.
6. The variance of the predicted error,  $Var(H - I_{RG}) = (1 - \rho_{HI_{RG}}^2)\sigma_H^2$ , is minimal. Note that  $Var(H - I_{RG}) = \sigma_{I_{RG}}^2 + \sigma_H^2 - 2\sigma_{HI_{RG}}$ , and when  $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$ ,  $\sigma_{I_{RG}}^2 = \sigma_{HI_{RG}}$ , whence  $Var(H - I_{RG}) = \sigma_H^2 - \sigma_{I_{RG}}^2 = (1 - \rho_{HI_{RG}}^2)\sigma_H^2$  is minimal.

The statistical RLGSI properties are equal to the statistical RLPSI properties. Thus the RLGSI is an application of the RLPSI to the genomic selection context.

### 1.1.3 Numerical Examples

To estimate the parameters associated with the RLGSI, we use the real data set described in Chap. Sect.1.8, where we found that, in the testing population, the

estimate of matrix  $\Gamma$  was  $\widehat{\Gamma} = \begin{bmatrix} 0.21 & 2.95 & 5.00 \\ 2.95 & 42.41 & 71.11 \\ 5.00 & 71.11 & 121.53 \end{bmatrix}$ . We use this matrix and the

GEBVs associated with the traits grain yield (GY, ton ha<sup>-1</sup>), ear height (EHT, cm), and plant height (PHT, cm) to illustrate the RLGSi theoretical results.

Suppose that on the RLGSi expected genetic gain per trait we impose one and two null restrictions using matrices  $\mathbf{U}'_1 = [1 \ 0 \ 0]$  and  $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

(see Chap. Sect. 1.3, for details about matrix  $\mathbf{U}$ ). We need to estimate the RLGSi vector of coefficients ( $\beta_{RG} = \mathbf{K}_G \mathbf{w}$ ) as  $\widehat{\beta}_{RG} = \widehat{\mathbf{K}}_G \mathbf{w}$ , where  $\widehat{\mathbf{K}}_G = [\mathbf{I}_3 - \widehat{\mathbf{Q}}_G]$  and  $\widehat{\mathbf{Q}}_G = \mathbf{U}(\mathbf{U}'\widehat{\Gamma}\mathbf{U})^{-1}\mathbf{U}'\widehat{\Gamma}$  are estimates of matrices  $\mathbf{K}_G = [\mathbf{I}_3 - \mathbf{Q}_G]$  and  $\mathbf{Q}_G = \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\Gamma$  respectively, and  $\mathbf{I}_3$  is an identity matrix  $3 \times 3$ . The estimated  $\mathbf{Q}_G$  matrices for restrictions  $\mathbf{U}'_1 = [1 \ 0 \ 0]$  and  $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$  were  $\widehat{\mathbf{Q}}_{G_1} = \mathbf{U}_1(\mathbf{U}'_1\widehat{\Gamma}\mathbf{U}_1)^{-1}$

$\mathbf{U}'_1\widehat{\Gamma} = \begin{bmatrix} 1.0 & 14.05 & 23.81 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$  and  $\widehat{\mathbf{Q}}_{G_2} = \mathbf{U}_2(\mathbf{U}'_2\widehat{\Gamma}\mathbf{U}_2)^{-1}\mathbf{U}'_2\widehat{\Gamma} = \begin{bmatrix} 1.0 & 0 & 11.18 \\ 0 & 1.0 & 0.90 \\ 0 & 0 & 0 \end{bmatrix}$  respec-

tively, whereas the estimated  $\mathbf{K}_G$  matrices for both restrictions were  $\widehat{\mathbf{K}}_{G_1} = [\mathbf{I}_3 - \widehat{\mathbf{Q}}_{G_1}]$

$$= \begin{bmatrix} 0 & -14.05 & -23.81 \\ 0 & 1.0 & 0 \\ 0 & 0 & 1.0 \end{bmatrix} \text{ and } \widehat{\mathbf{K}}_{G_2} = [\mathbf{I}_3 - \widehat{\mathbf{Q}}_{G_2}] = \begin{bmatrix} 0 & 0 & -11.18 \\ 0 & 0 & -0.90 \\ 0 & 0 & 1.0 \end{bmatrix}.$$

Let  $\mathbf{w}' = [5 \ -0.1 \ -0.1]$  be the vector of economic weights; then the estimated RLGSi vector of coefficients for one and two null restrictions were  $\widehat{\beta}'_{RG_1} = \mathbf{w}'\widehat{\mathbf{K}}'_{G_1} = [3.78 \ -0.1 \ -0.1]$  and  $\widehat{\beta}'_{RG_2} = \mathbf{w}'\widehat{\mathbf{K}}'_{G_2} = [1.12 \ 0.09 \ -0.1]$  respectively, and the estimated RLGSi for both restrictions can be written as  $\widehat{I}_{RG_1} = 3.78\text{GEBV}_1 - 0.1\text{GEBV}_2 - 0.1\text{GEBV}_3$  and  $\widehat{I}_{RG_2} = 1.12\text{GEBV}_1 + 0.09\text{GEBV}_2 - 0.1\text{GEBV}_3$ , where GEBV<sub>1</sub>, GEBV<sub>2</sub>, and GEBV<sub>3</sub> are the genomic estimated breeding values associated with traits GY, EHT, and PHT respectively in the testing population.

Table 6.1 presents 20 genotypes selected from a population of 380 genotypes and the GEBVs in the testing population ranked according to the estimated RLGSi values for one restriction, where  $\mathbf{U}'_1 = [1 \ 0 \ 0]$ . The estimated RLGSi values for genotypes 5 and 306 can be obtained as follows:  $\widehat{I}_{RG_5} = 3.78(-0.6) - 0.1(-8.67) - 0.1(15.97) = 0.196$  and  $\widehat{I}_{RG_{306}} = 3.78(0.13) - 0.1(1.31) - 0.1(1.66) = 0.194$  respectively. This procedure is valid for any number of genotypes and GEBVs in the testing population.

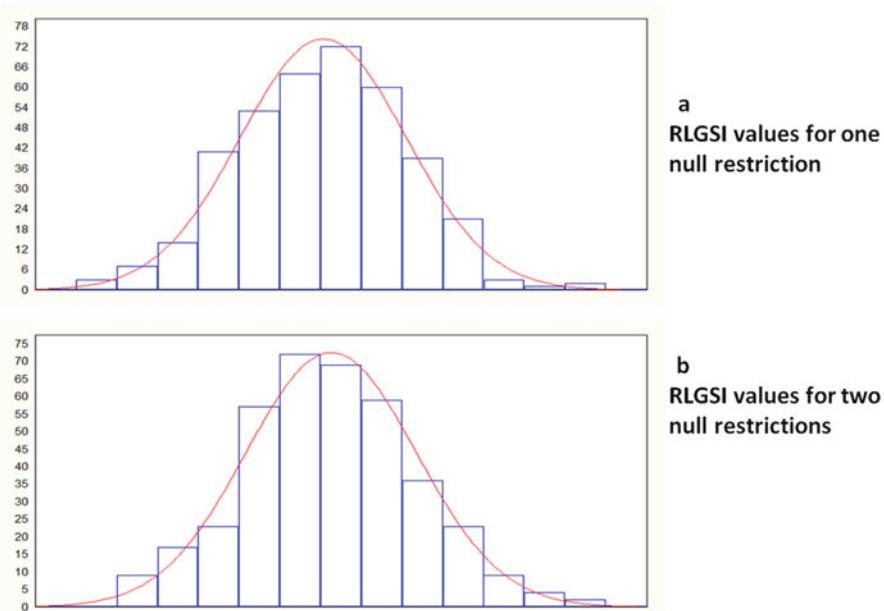
Assume a selection intensity of 10% ( $k_{IG} = 1.755$ ); then the estimated RLGSi selection response and expected genetic gain per trait not including the interval length were  $\widehat{R}_{RG_1} = k_{IG} \sqrt{\widehat{\beta}'_{RG_1}\widehat{\Gamma}\widehat{\beta}_{RG_1}} = 0.40$  and  $\widehat{\mathbf{E}}'_{RG_1} = k_I \frac{\widehat{\beta}'_{RG_1}\widehat{\Gamma}}{\sqrt{\widehat{\beta}'_{RG_1}\widehat{\Gamma}\widehat{\beta}_{RG_1}}}$

**Table 1.1** Number of genotypes selected from 380 genotypes of a real testing population; genomic estimated breeding values (GEBVs) associated with three traits: grain yield (GY, ton ha<sup>-1</sup>), ear height (EHT, cm), and plant height (PHT, cm) in the testing population, and estimated and ranked restricted linear genomic selection index (RLGSI) values obtained in the testing population for one null restriction

Number of genotypes	Estimated GEBVs in the testing population			Estimated RLGSI
	GEBV-GY	GEBV-EHT	GEBV-PHT	
5	-0.6	-8.67	-15.97	0.196
306	0.13	1.31	1.66	0.194
6	0.06	1.83	-1.13	0.157
349	0.37	4.34	8.12	0.153
142	-0.26	-5.47	-5.85	0.149
69	-0.11	-3.43	-2.16	0.143
24	0.03	-0.43	0.19	0.137
192	-0.8	-13.91	-17.7	0.137
33	-0.18	-1.44	-6.71	0.135
18	-0.43	-5.48	-12.08	0.131
21	-1.00	-16.11	-22.96	0.127
41	0.17	1.09	4.08	0.126
351	0.16	2.64	2.15	0.126
323	0.04	-0.79	1.04	0.126
158	-0.49	-8.95	-10.83	0.126
25	-0.24	-3.46	-6.86	0.125
338	0.37	3.88	8.89	0.122
316	-0.01	-0.51	-1.09	0.122
32	-0.19	-3.97	-4.43	0.122
204	-0.46	-7.41	-11.19	0.121

$= [0 \quad -1.42 \quad -2.58]$  respectively. For two restrictions, with  $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ , the estimated RLGSI selection response and expected genetic gains not including the interval length were  $\widehat{R}_{RG_2} = k_{I_G} \sqrt{\widehat{\beta}'_{RG_2} \widehat{\Gamma} \widehat{\beta}_{RG_2}} = 0.23$  and  $\widehat{\mathbf{E}}'_{RG_2} = k_I \frac{\widehat{\beta}'_{RG_2} \widehat{\Gamma}}{\sqrt{\widehat{\beta}'_{RG_2} \widehat{\Gamma} \widehat{\beta}_{RG_2}}} = [0 \quad 0 \quad -2.29]$  respectively. When the number of restrictions increases, the estimated RLGSI selection response value decreases, whereas the number of zeros increases in the estimated RLGSI expected genetic gain per trait. The number of zeros in the estimated RLGSI expected genetic gain per trait is equal to the number of restrictions imposed on RLGSI by matrix  $\mathbf{U}'$ , where each restriction appears as 1.

Figure 6.1 presents the frequency distribution of the estimated RLGSI values for one (Fig. 1.1a) and two null restrictions (Fig. 1.1b). For both restrictions the frequency distribution of the estimated RLGSI values approaches the normal distribution.



**Fig. 1.1** Distribution of 380 estimated restricted linear genomic selection index (RLGSI) values with one (a) and two (b) null restrictions respectively obtained in a real testing population for one selection cycle in one environment

Now we use the simulated data set described in Chap. Sect. 1, to compare RLPSI (restricted linear phenotypic selection index, Chap. for details) efficiency versus RLGSI efficiency. Table 1.2 presents the estimated RLPSI and RLGSI selection response for one, two, and three null restrictions imposed by matrices

$$\mathbf{U}'_1 = [1 \ 0 \ 0], \mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \text{ and } \mathbf{U}'_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ for five simulated}$$

selection cycles including and not including the interval between selection cycles. In each selection cycle, the sample size was equal to 500 genotypes, each with four repetitions and four traits, whereas the selection intensity was 10% ( $k_I = 1.755$ ); the interval lengths for the RLPSI and RLGSI were 4 and 1.5 years (Beyene et al. 2015) respectively.

Table 1.2 was divided in two parts. The first part presents the estimated RLPSI whereas the second part presents the estimated RLGSI selection responses. Columns 2, 3, and 4 in Table 1.2 present the estimated RLPSI and RLGSI selection responses not including the interval length, whereas columns 5, 6, and 7 present the estimated RLPSI and RLGSI selection response, including the interval length. The averages of the estimated RLPSI selection response not including the interval length for one, two, and three restrictions were 7.04, 5.50, and 3.90, whereas when the interval length was included, the averages were 1.76, 1.38, and 0.98 respectively. The averages of the estimated RLGSI selection response not including the interval length

**Table 1.2** Estimated restricted linear phenotypic selection index (RLPSI) and RLGSI selection responses for 1, 2, and 3 null restrictions for 5 simulated selection cycles including and not including the interval between selection cycles. The interval lengths for the RLPSI and the RLGSI were 4 and 1.5 years respectively

Cycle	Estimated RLPSI selection response					
	Not including interval length			Including interval length <sup>a</sup>		
	1	2	3	1	2	3
1	6.87	5.54	4.13	1.72	1.39	1.03
2	8.45	5.94	4.27	2.11	1.49	1.07
3	7.17	5.79	4.16	1.79	1.45	1.04
4	6.68	5.06	3.72	1.67	1.27	0.93
5	6.02	5.16	3.24	1.51	1.29	0.81
Average	7.04	5.50	3.90	1.76	1.38	0.98
Cycle	Estimated RLGSI selection response					
	Not including interval length			Including interval length <sup>b</sup>		
	1	2	3	1	2	3
1	6.41	5.58	4.71	4.28	3.72	3.14
2	5.04	3.47	2.47	3.36	2.32	1.65
3	4.76	3.36	2.22	3.17	2.24	1.48
4	4.51	3.07	2.28	3.01	2.05	1.52
5	4.46	3.10	2.26	2.97	2.07	1.51
Average	5.04	3.72	2.79	3.36	2.48	1.86

<sup>a</sup>The estimated RLPSI selection response was divided by 4

<sup>b</sup>The estimated RLGSI selection response was divided by 1.5

for one, two, and three restrictions were 5.04, 3.72, and 2.79, whereas when the interval length was included the averages were 3.36, 2.48, and 1.86 respectively. These results indicated that when the interval length was included in the estimation of the RLPSI and RLGSI selection response, RLGSI efficiency was greater than RLPSI efficiency, and vice versa, when the interval length was not included the RLPSI efficiency was greater than RLGSI efficiency.

Table 1.3 presents the estimated RLPSI (first part) and RLGSI (second part) expected genetic gain per trait not including the interval between selection cycles for one, two, and three null restrictions in five simulated selection cycles. In this case, RLPSI efficiency is greater than RLGSI efficiency because the averages of the estimated RLPSI expected genetic gain per trait were -2.52, 2.26, and 2.26 for one null restriction; 2.84 and 2.65 for two null restrictions; and 3.90 for three null restrictions. For the same set of restrictions, the averages of the estimated RLGSI expected genetic gain per trait were: -1.85, 1.13, and 2.06 for one null restriction; 1.52 and 2.19 for two null restrictions, and 2.79 for three null restrictions. However, divided by the interval length (4 years in the RLPSI), the averages of the estimated RLPSI expected genetic gain per trait were -0.63, 0.57, and 0.57 for one null restriction; 0.71 and 0.66 for two null restrictions, and 0.98 for three null restrictions. In a similar manner, dividing by the interval length (1.5 years in this case), the averages of the estimated RLGSI expected genetic gain per trait were -1.23, 0.75,

**Table 1.3** Estimated RLPSI and RLGSI expected genetic gain per trait for 1, 2, and 3 null restrictions for 5 simulated selection cycles (each with 4 traits) not including the interval length between selection cycles

Cycle	Estimated RLPSI expected genetic gain for one, two, and three null restrictions											
	1				2				3			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
1	0	-2.18	2.03	2.66	0	0	2.77	2.77	0	0	0	4.13
2	0	-3.41	2.33	2.71	0	0	2.87	3.07	0	0	0	4.27
3	0	-2.30	3.12	1.74	0	0	3.11	2.68	0	0	0	4.16
4	0	-2.88	1.42	2.38	0	0	2.35	2.70	0	0	0	3.72
5	0	-1.83	2.38	1.81	0	0	3.12	2.04	0	0	0	3.24
Average	0	-2.52	2.26	2.26	0	0	2.84	2.65	0	0	0	3.90
Cycle	Estimated RLGSI expected genetic gain for 1, 2, and 3 null restrictions											
	1				2				3			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
1	0	-1.41	1.29	3.72	0	0	1.89	3.70	0	0	0	4.71
2	0	-2.16	1.07	1.81	0	0	1.49	1.98	0	0	0	2.47
3	0	-1.94	1.24	1.57	0	0	1.58	1.78	0	0	0	2.22
4	0	-1.90	1.02	1.60	0	0	1.34	1.73	0	0	0	2.28
5	0	-1.83	1.02	1.61	0	0	1.33	1.77	0	0	0	2.26
Average	0	-1.85	1.13	2.06	0	0	1.52	2.19	0	0	0	2.79

and 1.37 for one restriction; 1.01 and 1.46 for two restrictions; and 1.86 for three restrictions.

Table 1.4 presents the estimated RLPSI heritability ( $\widehat{h}_{I_R}^2$ ) values, the estimated restricted linear genomic selection index (RLGSI) accuracy ( $\widehat{\rho}_{HI_{RG}}$ ) values, the values of  $W = \frac{\widehat{\rho}_{HI_{RG}}}{\widehat{h}_{I_R}} L_{RP}$  ( $L_{RP} = 4$ ), and the values of  $\widehat{p} = 100(\widehat{\lambda}_R - 1)$ , where  $\widehat{\lambda}_R = \widehat{\rho}_{HI_R} / \widehat{\rho}_{HI_{RG}}$  and  $\widehat{\rho}_{HI_R}$  is the estimated RLPSI accuracy, for one, two, and three restrictions for five simulated selection cycles. The RLGSI interval length was  $L_{RG} = 1.5$  whereas the averages of the values of  $W = \frac{\widehat{\rho}_{HI_{RG}}}{\widehat{h}_{I_R}} L_{RP}$  for each restriction were 1.22, 0.85, and 0.60; this means that the estimated Technow inequality (Technow et al. 2013),  $L_{RG} < \frac{\widehat{\rho}_{HI_{RG}}}{\widehat{h}_{I_R}} L_{RP}$  (Chap. 5, Eq. 5.18), was not true. Thus, according to the Technow inequality results, for this data set, RLGSI efficiency in terms of time was not greater than RLPSI efficiency. The inequality  $L_{RG} < \frac{\widehat{\rho}_{HI_G}}{\widehat{h}_{I_R}} L_{I_R}$  was not true because the estimated RLGSI accuracy was very low, whereas RLPSI heritability was high. Thus, note that the averages of the estimated RLGSI accuracy for one, two, and three null restrictions were 0.25, 0.19, and 0.14 respectively, and the averages of the estimated RLPSI heritability values were 0.70, 0.78 and 0.88, respectively. Thus, according to these results, because the estimated RLGSI accuracy is very low and

**Table 1.4** Estimated RLPSTI heritability ( $\hat{h}_{I_k}^2$ ), estimated RLGS1 accuracy ( $\hat{\rho}_{H_{I_k}}$ ), estimated values of  $W = \frac{\hat{\rho}_{H_{I_k}}}{\hat{h}_{I_k}} L_{RP}$  ( $L_{RP} = 4$ ), and values of  $\hat{p} = 100(\hat{j}_k - 1)$ ,

where  $\hat{j}_k = \hat{\rho}_{H_{I_k}}/\hat{\rho}_{H_{I_{k0}}}$ , and  $\hat{\rho}_{H_{I_k}}$  are the estimated RLPSTI accuracy values, for 1, 2, and 3 restrictions for five simulated selection cycles

Cycle	RLPSTI heritability			RLGS1 accuracy			Estimated values of $W$			Estimated values of $\hat{p}$		
	1	2	3	1	2	3	1	2	3	1	2	3
1	0.65	0.77	0.89	0.33	0.28	0.24	1.62	1.29	1.02	7.27	-1.40	-12.28
2	0.76	0.80	0.90	0.26	0.18	0.13	1.20	0.80	0.54	84.12	83.76	87.74
3	0.71	0.80	0.88	0.24	0.17	0.11	1.16	0.77	0.49	80.34	103.03	119.72
4	0.71	0.79	0.89	0.22	0.15	0.11	1.06	0.68	0.48	79.02	97.29	94.65
5	0.67	0.76	0.86	0.22	0.15	0.11	1.07	0.70	0.48	74.31	110.97	80.61
Average	0.70	0.78	0.88	0.25	0.19	0.14	1.22	0.85	0.60	65.01	78.73	74.09



RLPSI heritability is high, RLGSI efficiency was lower than RLPSI efficiency in terms of time.

The last three columns of Table 1.4, from left to right, present the estimated  $p$  values,  $\hat{p} = 100(\hat{\lambda}_R - 1)$ , for one, two, and three null restrictions in five simulated selection cycles. The average of the  $\hat{p}$  values indicates that for each of the three restrictions the RLPSI efficiency was 65.05%, 78.73%, and 74.09%, greater than RLGSI efficiency at predicting the net genetic merit. Thus, for this data set, the RLPSI was a better predictor of the net genetic merit than the RLGSI in each cycle.

## 1.2 The Predetermined Proportional Gain Linear Genomic Selection Index

### 1.2.1 Objective of the PPG-LGSI

Let  $\mathbf{d}' = [d_1 \ d_2 \ \dots \ d_r]$  be a vector  $1 \times r$  ( $r$  is the number of predetermined proportional gains) of the predetermined proportional gains imposed by the breeder, and assume that  $\mu_q$  is the population mean of the  $q$ th trait before selection. The objective of the predetermined proportional gain linear genomic selection index (PPG-LGSI) is to change  $\mu_q$  to  $\mu_q + d_q$  in the testing population, where  $d_q$  is a predetermined change in  $\mu_q$ . It is possible to solve this problem minimizing the mean squared difference between  $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$  and  $H = \mathbf{w}'\mathbf{g}$ ,  $E[(H - I_G)^2]$ , under the restriction  $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \theta_G\mathbf{d}$ , where  $\theta_G$  is a proportionality constant, or under the

restriction  $\mathbf{D}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$ , where  $\mathbf{D}' = \begin{bmatrix} d_r & 0 & \dots & 0 & -d_1 \\ 0 & d_r & \dots & 0 & -d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & d_r & -d_{r-1} \end{bmatrix}$  is a matrix

$(r-1) \times r$  (see Chap. 3 for details), and  $d_q$  ( $q = 1, 2, \dots, r$ ) is the  $q^{\text{th}}$  element of vector  $\mathbf{d}' = [d_1 \ d_2 \ \dots \ d_r]$ ;  $\mathbf{U}'$  is a matrix  $(t-1) \times t$  of 1s and 0s, and  $\boldsymbol{\Gamma} = \left\{ \sigma_{\gamma_{qq'}} \right\}$  ( $q, q' = 1, 2, \dots, t$ ,  $t$  = number of traits) is a covariance matrix of additive genomic breeding values,  $\boldsymbol{\gamma}' = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_t]$ .

### 1.2.2 The Maximized PPG-LGSI Parameters

In this subsection, we minimize  $E[(H - I_G)^2]$  under the restriction  $\mathbf{D}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$  and later under the restriction  $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \theta_G\mathbf{d}$ . Under the restriction  $\mathbf{D}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$ , it is necessary to minimize the function

$$f_P(\boldsymbol{\beta}, \mathbf{v}) = \boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta} + \mathbf{w}'\mathbf{C}\mathbf{w} - 2\mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta} + 2\mathbf{v}'\mathbf{D}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} \quad (1.7)$$

with respect to  $\beta$  and  $v' = [v_1 \ v_2 \ \dots \ v_{r-1}]$ , where  $v'$  is a vector of Lagrange multipliers. From a mathematical point of view, Eq. (1.7) is equal to Eq. (1.1); thus, the vector of coefficients  $\beta$  of the PPG-LGSI should be similar to the vector of coefficients of the RLGSI (Eq. 1.3), i.e., the PPG-LGSI vector of coefficients is equal to

$$\beta_{PG} = \mathbf{K}_P \mathbf{w}, \quad (1.8)$$

where now  $\mathbf{K}_P = [\mathbf{I}_t - \mathbf{Q}_P]$ ,  $\mathbf{Q}_P = \mathbf{U}\mathbf{D}(\mathbf{D}'\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{D}'\mathbf{U}'\mathbf{T}$ ,  $\mathbf{w}$  is a vector of economic weights, and  $\mathbf{I}_t$  is an identity matrix  $t \times t$ . When  $\mathbf{D}' = \mathbf{U}'$ ,  $\beta_{PG} = \beta_{RG}$  (the RLGSI vector of coefficients), and when  $\mathbf{U}'$  is a null matrix,  $\beta_{PG} = \mathbf{w}$  (the LGSI vector of coefficients). This means that the PPG-LGSI includes the RLGSI and the LGSI as particular cases.

Under the restriction  $\mathbf{U}'\mathbf{T}\beta = \theta_G \mathbf{d}$  the vector of coefficients of the PPG-LGSI can be written as

$$\beta_{PG} = \beta_{RG} + \theta_G \mathbf{U}(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1} \mathbf{d}, \quad (1.9)$$

where  $\beta_{RG} = \mathbf{K}_G \mathbf{w}$  (Eq. 1.3),  $\mathbf{K}_G = [\mathbf{I} - \mathbf{Q}_G]$ ,  $\mathbf{Q}_G = \mathbf{U}(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{U}'\mathbf{T}$ , and  $\mathbf{d}' = [d_1 \ d_2 \ \dots \ d_r]$  is the vector of the predetermined proportional gains imposed by the breeder. It can be shown that  $\theta_G$ , the proportionality constant, can be written as

$$\theta_G = \frac{\mathbf{d}'(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{U}'\mathbf{T}\mathbf{w}}{\mathbf{d}'(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{d}}. \quad (1.10)$$

When  $\theta_G = 0$ ,  $\beta_{PG} = \beta_{RG}$ , and when  $\mathbf{U}'$  is a null matrix,  $\beta_{PG} = \mathbf{w}$ . Equations (6.8) and (1.9) give the same results, that is, both equations express the same result in a different mathematical way.

The maximized selection response and expected genetic gain per trait of the PPG-LGSI can be written as

$$R_{PG} = \frac{k_I}{L_G} \sqrt{\beta_{PG}' \Gamma \beta_{PG}} \quad (1.11)$$

and

$$\mathbf{E}_{PG} = \frac{k_I}{L_G} \frac{\Gamma \beta_{PG}}{\sqrt{\beta_{PG}' \Gamma \beta_{PG}}}, \quad (1.12)$$

respectively, where  $L_G$  is the time required to complete a selection cycle using the PPG-LGSI. Equations (1.11) and (1.12) depend only on GEBV information.

### 1.2.3 Statistical Properties of the PPG-LGSI

Assuming that  $H = \mathbf{w}'\mathbf{g}$  and the PPG-LGSI ( $I_{PG} = \boldsymbol{\beta}'_{PG}\boldsymbol{\gamma}$ ) have bivariate joint normal distribution,  $\boldsymbol{\beta}_{PG} = \mathbf{K}_P\mathbf{w}$ ;  $\boldsymbol{\Gamma}$ ,  $\mathbf{C}$ , and  $\mathbf{w}$  are known, it can be shown that PPG-LGSI has the following statistical properties:

1. The vector  $\boldsymbol{\beta}_{PG} = \mathbf{K}_P\mathbf{w}$  minimizes the mean square error under the restriction  $\mathbf{D}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$ .
2. The variance of  $I_{PG} = \boldsymbol{\beta}'_{PG}\boldsymbol{\gamma}$  ( $\sigma_{I_{PG}}^2 = \boldsymbol{\beta}'_{PG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{PG}$ ) is equal to the covariance between  $I_{PG} = \boldsymbol{\beta}'_{PG}\boldsymbol{\gamma}$  and  $H = \mathbf{w}'\mathbf{g}$  ( $\sigma_{HI_{PG}} = \mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta}_{PG}$ ).
3. The maximized correlation between  $H$  and  $I_{PG}$  (also called PPG-LGSI accuracy) is equal to  $\rho_{HI_{PG}} = \frac{\sigma_{I_{PG}}}{\sigma_H}$ , where  $\sigma_{I_{PG}} = \sqrt{\boldsymbol{\beta}'_{PG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{PG}}$  and  $\sigma_H = \sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}$  are the standard deviations of  $I_{PG} = \boldsymbol{\beta}'_{PG}\boldsymbol{\gamma}$  and  $H = \mathbf{w}'\mathbf{g}$  respectively.
4. The variance of the predicted error,  $Var(H - I_{PG}) = (1 - \rho_{HI_{PG}}^2)\sigma_H^2$ , is minimal.

The statistical PPG-LGSI properties are equal to the statistical PPG-LPSI properties, then, the PPG-LGSI is an application of the PPG-LPSI to the genomic selection context.

### 1.2.4 Numerical Example

To illustrate the PPG-LGSI theory, we use the estimated matrix  $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.21 & 2.95 & 5.00 \\ 2.95 & 42.41 & 71.11 \\ 5.00 & 71.11 & 121.53 \end{bmatrix}$  and the GEBVs associated with the traits GY (ton  $ha^{-1}$ ), EHT (cm), and PHT (cm), described in Sect. 1.1.3.

It is necessary to estimate the PPG-LGSI vector of coefficients  $\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}_{RG} + \theta_g\mathbf{U}(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{d}$  (Eqs. 1.9 and 1.10). In Sect. 1.1.3, we showed that the estimated vectors of coefficients of  $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$  for the null restrictions  $\mathbf{U}'_1 = [1 \ 0 \ 0]$  and  $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$  were  $\widehat{\boldsymbol{\beta}}'_{RG1} = \mathbf{w}'\widehat{\mathbf{K}}'_G = [3.78 \ -0.1 \ -0.1]$  and  $\widehat{\boldsymbol{\beta}}'_{RG2} = \mathbf{w}'\widehat{\mathbf{K}}'_G = [1.12 \ 0.09 \ -0.1]$  respectively, where  $\mathbf{w}' = [5 \ -0.1 \ -0.1]$ . This means that to estimate  $\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}_{RG} + \theta_G\mathbf{U}(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{d}$ , we need only to estimate  $\theta_G\mathbf{U}(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{d}$  for both sets of restrictions.

Consider matrix  $\mathbf{U}'_1 = [1 \ 0 \ 0]$  and let  $d_1 = 7.0$  be the predetermined proportional gain restriction for trait 1. We can estimate  $\theta_G$  and  $\mathbf{U}(\mathbf{U}'\mathbf{T}\mathbf{U})^{-1}\mathbf{d}$  as

$$\widehat{\theta}_{G1} = \frac{7.0(\mathbf{U}'_1\widehat{\mathbf{T}}\mathbf{U}_1)^{-1}\mathbf{U}'_1\widehat{\mathbf{w}}}{7.0(\mathbf{U}'_1\widehat{\mathbf{T}}\mathbf{U}_1)^{-1}7.0} = 0.036 \quad \text{and} \quad \mathbf{U}_1(\mathbf{U}'_1\widehat{\mathbf{T}}\mathbf{U}_1)^{-1}7.0 = \begin{bmatrix} 33.333 \\ 0 \\ 0 \end{bmatrix},$$

whence the PPG-LGSI vector of coefficients was

$$\widehat{\boldsymbol{\beta}}_{PG1} = \widehat{\boldsymbol{\beta}}_{RG1} + \widehat{\theta}_{G1}\mathbf{U}_1(\mathbf{U}'_1\widehat{\mathbf{T}}\mathbf{U}_1)^{-1}7.0 = \begin{bmatrix} 5.0 \\ -0.1 \\ -0.1 \end{bmatrix}, \text{ and the estimated PPG-LGSI}$$

was  $\hat{I}_{PG_1} = 5.0\text{GEBV}_1 - 0.1\text{GEBV}_2 - 0.1\text{GEBV}_3$ . In a similar manner, we can estimate the PPG-LGSI vector of coefficients under restrictions  $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$  and  $\mathbf{d}'_2 = [7 \quad -3]$ . In this case,

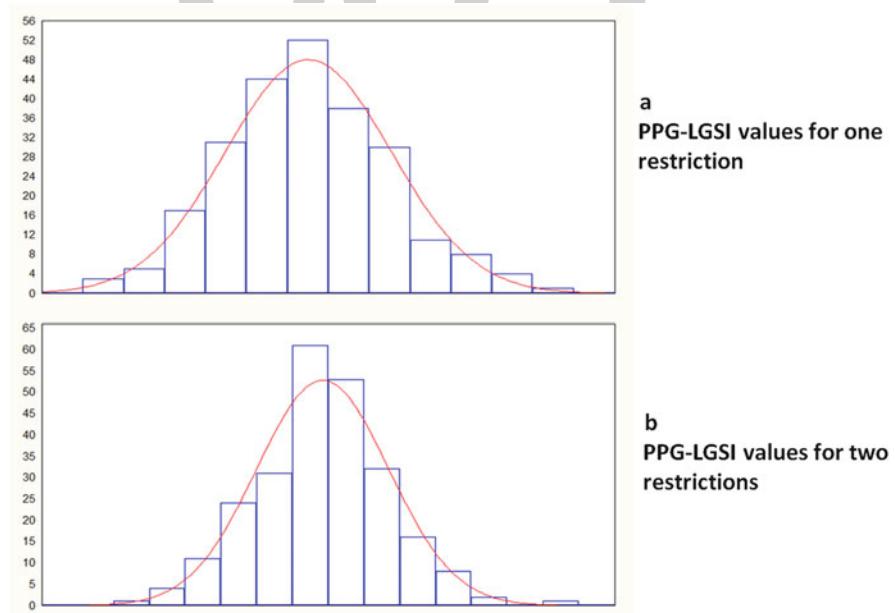
$$\hat{\beta}_{PG_2} = \hat{\beta}_{RG_2} + \hat{\theta}_{G_2} \mathbf{U}_2 (\mathbf{U}'_2 \hat{\Gamma} \mathbf{U}_2)^{-1} \mathbf{d}_2 = \begin{bmatrix} 4.97 \\ -0.18 \\ -0.10 \end{bmatrix} \text{ and the estimated PPG-LGSI}$$

$$\text{was } \hat{I}_{PG_2} = 4.97\text{GEBV}_1 - 0.18\text{GEBV}_2 - 0.1\text{GEBV}_3.$$

Figure 1.2 presents the frequency distribution of the estimated PPG-LGSI values for one (Fig. 1.2a) and two (Fig. 1.2b) predetermined restrictions,  $d = 7$  and  $\mathbf{d}' = [7 \quad -3]$  respectively, obtained in a real testing population for one selection cycle in one environment. For both restrictions, the frequency distribution of the estimated PPG-LGSI values approaches the normal distribution.

Assume a selection intensity of 10% ( $k_{IG} = 1.755$ ); then, for one predetermined restriction, where  $\mathbf{U}'_1 = [1 \quad 0 \quad 0]$  and  $d_1 = 7.0$ , the estimated PPG-LGSI selection response and expected genetic gain per trait, not including the interval length, were

$$\hat{R}_{PG_1} = k_{IG} \sqrt{\hat{\beta}'_{PG_1} \hat{\Gamma} \hat{\beta}_{PG_1}} = 1.05 \quad \text{and} \quad \hat{\mathbf{E}}'_{PG_1} = k_I \frac{\hat{\beta}'_{PG_1} \hat{\Gamma}}{\sqrt{\hat{\beta}'_{PG_1} \hat{\Gamma} \hat{\beta}_{PG_1}}} = [0.74 \quad 9.92 \quad 16.54]$$



**Fig. 1.2** Distribution of 380 estimated predetermined proportional gain linear genomic selection index (PPG-LGSI) values with one (a) and two (b) predetermined restrictions,  $d = 7$  and  $\mathbf{d}' = [7 \quad -3]$  respectively, obtained in a real testing population for one selection cycle in one environment

respectively. For two restrictions, with  $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$  and  $\mathbf{d}' = [7 \ -3]$ , the estimated RLGSI selection response and expected genetic gains, not including the interval length, were  $\widehat{R}_{PG_2} = k_{IG} \sqrt{\widehat{\beta}'_{PG_2} \widehat{\Gamma} \widehat{\beta}_{G_2}} = 0.52$  and  $\widehat{\mathbf{E}}'_{PG_2} = k_I \frac{\widehat{\beta}'_{PG_2} \widehat{\Gamma}}{\sqrt{\widehat{\beta}'_{PG_2} \widehat{\Gamma} \widehat{\beta}_{PG_2}}} = [0.11 \ -0.05 \ 0.14]$  respectively.

Now, we use the simulated data set described in Chap. 2, Sect. 2.8.1 to compare PPG-LGSI efficiency versus predetermined proportional gain linear phenotypic selection index (PPG-LPSI) efficiency. Let  $\mathbf{U}'_1 = [1 \ 0 \ 0]$ ,  $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ , and  $\mathbf{U}'_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$  be the matrices and  $d'_1 = 7$ ,  $\mathbf{d}'_2 = [7 \ -3]$ , and  $\mathbf{d}'_3 = [7 \ -3 \ 5]$  the vectors for one, two, and three predetermined restrictions respectively. Table 1.5 presents the estimated PPG-LPSI and PPG-LGSI selection response for each predetermined restriction in five simulated selection cycles including and not including the interval between selection cycles (4 years for the PPG-LPSI and 1.5 years for the PPG-LGSI); estimated PPG-LPSI and PPG-LGSI accuracy; and estimated variance of the predicted error (VPE). In each selection cycle, the sample size was equal to 500 genotypes, each with four repetitions and four traits. The selection intensity was 10% ( $k_J = 1.755$ ).

The averages of the estimated PPG-LPSI selection response not including the interval length were 15.14, 14.87, and 13.30, whereas when the interval length was included, the average selection responses were 3.79, 3.72, and 3.33, for one, two, and three predetermined restrictions respectively (Table 1.5). The averages of the estimated PPG-LGSI selection responses not including the interval length for one, two, and three predetermined restrictions were 14.48, 13.47, and 11.26 respectively, and when the interval length was included, the selection responses were 9.65, 8.98, and 7.51 respectively (Table 1.5). These results indicate that when the interval length was included in the estimation of the PPG-LPSI and PPG-LGSI selection responses, PPG-LGSI efficiency was greater than PPG-LPSI efficiency, and vice versa, when the interval length was not included in the PPG-LPSI and PPG-LGSI selection responses, PPG-LPSI efficiency was higher than PPG-LGSI efficiency.

The averages of the estimated VPE values of the PPG-LPSI for one, two, and three predetermined restrictions were 22.42, 30.56, and 41.17 respectively, whereas the estimated VPE values of the PPG-LGSI (see Sect. 6.2.3 for details) were 59.80, 66.95, and 83.98, respectively, that is, in all selection cycles, the VPE of the PPG-LPSI was lower than that of the PPG-LGSI. This means that for this data set, the PPG-LPSI was a better predictor of the net genetic merit than the PPG-LGSI. These results can be explained by observing that the averages of the estimated PPG-LPSI accuracies were 0.88, 0.86, and 0.77, whereas the estimated PPG-LGSI accuracies were 0.65, 0.68, and 0.57 for each predetermined restriction, that is, the estimated PPG-LGSI accuracies were lower than the estimated PPG-LPSI accuracies for this data set.



**Table 1.5** Estimated predetermined proportional gain linear phenotypic and genomic selection index (PPG-LPSI and PPG-LGSI respectively) selection responses for 1, 2 and 3 predetermined restrictions for five simulated selection cycles including and not including the interval between selection cycles (4 years for the PPG-LPSI and 1.5 years for the PPG-LGSI); estimated PPG-LPSI and PPG-LGSI accuracy and estimated variance of the predicted error (VPE)

Estimated PPG-LPSI selection response								
Cycle	Not including interval length			Including interval length <sup>a</sup>			Estimated PPG-LPSI accuracy	Estimated PPG-LPSI VPE
	1	2	3	1	2	3		
1	17.81	16.72	15.23	4.45	4.18	3.81	0.91	0.77
2	15.69	15.59	14.39	3.92	3.90	3.60	0.88	0.81
3	14.22	14.16	13.18	3.56	3.54	3.30	0.87	0.80
4	14.34	14.33	11.56	3.59	3.58	2.89	0.86	0.70
5	13.64	13.56	12.16	3.41	3.39	3.04	0.86	0.76
Average	15.14	14.87	13.30	3.79	3.72	3.33	0.88	0.77

Estimated PPG-LGSI selection response								
Cycle	Not interval length			Including interval length			Estimated PPG-LGSI accuracy	Estimated PPG-LGSI VPE
	1	2	3	1	2	3		
1	21.22	17.83	16.66	14.15	11.89	11.10	0.65	0.91
2	13.90	13.52	10.65	9.27	9.01	7.10	0.72	0.70
3	13.59	13.15	10.70	9.06	8.77	7.13	0.70	0.67
4	12.30	11.84	9.36	8.20	7.89	6.24	0.61	0.59
5	11.38	11.03	8.96	7.59	7.35	5.97	0.56	0.54
Average	14.48	13.47	11.26	9.65	8.98	7.51	0.65	0.68

<sup>a</sup>The estimated LPSI selection response was divided by 4 and the estimated LGS selection response was divided by 1.5

**Table 1.6** Estimated PPG-LPSI heritability ( $\widehat{h}_P^2$ ), values of  $W_P = \frac{\widehat{\rho}_{H_{LG}}}{\widehat{h}_P} L_P$  ( $L_P = 4$ ), and the ratio of the estimated PPG-LPSI accuracy ( $\widehat{\rho}_{H_P}$ ) to the estimated PPG-LGSI accuracy ( $\widehat{\rho}_{H_{PG}}$ ):  $\widehat{\lambda}_P = \widehat{\rho}_{H_P}/\widehat{\rho}_{H_{PG}}$ , and values of  $\widehat{p} = 100(\widehat{\lambda}_P - 1)$  for 1, 2 and 3 predetermined restrictions for five simulated selection cycles

Cycle	PPG-LPSI heritability			Values of $W_P$			Estimated ratio values ( $\widehat{p}$ )		
	1	2	3	1	2	3	1	2	3
1	0.84	0.77	0.83	4.71	4.13	3.72	-18.62	-6.71	-10.20
2	0.80	0.78	0.83	3.22	3.17	2.42	18.30	20.54	32.04
3	0.77	0.76	0.8	3.18	3.09	2.45	19.89	21.59	31.42
4	0.76	0.75	0.78	2.80	2.71	2.10	29.16	31.84	33.75
5	0.75	0.75	0.79	2.57	2.49	1.97	35.26	36.55	42.35
Average	0.72	0.71	0.76	3.29	3.12	2.53	16.80	20.76	25.87

Table 1.6 presents the estimated predetermined PPG-LPSI heritability ( $\widehat{h}_P^2$ ) values,  $W_P = \frac{\widehat{\rho}_{H_{LG}}}{\widehat{h}_P} L_P$  ( $L_P = 4$ ) values, and ratio of the estimated PPG-LPSI accuracy ( $\widehat{\rho}_{H_P}$ ) to the estimated PPG-LGSI accuracy ( $\widehat{\rho}_{H_{PG}}$ ), i.e.,  $\widehat{\lambda}_P = \widehat{\rho}_{H_P}/\widehat{\rho}_{H_{PG}}$ , and, finally, values of  $\widehat{p} = 100(\widehat{\lambda}_P - 1)$  for one, two, and three null restrictions for five simulated selection cycles.

The averages of the  $W_P$  values for one, two, and three null restrictions were 3.29, 3.12, and 2.53, respectively, whereas the PPG-LGSI interval length was 1.5 ( $L_G = 1.5$ ). This means that the estimated Technow inequality,  $L_G < \frac{\widehat{\rho}_{H_{LG}}}{\widehat{h}_P} L_P$  (see Chap. 5, Eq. 5.18) was true. Thus, PPG-LGSI efficiency in terms of time was greater than PPG-LPSI efficiency for this data set. These results coincide with those obtained earlier in this chapter, when we compared PPG-LGSI efficiency versus PPG-LPSI efficiency in terms of interval length. However, the average values of  $\widehat{p} = 100(\widehat{\lambda}_P - 1)$  (see Chap. 5) were, in percentage terms, 16.80%, 20.76%, and 25.85% for each restriction. These latter results indicate that for this data set, the PPG-LPSI was a better predictor of the net genetic merit than the PPG-LGSI. This is because the estimated PPG-LPSI accuracies were higher than the estimated PPG-LPSI accuracies for this data set. We found similar results when we compared the PPG-LPSI VPE versus PPG-LGSI VPE (Table 1.5).

### 1.3 The Combined Restricted Linear Genomic Selection Index

The combined restricted linear genomic selection index (CRLGSI) is based on the RLPSI (Chap.) and combined linear genomic selection index (CLGSI, Chap.) theory. In the RLPSI, 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 the CRLGSI, but in the latter case, it is necessary to impose  $2r$  restrictions, i.e., we need to fix  $r$  traits and their associated  $r$  GEBVs to obtain results similar to those obtained with the RLPSI. This is the main difference between the CRLGSI and the RLPSI.

It can be shown that  $\text{Cov}(\mathbf{I}_C, \mathbf{a}_C) = \Psi_C \boldsymbol{\beta}_C$  is the covariance between the breeding value vector ( $\mathbf{a}'_C = [\mathbf{g}' \quad \boldsymbol{\gamma}']$ ) and the CLGSI,  $I_C = \boldsymbol{\beta}'_C \mathbf{t}_C$  (see Chap. 5 for details), where  $\mathbf{t}'_C = [\mathbf{y}' \quad \boldsymbol{\gamma}']$ . In the CRLGSI, we want some covariances between the linear combinations of  $\mathbf{a}_C$  ( $\mathbf{U}'_C \mathbf{a}_C$ ) and CLGSI to be zero, i.e.,  $\text{Cov}(\mathbf{I}_C, \mathbf{U}'_C \mathbf{a}_C) = \mathbf{U}'_C \Psi_C \boldsymbol{\beta}_C = \mathbf{0}$ , where  $\mathbf{U}'_C$  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 that the trait and its GEBV have no restrictions) and  $\Psi_C = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$  is a block covariance matrix of  $\mathbf{a}'_C = [\mathbf{g}' \quad \boldsymbol{\gamma}']$  where  $\mathbf{C}$  and  $\boldsymbol{\Gamma}$  are the covariance matrices of breeding ( $\mathbf{g}$ ) and genomic ( $\boldsymbol{\gamma}$ ) values respectively. This problem can be solved by minimizing the mean squared difference between the CLGSI and  $H$  ( $E[(H - I_C)^2]$ ) under the restriction  $\mathbf{U}'_C \Psi_C \boldsymbol{\beta}_C = \mathbf{0}$  similar to the RLGSI in Sect. 1.1.

### 1.3.1 The Maximized CRLGSI Parameters

Let  $\mathbf{T}_C = \begin{bmatrix} \mathbf{P} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$  be the block covariance matrix of  $\mathbf{t}'_C = [\mathbf{y}' \quad \boldsymbol{\gamma}']$  where  $\mathbf{P}$  and  $\boldsymbol{\Gamma}$  are the covariance matrices of phenotypic ( $\mathbf{y}$ ) and genomic ( $\boldsymbol{\gamma}$ ) values respectively. Based on the Eq. (1.1) result, it can be shown that the CRLGSI vector of coefficients that minimizes  $E[(H - I_C)^2]$  under the restriction  $\mathbf{U}'_C \Psi_C \boldsymbol{\beta}_C = \mathbf{0}$  is

$$\boldsymbol{\beta}_{CR} = \mathbf{K}_C \boldsymbol{\beta}_C, \quad (1.13)$$

where  $\mathbf{K}_C = [\mathbf{I} - \mathbf{Q}_C]$ ,  $\mathbf{Q}_C = \mathbf{T}_C^{-1} \Phi_C (\Phi'_C \mathbf{T}_C^{-1} \Phi_C)^{-1} \Phi'_C$ ,  $\Phi_C = \mathbf{U}'_C \Psi_C$ , and  $\boldsymbol{\beta}_C = \mathbf{T}_C^{-1} \Psi_C \mathbf{a}_C$  (the vector of coefficients of the CLGSI, see Chap. 5 for details);  $\mathbf{T}_C^{-1}$  is the inverse of matrix  $\mathbf{T}_C$ , and  $\mathbf{I}$  is an identity matrix  $2t \times 2t$ . When no restrictions are imposed on any of the traits,  $\mathbf{U}'_C$  is a null matrix and  $\boldsymbol{\beta}_{CR} = \boldsymbol{\beta}_C$  (the vector of coefficients of the CLGSI). That is, the CRLGSI is more general than the CLGSI. Similar to the RLPSI and the RLGSI, matrices  $\mathbf{K}_C$  and  $\mathbf{Q}_C$  are idempotent ( $\mathbf{K}_C = \mathbf{K}_C^2$  and  $\mathbf{Q}_C = \mathbf{Q}_C^2$ ) and orthogonal ( $\mathbf{K}_C \mathbf{Q}_C = \mathbf{Q}_C \mathbf{K}_C = \mathbf{0}$ ), that is,  $\mathbf{K}_C$  and  $\mathbf{Q}_C$  are projectors. Thus, we can assume that the CRLGSI has similar properties to those described for the RLPSI (see Chap. for details)

when matrices  $\Psi_C = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$  and  $\mathbf{T}_C = \begin{bmatrix} \mathbf{P} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$  are known.

The maximized selection response and the optimized expected genetic gain per trait of the CRLGSI can be written as



$$R_{CR} = \frac{k_I}{L_I} \sqrt{\boldsymbol{\beta}'_{CR} \mathbf{T}_C \boldsymbol{\beta}_{CR}} \quad (1.14)$$

and

$$\mathbf{E}_{CR} = \frac{k_I}{L_I} \frac{\boldsymbol{\Psi} \boldsymbol{\beta}_{CR}}{\sqrt{\boldsymbol{\beta}'_{CR} \mathbf{T}_C \boldsymbol{\beta}_{CR}}}, \quad (1.15)$$

respectively. Although in the RLGSI and the PPG-LGSI the interval between selection cycles is denoted as  $L_G$ , in the CRLGSI it is denoted as  $L_I$ . This is because the RLPSI and the CRLGSI should have the same interval between selection cycles.

### 1.3.2 Numerical Examples

To illustrate the CRLGSI theoretical results, we use a real training 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). Matrices  $\mathbf{P}$  and  $\mathbf{C}$  were estimated based on Eqs. (2.22) to (2.24) described in Chap. The

estimated matrices were  $\hat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}$  and

$\hat{\mathbf{C}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$ . In a similar manner, we estimated matrix  $\Gamma$  using

Eqs. (5.21) to (5.23)s decribed in Chap. The estimated matrix was

$$\hat{\Gamma} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}.$$

To estimate the CRLGSI and its associated parameters (selection response, expected genetic gain per trait, etc.), we need to obtain matrices  $\hat{\mathbf{T}}_C = \begin{bmatrix} \hat{\mathbf{P}} & \hat{\Gamma} \\ \hat{\Gamma} & \hat{\Gamma} \end{bmatrix}$

and  $\hat{\Psi}_C = \begin{bmatrix} \hat{\mathbf{C}} & \hat{\Gamma} \\ \hat{\Gamma} & \hat{\Gamma} \end{bmatrix}$  using phenotypic and genomic information and the estimated CRLGSI vector of coefficients  $\hat{\boldsymbol{\beta}}_{CR} = \hat{\mathbf{K}}_C \hat{\boldsymbol{\beta}}_C$ , where  $\hat{\mathbf{K}}_C = [\mathbf{I} - \hat{\mathbf{Q}}_C]$ ,  $\hat{\mathbf{Q}}_C = \hat{\mathbf{T}}_C^{-1} \hat{\Phi}_C (\hat{\Phi}'_C \hat{\mathbf{T}}_C^{-1} \hat{\Phi}_C)^{-1} \hat{\Phi}'_C$ ,  $\hat{\Phi}_C = \mathbf{U}'_C \hat{\Psi}_C$ , and  $\hat{\boldsymbol{\beta}}_C = \hat{\mathbf{T}}_C^{-1} \hat{\Psi}_C \mathbf{a}_C$ .

We have indicated that the main difference between the RLGSI and the CRLGSI is 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 (*Zea mays*)  $F_2$  population described earlier and suppose that we restrict trait GY; then, matrix  $\mathbf{U}'_C$  should be constructed as  $\mathbf{U}'_{C_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ . If we restrict traits GY and EHT, matrix  $\mathbf{U}'_C$  should

be constructed as  $\mathbf{U}'_{C_2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ , etc. The procedure for obtaining

matrices  $\widehat{\mathbf{K}}_C = [\mathbf{I} - \widehat{\mathbf{Q}}_C]$ ,  $\widehat{\mathbf{Q}}_C = \widehat{\mathbf{T}}_C^{-1}\widehat{\Phi}_C(\widehat{\Phi}'_C\widehat{\mathbf{T}}_C^{-1}\widehat{\Phi}_C)^{-1}\widehat{\Phi}'_C$ , and  $\widehat{\Phi}_C = \mathbf{U}'_C\widehat{\Psi}_C$  is similar to that described in Chap. 3.

Let  $\mathbf{w}' = [5 \quad -0.1 \quad -0.1 \quad 0 \quad 0 \quad 0]$  be the vector of economic weights and assume that we restrict trait GY; in this case, according to the estimated matrices  $\widehat{\mathbf{P}}$ ,  $\widehat{\mathbf{C}}$ , and  $\widehat{\Gamma}$  described earlier, the estimated CRLGSI vector of coefficients was  $\widehat{\beta}'_{RG} = [0.076 \quad -0.004 \quad -0.018 \quad 2.353 \quad -0.096 \quad -0.082]$ , whence the estimated CRLGSI can be written as

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

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 two or more restrictions.

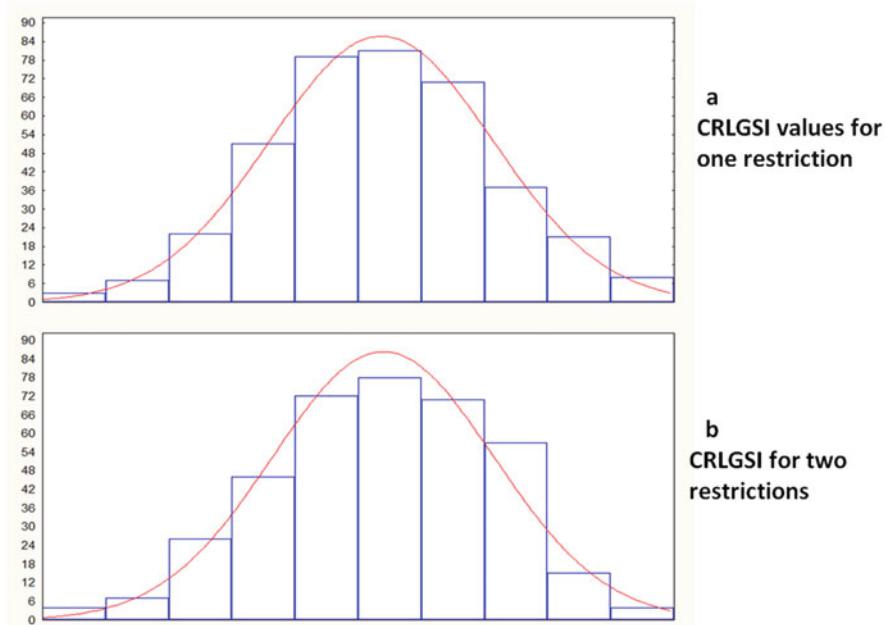
Figure 1.3 presents the frequency distribution of the estimated CRLGSI values for one (Fig. 1.3a) and two null restrictions (Fig. 1.3b) using matrices  $\mathbf{U}'_{C_1}$  and  $\mathbf{U}'_{C_2}$ , and the real data set of the  $F_2$  population. For both restrictions, the frequency distribution of the estimated CRLGSI values approaches normal distribution.

Suppose a selection intensity of 10% ( $k_I = 1.755$ ), matrix  $\mathbf{U}'_{C_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$  and that the vector of economic weights is  $\mathbf{w}' = [5 \quad -0.1 \quad -0.1 \quad 0 \quad 0 \quad 0]$ ; then, according to the estimated matrices  $\widehat{\mathbf{P}}$ ,  $\widehat{\mathbf{C}}$ , and  $\widehat{\Gamma}$  described earlier, the estimated CRLGSI selection response and the estimated CRLGSI expected genetic gain per trait were  $\widehat{R}_{CR} = k_I \sqrt{\widehat{\beta}'_{CR}\widehat{\mathbf{T}}_C\widehat{\beta}_{CR}} = 0.96$  and  $\widehat{\mathbf{E}}'_{CR} = k_I \frac{\widehat{\beta}'_{CR}\widehat{\Psi}}{\sqrt{\widehat{\beta}'_{CR}\widehat{\mathbf{T}}_C\widehat{\beta}_{CR}}} = [0 \quad -3.53 \quad -6.03 \quad 0 \quad -2.93 \quad -4.87]$  respectively,

whereas the estimated CRLGSI accuracy was  $\widehat{\rho}_{HI_{CR}} = \frac{\widehat{\sigma}_{I_{CR}}}{\widehat{\sigma}_H} = 0.51$

Now, we use the simulated data described in Chapter, Sect. 2.8.1 to compare CRLGSI efficiency versus RLGSI efficiency. The criteria for this comparison are the Technow inequality and the ratio of the estimated CRLGSI accuracy ( $\widehat{\rho}_{HI_{CR}}$ ) to the estimated RLGSI accuracy ( $\widehat{\rho}_{HI_R}$ ) expressed as percentages i.e.,  $\widehat{p} = 100(\widehat{\lambda}_{CR} - 1)$ , where  $\widehat{\lambda}_P = \widehat{\rho}_{HI_{CR}}/\widehat{\rho}_{HI_R}$ , for one, two, and three null restrictions for five simulated selection cycles.

Table 1.7 presents the estimated CRLGSI heritability ( $\widehat{h}_C^2$ ), the estimated RLGSI accuracy ( $\widehat{\rho}_{HI_R}$ ), the values of  $W_C = \frac{\widehat{\rho}_{HI_R}}{\widehat{h}_I} L_I$  ( $L_I = 4$ ), and the values of  $h_I$



**Fig. 1.3** Distribution of 244 estimated combined restricted linear genomic selection index (CRLGSI) values with one (a) and two (b) null restrictions respectively obtained in a real training population for one selection cycle in one environment

$\hat{p} = 100(\hat{\lambda}_{CR} - 1)$ , where  $\hat{\lambda}_{CR} = \hat{\rho}_{HI_{CR}}/\hat{\rho}_{HI_R}$  and  $\hat{\rho}_{HI_{CR}}$  is the estimated CRLGSI accuracy, for one, two, and three null restrictions for five simulated selection cycles. The averages of the  $W_C = \frac{\hat{\rho}_{HI_R}}{\hat{\rho}_{HI_{CR}}} L_I$  values for one, two, and three null restrictions were 1.26, 0.92, and 0.59 respectively, whereas the RLGSI interval length was 1.5 ( $L_G = 1.5$ ). This means that the estimated Technow inequality ( $L_G < \frac{\hat{\rho}_{HI_G}}{\hat{\rho}_{HI_{CR}}} L_I$ ) was not true. Thus, for this data set, RLGSI efficiency in terms of time is not greater than CRLGSI efficiency. The inequality  $L_G < \frac{\hat{\rho}_{HI_G}}{\hat{\rho}_{HI_{CR}}} L_I$  was not true because the estimated RLGSI accuracy was very low, whereas CRLGSI heritability was high. Thus, note that the averages of the estimated RLGSI accuracy for one, two, and three null restrictions were 0.25, 0.19, and 0.14 respectively, whereas the averages of the estimated CRLGSI heritability values were 0.72, 0.75, and 0.89 respectively. Thus, according to these results, when the estimated RLGSI accuracy is very low and the estimated CRLGSI heritability is high, RLGSI efficiency will be lower than CRLGSI efficiency in terms of time.



**Table 1.7** Estimated combined restricted linear genomic selection index (CRLGSI) heritability ( $\hat{h}_I^2$ ), estimated RLGSI accuracy ( $\hat{\rho}_{H_{Ik}}$ ), values of  $W_C = \frac{\hat{\rho}_{H_{Ik}} L_I}{\hat{h}_I}$  ( $L_I = 4$ ), and values of  $\hat{p} = 100(\hat{\lambda}_{CR} - 1)$ , where  $\hat{\lambda}_{CR} = \hat{\rho}_{H_{ICR}}/\hat{\rho}_{H_{Ik}}$  and  $\hat{\rho}_{H_{ICR}}$  is the estimated CRLGSI accuracy, for 1, 2, and 3 null restrictions for five simulated selection cycles.

Cycle	CRLGSI heritability			RLGSI accuracy			Values of $W_C$			Values of $\hat{p}$		
	1	2	3	1	2	3	1	2	3	1	2	3
1	0.39	0.41	0.90	0.33	0.28	0.24	2.11	1.75	1.01	15.15	50.00	8.33
2	0.82	0.84	0.89	0.26	0.18	0.13	1.15	0.79	0.55	46.15	72.22	61.54
3	0.80	0.84	0.88	0.24	0.17	0.11	1.07	0.74	0.47	66.67	82.35	81.82
4	0.82	0.85	0.89	0.22	0.15	0.11	0.97	0.65	0.47	68.18	93.33	72.73
5	0.77	0.82	0.91	0.22	0.15	0.11	1.00	0.66	0.46	72.73	93.33	81.82
Average	0.72	0.75	0.89	0.25	0.19	0.14	1.26	0.92	0.59	53.78	78.25	61.25



The last three columns of Table 1.7, from left to right, present the average of the values of  $\hat{p} = 100(\hat{\lambda}_{CR} - 1)$ , for one, two, and three null restrictions of five simulated selection cycles. According to these results, CRLGSI efficiency was 53.78%, 78.25%, and 61.25% higher than RLGSI efficiency. Thus, for this data set, the CRLGSI was a better predictor of the net genetic merit than the RLGSI.

## 1.4 The Combined Predetermined Proportional Gains Linear Genomic Selection Index

In the PPG-LPSI described in Chap. the vector of the PPG (predetermined proportional gains) was  $\mathbf{d}' = [d_1 \ d_2 \ \dots \ d_r]$ . However, because the combined predetermined proportional gains LGSI (CPPG-LGSI) uses phenotypic and GEBV information jointly to predict the net genetic merit, the vector of the PPG ( $\mathbf{d}_C$ ) should be twice the standard vector  $\mathbf{d}'$ , that is,  $\mathbf{d}'_C = [d_1 \ d_2 \ \dots \ d_r \ d_{r+1} \ d_{r+2} \ \dots \ d_{2r}]$ , 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. In addition, in the CPPG-LGSI, 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-LPSI described in Chap. and the CPPG-LGSI.

### 1.4.1 The Maximized CPPG-LGSI Parameters

It can be shown that the vector of coefficients of the CPPG-LGSI can be written as

$$\boldsymbol{\beta}_{CP} = \boldsymbol{\beta}_{CR} + \theta_{CP} \boldsymbol{\delta}_{CP}, \quad (1.16)$$

where

$$\theta_{CP} = \frac{\boldsymbol{\beta}'_C \boldsymbol{\Phi}_C (\boldsymbol{\Phi}'_C \widehat{\mathbf{T}}_C^{-1} \boldsymbol{\Phi}_C)^{-1} \mathbf{d}_C}{\mathbf{d}'_C (\boldsymbol{\Phi}'_C \widehat{\mathbf{T}}_C^{-1} \boldsymbol{\Phi}_C)^{-1} \mathbf{d}_C} \quad (1.17)$$

is a proportionality constant. In addition, in Eq. (1.16),  $\boldsymbol{\beta}_{CR} = \mathbf{K}_C \boldsymbol{\beta}_C$  is the vector of coefficients of the CRLGSI (Eq. 1.13),  $\boldsymbol{\delta}_{CP} = \mathbf{T}_C^{-1} \boldsymbol{\Phi}_C (\boldsymbol{\Phi}'_C \widehat{\mathbf{T}}_C^{-1} \boldsymbol{\Phi}_C)^{-1} \mathbf{d}_C$ ,  $\boldsymbol{\Phi}'_C = \mathbf{U}'_C \boldsymbol{\Psi}_C$ , and  $\boldsymbol{\beta}_C = \mathbf{T}_C^{-1} \boldsymbol{\Psi}_C \mathbf{a}_C$  (the vector of coefficients of the CLGSI). When  $\theta_{CP} = 0$ ,  $\boldsymbol{\beta}_{CP} = \boldsymbol{\beta}_{CR}$ , and if  $\theta = 0$  and  $\mathbf{U}'_C$  is the null matrix, then  $\boldsymbol{\beta}_{CR} = \boldsymbol{\beta}_C$ .

Thus, the CPPG-LGSI is more general than the CRLGSI and the CLGSI, and includes the latter two indices as particular cases. In addition, it can be shown that the CPPG-LGSI has the same properties as the PPG-LPSI described in Chap. 3.

The maximized selection response and the expected genetic gain per trait of the CPPG-LGSI can be written as

$$R_{CP} = \frac{k_I}{L_I} \sqrt{\boldsymbol{\beta}'_{CP} \mathbf{T}_C \boldsymbol{\beta}_{CP}} \quad (1.18)$$

and

$$\mathbf{E}_{CP} = \frac{k_I}{L_I} \frac{\Psi \boldsymbol{\beta}_{CP}}{\sqrt{\boldsymbol{\beta}'_{CP} \mathbf{T}_C \boldsymbol{\beta}_{CP}}}, \quad (1.19)$$

respectively. Although in the RLGSI and the PPG-LGSI the interval between selection cycles is denoted as  $L_G$ , in the CPPG-LGSI it is denoted as  $L_I$ . This is because the RLPSI and the CPPG-LGSI should have the same interval between selection cycles because they use phenotypic information to predict the net genetic merit.

### 1.4.2 Numerical Examples

Similar to the CRLGSI, to illustrate the CPPG-LGSI results we use the real training maize (*Zea mays*)  $F_2$  population with 248 genotypes, 233 molecular markers, and three traits—GY (ton  $ha^{-1}$ ), EHT (cm), and PHT

(cm)—where  $\hat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}$ ,  $\hat{\mathbf{C}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$ , and

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

respectively.

We can obtain the estimated CPPG-LGSI vector of coefficients as  $\hat{\boldsymbol{\beta}}_{CP} = \hat{\boldsymbol{\beta}}_{CR} + \hat{\theta}_{CP} \hat{\boldsymbol{\delta}}_{CP}$  (Eq. 1.16). Suppose that we restrict trait GY and its associated GEBV with matrix  $\mathbf{U}'_{C1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$  and the vector of predetermined restriction  $\mathbf{d}'_C = [7 \ 3.5]$ . In Sect. 1.3.2, we showed that the estimated CRLGSI vector of coefficients was  $\hat{\boldsymbol{\beta}}'_{CR} = [0.076 \ -0.004 \ -0.018 \ 2.353 \ -0.096 \ -0.082]$ ; then, we only need to calculate  $\hat{\theta}_{CP}$  and  $\hat{\boldsymbol{\delta}}_{CP}$  to obtain the vector of coefficients  $\hat{\boldsymbol{\beta}}_{CP}$ .

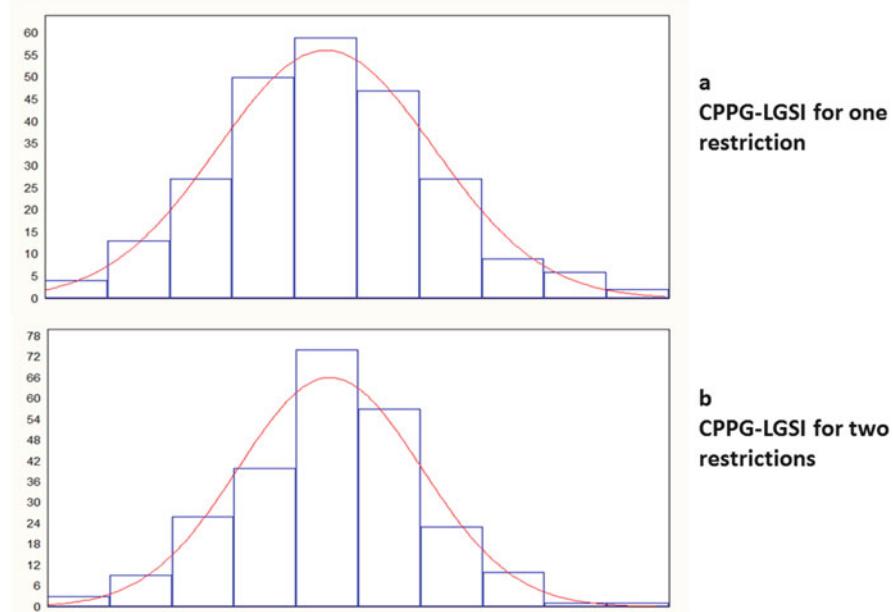
Let  $\mathbf{w}' = [5 \ -0.1 \ -0.1 \ 0 \ 0 \ 0]$  be the vector of economic weights. It can be shown that  $\hat{\theta}_{CP} = 0.00030$  is the estimated value of the proportionality constant and  $\hat{\boldsymbol{\delta}}_{CP} = [0.56 \ -77.28 \ 40.89 \ 49.44 \ 77.28 \ -40.89]$ . Thus, the estimated CPPG-LGSI vector of coefficients was  $\hat{\boldsymbol{\beta}}'_{CR} = [0.76 \ -0.030 \ -0.004 \ 2.369 \ -0.070 \ -0.096]$ , whence the estimated CPPG-LGSI can be written as

$$\hat{I}_{CP} = 0.076GY - 0.03EHT - 0.004PHT + 2.369GEBV_{GY} - 0.070GEBV_{EHT} - 0.096GEBV_{PHT},$$

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 two or more restrictions. Note that because  $\hat{\theta}_{CP} = 0.0003$  is very small, the estimated CPPG-LGSI and CRLGSI values were very similar.

Figure 1.4 presents the frequency distribution of the estimated CPPG-LGSI values for one (Fig. 1.4a) and two predetermined restrictions (Fig. 1.4b) using

matrices  $\mathbf{U}'_{C_1}$  and  $\mathbf{U}'_{C_2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ , the vectors of the PPG



**Fig. 1.4** Distribution of 244 estimated combined predetermined proportional gain linear genomic selection index (CPPG-LGSI) values with one (a) and two (b) predetermined restrictions,  $d = 7$  and  $\mathbf{d}' = [7 \ -3]$  respectively, obtained in a real training population for one selection cycle in one environment

$\mathbf{d}'_{C1} = [7 \ 3.5]$  and  $\mathbf{d}'_{C2} = [7 \ -3 \ 3.5 \ -1.5]$ , and the real data set F<sub>2</sub>. For both restrictions, the frequency distribution of the estimated CPPG-LGSI values approaches normal distribution.

Suppose a selection intensity of 10% ( $k_I = 1.755$ ) and that we restrict trait GY and its associated GEBV. The estimated CPPG-LGSI selection response and expected genetic gain per trait were  $\widehat{R}_{CP} = k_I \sqrt{\widehat{\mathbf{P}}'_{CP} \widehat{\mathbf{T}}_C \widehat{\mathbf{P}}_{CP}} = 0.98$  and  $\widehat{\mathbf{E}}'_{CP} = k_I \frac{\widehat{\mathbf{P}}'_{CP} \widehat{\Psi}}{\sqrt{\widehat{\mathbf{P}}'_{CP} \widehat{\mathbf{T}} \widehat{\mathbf{P}}_{CP}}} = [0.007 \ -3.647 \ -5.760 \ 0.004 \ -2.829 \ -4.711]$  respectively, whereas the estimated CPPG-LGSI accuracy was  $\widehat{\rho}_{HI_{CP}} = \frac{\widehat{\sigma}_{I_{CP}}}{\widehat{\sigma}_H} = 0.52$ . Once again, because  $\widehat{\theta}_{CP} = 0.0003$ , the latter results are very similar to the CRLGSI results.

Now, we use the simulated data described in Chap. to compare CPPG-LGSI efficiency versus PPG-LGSI efficiency. The criteria for this comparison are the Technow inequality (Chapter.) and the ratio of CPPG-LGSI accuracy ( $\rho_{HI_{CP}}$ ) to PPG-LGSI accuracy ( $\rho_{HI_p}$ ) expressed as percentages (Chapter)  $\widehat{p} = 100(\lambda_{CP} - 1)$ , where  $\lambda_{CP} = \widehat{\rho}_{HI_{CP}}/\widehat{\rho}_{HI_p}$  for one, two, and three null restrictions in five simulated selection cycles.

Table 1.8 presents the estimated CPPG-LGSI heritability ( $\widehat{h}_I^2$ ), the estimated PPG-LGSI accuracy ( $\widehat{\rho}_{HI_{CP}}$ ), values of  $W_{CP} = \frac{\widehat{\rho}_{HI_G}}{\widehat{h}_I} L_I$  ( $L_I = 4$ ) and  $\widehat{p} = 100(\widehat{\lambda}_{CP} - 1)$ , where  $\widehat{\lambda}_p = \widehat{\rho}_{HI_{CP}}/\widehat{\rho}_{HI_p}$  and  $\widehat{\rho}_{HI_p}$  is the estimated CPPG-LGSI accuracy, for one, two, and three null restrictions in five simulated selection cycles. The averages of the estimated  $W_{CP}$  values for one, two, and three predetermined restrictions were 3.60, 3.31, and 2.50 respectively, whereas the PPG-LGSI interval length was 1.5 ( $L_G = 1.5$ ). This means that the estimated Technow inequality,  $L_G < \frac{\widehat{\rho}_{HI_G}}{\widehat{h}_I} L_I$ , was true. Thus, for this data set, PPG-LGSI efficiency is greater than CPPG-LGSI efficiency in terms of time.

The last three columns of Table 1.8, from left to right, present the values of  $\widehat{p} = 100(\widehat{\lambda}_{CP} - 1)$ , for one, two, and three null restrictions in five simulated selection cycles. The average values of  $\widehat{p} = 100(\widehat{\lambda}_{CP} - 1)$  for each of the three restrictions, in percentage terms, were 37.19%, 32.82%, and 37.08% respectively. This means that the CPPG-LGSI efficiency was greater than PPG-LGSI efficiency at predicting the net genetic merit.

**Table 1.8** Estimated combined predetermined proportional gain linear genomic selection index (CPPG-LGSI) heritability ( $\hat{h}_I^2$ ), estimated PPG-LGSI accuracy ( $\hat{\rho}_{HICP}$ ), values of  $W_{CP} = \frac{\hat{\rho}_{HICP}}{\hat{h}_I} L_I$  ( $L_I = 4$ ), and  $\hat{p} = 100(\hat{\lambda}_{CP} - 1)$ , where  $\hat{\lambda}_p = \hat{\rho}_{HICP}/\rho_{HI_p}$  and  $\hat{\rho}_{HI_p}$  is the estimated CPPG-LGSI accuracy, for one, two, and three null restrictions for five simulated selection cycles

Cycle	CPPG-LGSI heritability			PPG-LGSI accuracy			Values of $W_{CP}$			Values of $\hat{p}$
	1	2	3	1	2	3	1	2	3	
1	0.41	0.41	0.85	0.65	0.91	0.85	6.25	5.68	3.69	24.62
2	0.75	0.86	0.84	0.72	0.70	0.55	3.33	3.02	2.40	25.00
3	0.78	0.85	0.81	0.70	0.67	0.55	3.17	2.91	2.44	30.00
4	0.78	0.84	0.82	0.61	0.59	0.46	2.76	2.57	2.03	49.18
5	0.80	0.83	0.82	0.56	0.54	0.44	2.50	2.37	1.94	57.14
Average	0.70	0.76	0.83	0.65	0.68	0.57	3.60	3.31	2.50	37.19
										32.82
										37.08

## 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
- 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



## An Approach to Selection Index Method Based on Eigenanalysis



**Abstract** Based on the canonical correlation, on the singular value decomposition (SVD), and on the linear phenotypic selection indices theory, we describe the eigen selection index method (ESIM), the restricted ESIM (RESIM), and the predetermined proportional gain ESIM (PPG-ESIM), which use only phenotypic information to predict the net genetic merit. The ESIM is an unrestricted linear selection index, but the RESIM and PPG-ESIM are linear selection indices that allow null and predetermined restrictions respectively to be imposed on the expected genetic gains of some traits, whereas the rest remain without any restrictions. The aims of the three indices are to predict the unobservable net genetic merit values of the candidates for selection, maximize the selection response, and the accuracy, and provide the breeder with an objective rule for evaluating and selecting several traits simultaneously. Their main characteristics are: they do not require the economic weights to be known, the first multi-trait heritability eigenvector is used as its vector of coefficients; and because of the properties associated with eigen analysis, it is possible to use the theory of similar matrices to change the direction and proportion of the expected genetic gain values without affecting the accuracy. We describe the foregoing three indices and validate their theoretical results using real and simulated data.

### 2.1 The Linear Phenotypic Eigen Selection Index Method

The conditions described in Chapter. for the linear phenotypic selection index (LPSI) are necessary and sufficient for constructing the linear phenotypic eigen selection index method (ESIM). The ESIM index can be written as  $I = \mathbf{b}'\mathbf{y}$ , where  $\mathbf{b}' = [b_1 \ b_2 \ \dots \ b_t]$  is the unknown index vector of coefficients,  $t$  is the number of traits, and  $\mathbf{y}' = [y_1 \ y_2 \ \dots \ y_t]$  is a known vector of trait phenotypic values. The objectives of ESIM are:

1. To predict the net genetic merit  $H = \mathbf{w}'\mathbf{g}$ , where  $\mathbf{g}' = [g_1 \ g_2 \ \dots \ g_t]$  is the unknown vector of true breeding values for an individual and  $\mathbf{w}' = [w_1 \ w_2 \ \dots \ w_t]$  is a vector of unknown economic weights.



2. To maximize the ESIM selection response and the accuracy.
3. To select individuals with the highest  $H$  values in each selection cycle as parents of the next generation.
4. To provide the breeder with an objective rule for evaluating and selecting several traits simultaneously.

Although in the context of the LPSI  $\mathbf{w}$  is a known and fixed vector of economic weights, in the ESIM  $\mathbf{w}$  is fixed, but unknown and its values must be estimated in each selection cycle. This latter assumption is the fundamental difference between the ESIM and the LPSI and implies that the ESIM is more general than the LPSI. Thus, when  $\mathbf{w}$  is known, the LPSI and ESIM give the same results.

### 2.1.1 The ESIM Parameters

The theoretical ESIM selection response can be written as

$$R_I = k_I \sigma_H \rho_{HI}, \quad (2.1)$$

where  $k_I$  is the standardized selection differential (or selection intensity),  $\sigma_H = \sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}}$  is the standard deviation of  $H$ ,  $\rho_{HI} = \frac{\mathbf{w}' \mathbf{C} \mathbf{b}}{\sqrt{\mathbf{w}' \mathbf{C} \mathbf{w}} \sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}}$  is the correlation, and  $\mathbf{w}' \mathbf{C} \mathbf{b} = \sigma_{HI}$  the covariance between  $H$  and  $I$  respectively,  $\sigma_I = \sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}$  is the standard deviation of  $I$ ,  $\mathbf{C}$  is the covariance matrix of the true breeding values ( $\mathbf{g}$ ), and  $\mathbf{P}$  is the covariance matrix of the trait phenotypic values ( $\mathbf{y}$ ).

In the ESIM, it is assumed that  $k_I$  and  $\sigma_H$  are fixed, and that  $\mathbf{C}$  and  $\mathbf{P}$  are known; thus, to maximize Eq. (2.1), it is necessary to maximize  $\rho_{HI}^2 = \frac{(\mathbf{w}' \mathbf{C} \mathbf{b})^2}{(\mathbf{w}' \mathbf{C} \mathbf{w})(\mathbf{b}' \mathbf{P} \mathbf{b})}$  with respect to vectors  $\mathbf{b}$  and  $\mathbf{w}$  under the restrictions  $\sigma_H^2 = \mathbf{w}' \mathbf{C} \mathbf{w}$ ,  $\sigma_I^2 = \mathbf{b}' \mathbf{P} \mathbf{b}$ , and  $0 < \sigma_H^2, \sigma_I^2 < \infty$ , where  $\sigma_H^2 = \mathbf{w}' \mathbf{C} \mathbf{w}$  is the variance of  $H = \mathbf{w}' \mathbf{g}$  and  $\sigma_I^2 = \mathbf{b}' \mathbf{P} \mathbf{b}$  is the variance of  $I = \mathbf{b}' \mathbf{y}$ . That is, it is necessary to maximize the function

$$f(\mathbf{b}, \mathbf{w}, \mu, \phi) = (\mathbf{w}' \mathbf{C} \mathbf{b})^2 - \mu(\mathbf{b}' \mathbf{P} \mathbf{b} - \sigma_I^2) - \phi(\mathbf{w}' \mathbf{C} \mathbf{w} - \sigma_H^2) \quad (2.2)$$

with respect to  $\mathbf{b}$ ,  $\mathbf{w}$ ,  $\mu$ , and  $\phi$ , where  $\mu$  and  $\phi$  are Lagrange multipliers. The derivative results of Eq. (2.2) with respect to  $\mathbf{b}$ ,  $\mathbf{w}$ ,  $\mu$ , and  $\phi$  are:

$$(\mathbf{w}' \mathbf{C} \mathbf{b}) \mathbf{C} \mathbf{w} - \mu \mathbf{P} \mathbf{b} = \mathbf{0}, \quad (2.3)$$

$$(\mathbf{w}' \mathbf{C} \mathbf{b}) \mathbf{C} \mathbf{b} - \phi \mathbf{C} \mathbf{w} = \mathbf{0}, \quad (2.4)$$

$$\mathbf{b}' \mathbf{P} \mathbf{b} = \sigma_I^2 \text{ and } \mathbf{w}' \mathbf{C} \mathbf{w} = \sigma_H^2, \quad (2.5)$$

respectively, where Eq. (2.5) denotes the restrictions imposed for maximizing  $\rho_{HI}^2$ . It can be shown that  $\mathbf{w}' \mathbf{C} \mathbf{b} = \sqrt{\mu \sigma_I^2} = \sqrt{\phi \sigma_H^2} = \theta^{1/2}$ ; then, Eqs. (2.3) and (2.4) can be written as



$$\theta^{1/2} \mathbf{Cw} - \frac{\theta}{\sigma_I^2} \mathbf{Pb} = \mathbf{0} \quad (2.6)$$

and

$$\theta^{1/2} \mathbf{Cb} - \frac{\theta}{\sigma_H^2} \mathbf{Cw} = \mathbf{0}, \quad (2.7)$$

respectively. Equation (2.6) is equal to  $\mathbf{Cw} = \frac{\theta^{1/2}}{\sigma_I^2} \mathbf{Pb}$ ; then, vector  $\mathbf{w}$  can be written as

$$\mathbf{w}_E = \frac{\theta^{1/2}}{\sigma_I^2} \mathbf{C}^{-1} \mathbf{Pb}. \quad (2.8)$$

By the result of Eq. (2.8), the net genetic merit in the ESIM context is  $H_E = \mathbf{w}'_E \mathbf{g}$  and the correlation between  $H_E$  and  $I$  is  $\rho_{H_E I} = \frac{\mathbf{w}'_E \mathbf{C} \mathbf{b}}{\sqrt{\mathbf{w}'_E \mathbf{C} \mathbf{w}_E} \sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}} = \frac{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}}}$ . Now, it is necessary to find the vector  $\mathbf{b}$  that maximizes  $\rho_{H_E I}$ , which should be the ESIM index vector of coefficients. Substituting  $\mathbf{w}$  with  $\mathbf{w}_E$  in Eq. (2.7), we get

$$\mathbf{Cb} - \frac{(\mathbf{w}'_E \mathbf{C} \mathbf{b})^2}{\sigma_I^2 \sigma_{H_E}^2} \mathbf{Pb} = \mathbf{0}, \quad (2.9)$$

where  $\frac{(\mathbf{w}'_E \mathbf{C} \mathbf{b})^2}{\sigma_I^2 \sigma_{H_E}^2} = \rho_{H_E I}^2$  is the square of the correlation between ESIM and  $H_E = \mathbf{w}'_E \mathbf{g}$ . Let  $\rho_{H_E I}^2 = \lambda_E^2$ , then Eq. (2.9) can be written as

$$(\mathbf{P}^{-1} \mathbf{C} - \lambda_E^2 \mathbf{I}) \mathbf{b}_E = \mathbf{0}, \quad (2.10)$$

and the optimized ESIM index is  $I_E = \mathbf{b}'_E \mathbf{y}$ . Note that in Eq. (2.10)  $\mathbf{P}^{-1} \mathbf{C}$  is the multi-trait heritability. By Eqs. (2.8) and (2.10), the maximized correlation between  $H_E = \mathbf{w}'_E \mathbf{g}$  and  $I_E = \mathbf{b}'_E \mathbf{y}$  (or ESIM accuracy) can be written as

$$\rho_{H_E I_E} = \frac{\sigma_{I_E}}{\sigma_{H_E}}, \quad (2.11)$$

where  $\sigma_{I_E} = \sqrt{\mathbf{b}'_E \mathbf{P} \mathbf{b}_E}$  is the standard deviation of the variance of  $I_E = \mathbf{b}'_E \mathbf{y}$ , and  $\sigma_{H_E} = \sqrt{\mathbf{b}'_E \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_E}$  is the standard deviation of the variance of  $H_E = \mathbf{w}'_E \mathbf{g}$ . Hereafter, we write Eq. (2.11) as  $\rho_E = \rho_{H_E I_E}$  or  $\lambda_E = \rho_{H_E I_E}$  to simplify the notation.

An additional restriction on Eq. (2.10) is  $\mathbf{b}' \mathbf{b} = 1$ , because  $\rho_{H_E I_E}$  is invariant to the scale change and because if  $\mathbf{b}_E$  is an eigenvector of the multi-trait heritability matrix  $\mathbf{P}^{-1} \mathbf{C}$ , vector  $\alpha \mathbf{b}_E$  is also an eigenvector of  $\mathbf{P}^{-1} \mathbf{C}$  for all real values of  $\alpha$  (Mardia et al. 1982). This means that in the ESIM the magnitude of an eigenvector is unimportant;



only the direction matters (Watkins 2002). Equation (2.10) can also be written as  $\mathbf{C}\mathbf{b}_E = \lambda_E^2 \mathbf{P}\mathbf{b}_E$ , which is called the *generalized eigenvalue problem* (Watkins 2002). In the latter case,  $\mathbf{b}_E$  is called a *generalized eigenvector* and  $\lambda_E^2$  a *generalized eigenvalue*. The generalized eigenvalues may not exist; that is, they may be infinite. However, if  $\mathbf{P}$  is positive definite and has the same size as  $\mathbf{C}$ , all eigenvalues of  $\mathbf{P}^{-1}\mathbf{C}$  exist and are finite (Gentle 2007). Matrix  $\mathbf{P}$  is symmetric and positive definite and its eigenvalues are different with a probability of 1 if the number of genotypes is higher than the number of traits (Okamoto 1973).

If the heritability of the ESIM is  $h_I^2 = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\mathbf{b}'\mathbf{P}\mathbf{b}}$ , then another way of writing Eq. (2.1) is

$$R_I = k_I \sigma_I h_I^2 = k_I \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}}, \quad (2.12)$$

which is similar to the univariate breeder's equation (see Chapter.). All the parameters of Eq. (2.12) were defined earlier.

The derivative of the ratio  $\frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}}$  (Eq. 2.12) with respect to  $\mathbf{b}$  can be written as  $2(\mathbf{b}'\mathbf{P}\mathbf{b})^{1/2}\mathbf{C}\mathbf{b} - (\mathbf{b}'\mathbf{P}\mathbf{b})^{-1/2}(\mathbf{b}'\mathbf{C}\mathbf{b})\mathbf{P}\mathbf{b} = \mathbf{0}$ , and, except by a proportionality constant, the result is

$$\left( \mathbf{P}^{-1}\mathbf{C} - h_{I_E}^2 \mathbf{I} \right) \mathbf{b}_E = \mathbf{0}, \quad (2.13)$$

where  $h_{I_E}^2 = \frac{\mathbf{b}'_E \mathbf{C} \mathbf{b}_E}{\mathbf{b}'_E \mathbf{P} \mathbf{b}_E}$  is the maximized ESIM heritability. Let  $\lambda_E^2 = \rho_E^2 = h_{I_E}^2$ , then Eq. (2.13) is equal to Eq. (2.10) and can be written as  $\mathbf{b}'_E \mathbf{C} \mathbf{b}_E = \lambda_E^2 \mathbf{b}'_E \mathbf{P} \mathbf{b}_E$ , whence the maximized  $\rho_E^2$  in terms of  $h_{I_E}^2$  is

$$\rho_E^2 = \frac{\mathbf{b}'_E \mathbf{C} \mathbf{b}_E}{\mathbf{b}'_E \mathbf{P} \mathbf{b}_E}, \quad (2.14)$$

which should give a equivalent result to that of Eq. (2.11).

By Eq. (2.11) and  $\sigma_{H_E} = \sqrt{\mathbf{b}'_E \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_E}$ , the maximized ESIM selection response and expected genetic gain per trait can be written as

$$R_E = k_I \sqrt{\mathbf{b}'_E \mathbf{P} \mathbf{b}_E} \quad (2.15)$$

and

$$\mathbf{E}_E = k_I \frac{\mathbf{C} \mathbf{b}_E}{\sqrt{\mathbf{b}'_E \mathbf{P} \mathbf{b}_E}}, \quad (2.16)$$

respectively. Equations (2.15) and (2.16) do not require the economic weights to be known. In the original derivation of the ESIM, Cerón-Rojas et al. (2008) imposed the

restrictions  $\sigma_{H_E}^2 = 1$  and  $\sigma_{I_E}^2 = 1$ . Under these restrictions,  $\lambda_E = \mathbf{w}'_E \mathbf{C} \mathbf{b}_E$  and Eq. (2.15) can be written as  $R_E = k_I \lambda_E$ . When  $\sigma_{H_E}^2 \neq 1$  Eq. (2.15) is equal to  $R_E = k_I \sigma_{H_E} \lambda_E$ , where  $\sigma_{H_E} = \sqrt{\mathbf{b}'_E \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_E}$  and  $\lambda_E^2 = \rho_E^2 = h_{I_E}^2$ .

Let  $\mathbf{T} = \mathbf{P}^{-1} \mathbf{C}$  and  $\lambda_E^2 = h_{I_E}^2$ ; then, Eq. (2.13) can be written as  $\mathbf{T} \mathbf{b}_E = \lambda_E^2 \mathbf{I} \mathbf{b}_E$ , where  $\mathbf{I} = \mathbf{F}^{-1} \mathbf{F}$  is an identity matrix of size  $t \times t$  ( $t$  = number of traits), and  $\mathbf{F} = \text{diag}\{f_1 \ f_1 \ \cdots \ f_t\}$  is a diagonal matrix with values equal to any real number, except zero values. Thus, another way of writing Eqs. (2.10) and (2.13) is

$$(\mathbf{T}_2 - \lambda_E^2 \mathbf{I}) \boldsymbol{\beta} = \mathbf{0}, \quad (2.17)$$

where  $\mathbf{T}_2 = \mathbf{F} \mathbf{T} \mathbf{F}^{-1}$  and  $\boldsymbol{\beta} = \mathbf{F} \mathbf{b}_E$ ;  $\mathbf{T}$  and  $\mathbf{T}_2 = \mathbf{F} \mathbf{T} \mathbf{F}^{-1}$  are similar matrices and both have the same eigenvalues but different eigenvectors (Harville 1997). When the  $\mathbf{F}$  values are only 1s, vector  $\mathbf{b}_E$  is not affected; when the  $\mathbf{F}$  values are only  $-1$ s, vector  $\mathbf{b}_E$  changes its direction, and if the  $\mathbf{F}$  values are different from 1 and  $-1$ , matrix  $\mathbf{F}$  changes the proportional values of  $\mathbf{b}_E$ . In practice,  $\mathbf{b}_E$  is first obtained from Eq. (2.13) and then multiplied by matrix  $\mathbf{F}$  to obtain  $\boldsymbol{\beta} = \mathbf{F} \mathbf{b}_E$ ; that is,  $\boldsymbol{\beta}$  is a linear transformation of  $\mathbf{b}_E$ . Matrix  $\mathbf{T}_2 = \mathbf{F} \mathbf{T} \mathbf{F}^{-1}$  is called the *similarity transformation*, and matrix  $\mathbf{F}$  is called the *transforming matrix* (Watkins 2002). Cerón-Rojas et al. (2006) introduced an alternative procedure for modifying the  $\mathbf{b}_E$  signs that is a particular case of Eq. (2.17). Vector  $\boldsymbol{\beta} = \mathbf{F} \mathbf{b}_E$  can substitute  $\mathbf{b}_E$  in Eqs. (2.15) and (2.16); and in this case, the optimized ESIM index should be written as  $I_E = \boldsymbol{\beta}' \mathbf{y}$ .

## 2.1.2 Statistical ESIM Properties

The ratio of the index accuracies and the variance of the predicted error (VPE) are good criteria for comparing the index efficiencies for predicting the net genetic merit (see Chapter. for details). In Eq. (2.11), we obtained the accuracy of the ESIM; now, we derive the VPE of the ESIM.

The variance of  $I_E = \mathbf{b}'_E \mathbf{y}$  ( $\sigma_{I_E}^2$ ) and the covariance between  $H_E = \mathbf{w}'_E \mathbf{g}$  and  $I_E = \mathbf{b}'_E \mathbf{y}$  ( $\sigma_{H_E I_E}$ ) are the same, that is,

$$\sigma_{I_E}^2 = \mathbf{b}'_E \mathbf{P} \mathbf{b}_E \text{ and } \sigma_{H_E I_E} = \mathbf{w}'_E \mathbf{C} \mathbf{b}_E = \mathbf{b}'_E \mathbf{P} \mathbf{C}^{-1} \mathbf{C} \mathbf{b}_E = \mathbf{b}'_E \mathbf{P} \mathbf{b}_E, \quad (2.18)$$

respectively; that is,  $\sigma_{I_E}^2 = \sigma_{H_E I_E}$ . By Eq. (2.18), the VPE of the ESIM can be written as

$$E[(H_E - I_E)^2] = \sigma_{H_E}^2 + \sigma_{I_E}^2 - 2\sigma_{H_E I_E} = \sigma_{H_E}^2 - \sigma_{I_E}^2 = (1 - \rho_E^2)\sigma_{H_E}^2. \quad (2.19)$$



The relative effectiveness of  $I_E = \mathbf{b}'_E \mathbf{y}$  in predicting  $H_E = \mathbf{w}'_E \mathbf{g}$  is the ratio of  $(1 - \rho_E^2) \sigma_{H_E}^2$  over  $\sigma_{H_E}^2$ , i.e.,  $1 - \rho_E^2$ ; thus, the greater  $\rho_E^2$  is, the more effective  $I_E = \mathbf{b}'_E \mathbf{y}$  is at predicting  $H_E = \mathbf{w}'_E \mathbf{g}$ . The mean squared effect of  $I_E$  on  $H_E$ , or the total variance of  $H_E$  explained by  $I_E$  is

$$\sigma_{I_E}^2 = \rho_E^2 \sigma_{H_E}^2, \quad (2.20)$$

and the relative mean squared effect can be measured by  $\rho_E^2$  (Anderson 2003). If in Eq. (2.20)  $\rho_E^2 = 1$ ,  $\sigma_{I_E}^2 = \sigma_{H_E}^2$ , and if  $\rho_E^2 = 0$ ,  $\sigma_{I_E}^2 = 0$ . That is, the variance of  $H_E$  explained by  $I_E$  is proportional to  $\rho_E^2$ , and when  $\rho_E^2$  is close to 1,  $\sigma_{I_E}^2$  is close to  $\sigma_{H_E}^2$ , and if  $\rho_E^2$  is close to 0,  $\sigma_{I_E}^2$  is close to 0. All these results are valid for any index associated with the ESIM, such as the restricted ESIM (RESIM) and the predetermined proportional gains ESIM (PPG-ESIM), which are described in the following sections of this chapter.

### 2.1.3 The ESIM and the Canonical Correlation Theory

Canonical correlation theory describes the associations between two sets of variables (Hotelling 1935, 1936) and searches for linear combinations, called *canonical variables*, of each of two sets of variables having maximal correlation. The vector of coefficient of these linear combinations is called the *canonical vector* and the correlations between the canonical variables is called the *canonical correlation* (Wilms and Croux 2016).

To see how the ESIM and the canonical correlation theory are related, note that vectors  $\mathbf{y}$  and  $\mathbf{g}$  (Eq. 2.1) can be ordered in a new vector  $\mathbf{x}$  as  $\mathbf{x}' = [\mathbf{y}' \quad \mathbf{g}']$ , whence the covariance matrix of  $\mathbf{x}$  is  $\begin{bmatrix} \mathbf{P} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} \end{bmatrix}$ . One measure of the association between the  $j$ th linear combination of  $\mathbf{y}$  ( $I_E = \mathbf{b}'_E \mathbf{y}$ ) and the  $j$ th linear combination of  $\mathbf{g}$  ( $H_E = \mathbf{w}'_E \mathbf{g}$ ) is the  $j$ th canonical correlation ( $\lambda_j$ ) value obtained from equation  $(\mathbf{P}^{-1}\mathbf{C} - \lambda_j^2 \mathbf{I})\mathbf{b}_{Ej} = \mathbf{0}$ , where  $\mathbf{b}_{Ej}$  is the  $j$ th *canonical vector* ( $j = 1, 2 \dots, t$ ) of matrix  $\mathbf{P}^{-1}\mathbf{C}$ , and  $\mathbf{w}_{Ej} = \mathbf{C}^{-1}\mathbf{P}\mathbf{b}_{Ej}$ . Thus, in the canonical correlation context,  $I_E = \mathbf{b}'_E \mathbf{y}$  and  $H_E = \mathbf{w}'_E \mathbf{g}$  are *canonical variables*.

In the ESIM, the first eigenvector ( $\mathbf{b}_{E1}$ ) of matrix  $\mathbf{P}^{-1}\mathbf{C}$  should be used on  $I_E = \mathbf{b}'_{E1} \mathbf{y}$ ; the first eigenvalue ( $\lambda_1^2$ ) and  $\mathbf{b}_{E1}$  of  $\mathbf{P}^{-1}\mathbf{C}$  should be used on the ESIM selection response and on the ESIM expected genetic gain per trait, because, in this case, the ESIM has maximum accuracy compared with other indices, such as the LPSI. The latter results in this subsection imply that the sampling statistical properties associated with the canonical correlation theory are also valid for the ESIM.



## 2.1.4 Estimated ESIM Parameters and Their Sampling Properties

The estimated covariance matrix of the true breeding values ( $\mathbf{C}$ ) and that of the trait phenotypic values ( $\mathbf{P}$ ) are denoted as  $\widehat{\mathbf{C}}$  and  $\widehat{\mathbf{P}}$  respectively; they can be obtained by restricted maximum likelihood using Eqs. (2.22) to (2.24) described in Chapter. With matrices  $\widehat{\mathbf{C}}$  and  $\widehat{\mathbf{P}}$ , we constructed matrix  $\widehat{\mathbf{T}} = \widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}}$  and equation

$$(\widehat{\mathbf{T}} - \widehat{\lambda}_{Ej}^2 \mathbf{I})\widehat{\mathbf{b}}_{Ej} = \mathbf{0}, \quad (2.21)$$

$j = 1, 2, \dots, t$ , where  $t$  is the number of traits in the ESIM index. Note that  $\widehat{\lambda}_{Ej}^2$  is positive only if  $\widehat{\mathbf{P}}$  is positive definite (all eigenvalues positive) and  $\widehat{\mathbf{C}}$  is positive semidefinite (no negative eigenvalues); in addition, as  $\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}}$  is an asymmetric matrix, the values of  $\widehat{\mathbf{b}}_{Ej}$  and  $\widehat{\lambda}_{Ej}^2$  should be obtained using the singular value decomposition (SVD) theory (Anderson 2003).

Matrix  $\widehat{\mathbf{T}}$  is square and asymmetric of order  $t \times t$  and rank  $q \leq \min(p, c)$ , where  $p$  and  $c$  denote the rank of  $\widehat{\mathbf{P}}^{-1}$  and  $\widehat{\mathbf{C}}$  respectively; the rank of  $\widehat{\mathbf{T}}$  is equal to  $c$  only if  $\widehat{\mathbf{C}}$  is square and nonsingular. Thus, matrix  $\widehat{\mathbf{T}}$  has a maximum of  $q$  eigenvalues different from zero (Rao 2002). In addition,  $\widehat{\mathbf{T}}\widehat{\mathbf{T}}'$  and  $\widehat{\mathbf{T}}'\widehat{\mathbf{T}}$  are symmetric matrices, but  $\widehat{\mathbf{T}}\widehat{\mathbf{T}}' \neq \widehat{\mathbf{T}}'\widehat{\mathbf{T}}$ . Using the SVD theory, matrix  $\widehat{\mathbf{T}}$  can be written as

$$\widehat{\mathbf{T}} = \mathbf{V}_1 \mathbf{L}^{1/2} \mathbf{V}_2', \quad (2.22)$$

where  $\mathbf{V}_1$  ( $\mathbf{V}_1' \mathbf{V}_1 = \mathbf{V}_1 \mathbf{V}_1' = \mathbf{I}_q$ ) and  $\mathbf{V}_2$  ( $\mathbf{V}_2' \mathbf{V}_2 = \mathbf{V}_2 \mathbf{V}_2' = \mathbf{I}_q$ ) are matrices with the eigenvectors of matrices  $\widehat{\mathbf{T}}\widehat{\mathbf{T}}'$  and  $\widehat{\mathbf{T}}'\widehat{\mathbf{T}}$  respectively;  $\mathbf{L}^{1/2}$  is a diagonal matrix with the square root of the eigenvalues ( $\widehat{\lambda}_{E_1}^2 \geq \widehat{\lambda}_{E_2}^2 \geq \dots \geq \widehat{\lambda}_{E_q}^2 > 0$ ) of either  $\widehat{\mathbf{T}}\widehat{\mathbf{T}}'$  or  $\widehat{\mathbf{T}}'\widehat{\mathbf{T}}$  (the eigenvalues of  $\widehat{\mathbf{T}}\widehat{\mathbf{T}}'$  and  $\widehat{\mathbf{T}}'\widehat{\mathbf{T}}$  are the same). The entries  $\widehat{\lambda}_{E_1}^2 \geq \widehat{\lambda}_{E_2}^2 \geq \dots \geq \widehat{\lambda}_{E_q}^2 > 0$  of  $\mathbf{L}^{1/2}$  are uniquely determined, and they are called the *singular values* of  $\widehat{\mathbf{T}}$ . The columns of  $\mathbf{V}_1$  are orthonormal vectors called *left singular vectors* of  $\widehat{\mathbf{T}}$ , and the columns of  $\mathbf{V}_2$  are called *right singular vectors* (Watkins 2002).

Estimators  $\widehat{\mathbf{b}}_{E_1}$  and  $\widehat{\lambda}_{E_1}^2$  of the first eigenvector  $\mathbf{b}_{E_1}$  and the first eigenvalue  $\lambda_{E_1}^2$  respectively are the first column of matrix  $\mathbf{V}_1$  and the first diagonal element of matrix  $\mathbf{L}^{1/2}$ . Thus, because  $\widehat{\mathbf{T}}\widehat{\mathbf{T}}'$  is a symmetric matrix, the maximum likelihood estimators  $\widehat{\lambda}_{E_1}^2$  and  $\widehat{\mathbf{b}}_{E_1}$  in the ESIM context can be obtained from

$$(\widehat{\mathbf{T}}\widehat{\mathbf{T}}' - \widehat{\mu}_j \mathbf{I})\widehat{\mathbf{b}}_{E_j} = \mathbf{0}, \quad (2.23)$$

where  $\widehat{\mu}_j = \widehat{\lambda}_{E_j}^4, j = 1, 2, \dots, t$ . In the asymptotic context,  $\widehat{\lambda}_{E_1}^2$  and  $\widehat{\mathbf{b}}_{E_1}$  are consistent and unbiased estimators (Anderson 2003).

The latter results allow the ESIM index ( $I_E = \mathbf{b}'_E \mathbf{y}$ ) as  $\widehat{I}_E = \widehat{\mathbf{b}}'_E \mathbf{y}$  to be estimated. The estimator of the maximized ESIM selection response and expected genetic gain



per trait are  $\widehat{R}_E = k_I \sqrt{\widehat{\mathbf{b}}'_{E_1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{E_1}}$  and  $\widehat{\mathbf{E}}_E = k_I \frac{\widehat{\mathbf{C}} \widehat{\mathbf{b}}_{E_1}}{\sqrt{\widehat{\mathbf{b}}'_{E_1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{E_1}}}$  respectively, whereas the estimator of the maximized ESIM accuracy is  $\widehat{\lambda}_{E_1}$ , which should be similar to the estimator of the square root of the maximized ESIM heritability.

In the asymptotic context, the estimator of  $\mathbf{b}_{Ej}$  ( $\widehat{\mathbf{b}}_{Ej}$ ) has multivariate normal distribution with expectation  $E(\widehat{\mathbf{b}}_{Ej}) = \mathbf{b}_{Ej}$  and variance

$$Var(\widehat{\mathbf{b}}_{Ej}) = \frac{1}{2n} \mathbf{b}_{Ej} \mathbf{b}'_{Ej} + \frac{1}{n} (1 - \lambda_{Ej}^2) \sum_{i \neq j}^t \frac{\lambda_{Ej}^2 + \lambda_{Ei}^2 - 2\lambda_{Ei}^2 \lambda_{Ej}^2}{(\lambda_{Ei}^2 - \lambda_{Ej}^2)^2} \mathbf{b}_{Ei} \mathbf{b}'_{Ei}, \quad (2.24)$$

and, for  $i \neq j$ , the covariance between  $\widehat{\mathbf{b}}_{Ei}$  and  $\widehat{\mathbf{b}}_{Ej}$  can be written as

$$Cov(\widehat{\mathbf{b}}_{Ei}, \widehat{\mathbf{b}}_{Ej}) = \frac{(1 - \lambda_{Ej}^2)(1 - \lambda_{Ei}^2)(\lambda_{Ei}^2 + \lambda_{Ej}^2)}{n(\lambda_{Ei}^2 - \lambda_{Ej}^2)^2} \mathbf{b}_{Ej} \mathbf{b}'_{Ei}, \quad (2.25)$$

where  $n$  is the number of individuals or genotypes (Anderson 1999). The variance of  $\widehat{\mathbf{b}}_{Ej}$  and the covariance between  $\widehat{\mathbf{b}}_{Ei}$  and  $\widehat{\mathbf{b}}_{Ej}$  depend not only on  $n$ , but also on eigenvalues  $\lambda_{Ei}^2$  and  $\lambda_{Ej}^2$ . Suppose that  $\lambda_{Ej}^2 > \lambda_{Ei}^2$ ; then, when  $\lambda_{Ej}^2$  is very close to 1,  $Var(\widehat{\mathbf{b}}_{Ej}) \approx \frac{1}{2n} \mathbf{b}_{Ej} \mathbf{b}'_{Ej}$  (" $\approx$ " denotes an approximation) and  $Cov(\widehat{\mathbf{b}}_{Ei}, \widehat{\mathbf{b}}_{Ej})$  is very close to 0. By the result of Eq. (2.24), the variance of the first eigenvector ( $\widehat{\mathbf{b}}_{E1}$ ) of  $\widehat{\mathbf{P}}^{-1} \widehat{\mathbf{C}}$  can be written as  $Var(\widehat{\mathbf{b}}_{E1}) = \frac{1}{2n} \mathbf{b}_{E1} \mathbf{b}'_{E1} + \frac{1}{n} (1 - \lambda_{E1}^2) \sum_{j=2}^t \frac{\lambda_{E1}^2 + \lambda_{Ej}^2 - 2\lambda_{E1}^2 \lambda_{Ej}^2}{(\lambda_{E1}^2 - \lambda_{Ej}^2)^2} \mathbf{b}_{Ej} \mathbf{b}'_{Ej}$ . If the first eigenvalue  $\lambda_{E1}^2$  of  $\mathbf{P}^{-1} \mathbf{C}$  is very close to 1 ( $\lambda_{E1}^2 \approx 1$ ),  $Var(\widehat{\mathbf{b}}_{E1}) = \frac{1}{2n} \mathbf{b}_{E1} \mathbf{b}'_{E1}$  and  $Cov(\widehat{\mathbf{b}}_{E1}, \widehat{\mathbf{b}}_{Ej}) \approx 0$ .

In the asymptotic context, the  $j$ th estimator ( $\widehat{\lambda}_{Ej}$ ) of the canonical correlations has normal distribution with expectation  $E(\widehat{\lambda}_{Ej}) \approx \lambda_{Ej}$  and variance

$$Var(\widehat{\lambda}_{Ej}) \approx \frac{(1 - \lambda_{Ej}^2)^2}{n}, \quad (2.26)$$

whereas the  $j$ th estimator of the square of the canonical correlations  $\widehat{\lambda}_{Ej}^2$  has normal distribution with expectation  $E(\widehat{\lambda}_{Ej}^2) \approx \lambda_{Ej}^2$  and variance

$$Var(\widehat{\lambda}_j^2) \approx \frac{4\lambda_{Ej}^2(1 - \lambda_{Ej}^2)^2}{n}. \quad (2.27)$$



In addition, for  $i \neq j$ , the correlation between  $\widehat{\lambda}_{Ej}^2$  and  $\widehat{\lambda}_{Ei}^2$  is zero, i.e.,  $\text{Corr}(\widehat{\lambda}_{Ei}^2, \widehat{\lambda}_{Ej}^2) = 0$  (Bilodeau and Brenner 1999; Muirhead 2005).

Equation (2.26) implies that under the restrictions  $\sigma_H^2 = 1$  and  $\sigma_I^2 = 1$ , the expectation and variance of  $\widehat{R}_E = k_I \widehat{\lambda}_{E1}$  are  $E(\widehat{R}_E) \approx k_I \lambda_{E1}$  and  $\text{Var}(\widehat{R}_E) \approx \frac{k_I^2(1-\lambda_{E1}^2)}{n}^2$  respectively. However, obtaining the expectation and variance of  $\widehat{R}_E = k_I \widehat{\sigma}_H \widehat{\lambda}_{E1}$  or  $\widehat{R}_E = k_I \sqrt{\widehat{\mathbf{b}}'_{E1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{E1}}$  is more difficult, because in both equations there are two estimators:  $\widehat{\sigma}_H$  and  $\widehat{\lambda}_1$  in the first one, and  $\widehat{\mathbf{P}}$  and  $\widehat{\mathbf{b}}_{E1}$  in the second one.

### 2.1.5 Numerical Examples

We compare ESIM efficiency versus LPSI efficiency using a real data set from commercial egg poultry lines obtained from Akbar et al. (1984). The estimated phenotypic ( $\widehat{\mathbf{P}}$ ) and genetic ( $\widehat{\mathbf{C}}$ ) covariance matrices among the rate of lay (RL, number of eggs), age at sexual maturity (SM, days) and egg weight (EW, kg), were

$$\widehat{\mathbf{P}} = \begin{bmatrix} 240.57 & -95.62 & 2.07 \\ -95.62 & 167.20 & 4.58 \\ 2.07 & 4.58 & 22.80 \end{bmatrix} \text{ and } \widehat{\mathbf{C}} = \begin{bmatrix} 29.86 & -17.90 & -4.13 \\ -17.90 & 18.56 & 1.49 \\ -4.13 & 1.49 & 9.24 \end{bmatrix} \text{ respectively.}$$

The number of genotypes and the vector of economic weights were  $n = 3330$  and  $\mathbf{w}' = [19.54 \quad -3.56 \quad 17.01]$  respectively, whereas the selection intensity was 10% ( $k_I = 1.755$ ) for both indices.

The estimated LPSI vector of coefficients was  $\widehat{\mathbf{b}}'_S = \mathbf{w}' \widehat{\mathbf{P}}^{-1} \widehat{\mathbf{C}} = [1.82 \quad -1.38 \quad 3.25]$ , whereas the estimated selection response, expected genetic gain per trait, accuracy, and heritability of the LPSI were  $\widehat{R}_S = 1.755 \sqrt{\widehat{\mathbf{b}}'_S \widehat{\mathbf{P}} \widehat{\mathbf{b}}_S} = 74.91$ ,  $\widehat{\mathbf{E}}'_S = 1.755 \frac{\widehat{\mathbf{b}}'_S \widehat{\mathbf{C}}}{\sqrt{\widehat{\mathbf{b}}'_S \widehat{\mathbf{P}} \widehat{\mathbf{b}}_S}} = [2.70 \quad -2.20 \quad 0.84]$ ,

$$\widehat{\rho}_S = \frac{\sqrt{\widehat{\mathbf{b}}'_S \widehat{\mathbf{P}} \widehat{\mathbf{b}}_S}}{\sqrt{\mathbf{w}' \widehat{\mathbf{C}} \mathbf{w}}} = 0.362, \text{ and } \widehat{h}_S^2 = \frac{\widehat{\mathbf{b}}'_S \widehat{\mathbf{C}} \widehat{\mathbf{b}}_S}{\widehat{\mathbf{b}}'_S \widehat{\mathbf{P}} \widehat{\mathbf{b}}_S} = 0.143 \text{ respectively.}$$

Note that because in the ESIM context  $\widehat{\mathbf{b}}'_E \widehat{\mathbf{b}}_E = 1$ , the best way of comparing ESIM results versus LPSI results is when the LPSI coefficient vector is normalized, i.e., when the LPSI coefficient vector is equal to  $\widehat{\mathbf{b}}_S^* = \widehat{\mathbf{b}}_S / \widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S$  and then  $\widehat{\mathbf{b}}'^*_S \widehat{\mathbf{b}}_S^* = 1$ ; however, it can be shown that the normalization process only affects the estimated LPSI selection response because in that case,  $\widehat{R}_S = 74.91$  is divided by  $\widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S$ . For example, for this data set result,  $\widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S = 15.76$ ; then, the estimated LPSI selection response using  $\widehat{\mathbf{b}}_S^* = \widehat{\mathbf{b}}_S / \widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S$  is  $\widehat{R}_S = \frac{74.91}{15.74} = 4.75$ , whereas the rest of the estimated LPSI parameters are the same. When  $0 < \widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S < 1$  and  $1 < \widehat{R}_S$ , the



values of  $\widehat{R}_S$  increase, but when  $1 < \widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S$ , the values of  $\widehat{R}_S$  decrease, as in the example.

The product  $\widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S$  does not affect  $\widehat{\rho}_S$  because it is invariant to scale change. Also,  $\widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S$  does not affect  $\widehat{h}_S^2$  and  $\widehat{\mathbf{E}}_S$  because  $\widehat{\mathbf{b}}'_S \widehat{\mathbf{b}}_S$  appears in the numerator and denominator of both estimated parameters.

In the ESIM, the sign and proportion of the expected genetic gain values for traits RL, SM, and EW should be in accordance with the breeder's interest. For example, if the breeder's interest is that the expected genetic gain per trait for RL should be positive and negative for SM, the sign and proportion of the values of the first eigenvector should be modified using a linear combination of the estimated first eigenvector  $\widehat{\mathbf{b}}_{E_1}$ , i.e.,  $\widehat{\beta} = \mathbf{F}\widehat{\mathbf{b}}_{E_1}$ , to achieve expected genetic gain per trait values in RL and SM according to the breeder's interest.

The information needed to obtain the estimated ESIM parameters are matrices  $\widehat{\mathbf{T}} = \widehat{\mathbf{P}}^{-1} \widehat{\mathbf{C}}$  and  $\widehat{\mathbf{T}}\widehat{\mathbf{T}}' = \begin{bmatrix} 0.0146 & -0.0073 & -0.0338 \\ -0.0073 & 0.0093 & 0.0041 \\ -0.0338 & 0.0041 & 0.2056 \end{bmatrix}$ . We need to find the eigenvalues and eigenvectors of equation  $(\widehat{\mathbf{T}}\widehat{\mathbf{T}}' - \widehat{\mu}_j \mathbf{I})\widehat{\mathbf{b}}_{E_j} = \mathbf{0}$ ,

where  $\widehat{\mu}_j = \widehat{\lambda}_{E_j}^4$ , to obtain matrices  $\mathbf{V}_1$  and  $\mathbf{L}^{1/2}$ , which form matrix  $\widehat{\mathbf{T}} = \mathbf{V}_1 \mathbf{L}^{1/2} \mathbf{V}_2'$ .

Matrix  $\mathbf{V}_1$  is equal to  $\mathbf{V}_1 = \begin{bmatrix} -0.1701 & 0.6818 & 0.7115 \\ 0.0259 & -0.7187 & 0.6948 \\ 0.9851 & 0.1366 & 0.1046 \end{bmatrix}$ , whereas the diagonal elements of matrix  $\mathbf{L}$  are 0.2115, 0.0155, and 0.0025, that is, matrix  $\mathbf{L}^{1/2} = \begin{bmatrix} 0.4599 & 0 & 0 \\ 0 & 0.1244 & 0 \\ 0 & 0 & 0.0498 \end{bmatrix}$ . Thus,  $\widehat{\mu}_1 = \widehat{\lambda}_{E_1}^4 = 0.2115$ ,  $\widehat{\lambda}_{E_1}^2 = 0.4599$ ,

and the estimated ESIM accuracy was  $\widehat{\lambda}_{E_1} = 0.6782$ . The estimated ESIM eigenvector of coefficients is the first column of matrix  $\mathbf{V}_1$ , i.e.,  $\widehat{\mathbf{b}}'_{E_1} = [-0.1701 \ 0.0259 \ 0.9851]$ , and the estimated ESIM index can be constructed as  $\widehat{I}_E = -0.1701\text{RL} + 0.0259\text{SM} + 0.9851\text{EW}$ .

The estimated ESIM selection response and expected genetic gain per trait were

$$\widehat{R}_E = 1.755 \sqrt{\widehat{\mathbf{b}}'_{E_1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{E_1}} = 9.54 \text{ and } \widehat{\mathbf{E}}'_E = 1.755 \frac{\widehat{\mathbf{b}}'_{E_1} \widehat{\mathbf{C}}}{\sqrt{\widehat{\mathbf{b}}'_{E_1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{E_1}}} = [-3.10 \ 1.61 \ 3.18]$$

respectively. Because the estimated LPSI selection response was  $\widehat{R}_S = \frac{74.91}{15.74} = 4.75$ , the estimated ESIM selection response was higher than the estimated LPSI response. In addition, the estimated LPSI expected genetic gain per trait was  $\widehat{\mathbf{E}}'_S = [2.70 \ -2.20 \ 0.84]$ . Now, suppose that the breeder's interest is to increase RL and decrease SM; then,  $\widehat{\mathbf{E}}'_S$  is a good result but  $\widehat{\mathbf{E}}'_E$  is wrong.

We can change the sign and proportion of  $\widehat{\mathbf{E}}'_E$  by transforming  $\widehat{\mathbf{b}}_{E_1}$  into  $\widehat{\beta} = \mathbf{F}\widehat{\mathbf{b}}_{E_1}$  using a convenient matrix  $\mathbf{F}$  such as  $\mathbf{F} = \begin{bmatrix} -9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ . In such a case



$\widehat{\beta}' = \widehat{\mathbf{b}}'_{E_1} \mathbf{F} = [1.531 \quad 0.026 \quad 0.981]$ ,  $\widehat{R}_E = 1.755\sqrt{\widehat{\beta}' \widehat{\mathbf{P}} \widehat{\beta}} = 42.44$ , and  $\widehat{\mathbf{E}}'_E = 1.755 \frac{\widehat{\beta}' \widehat{\mathbf{C}}}{\sqrt{\widehat{\beta}' \widehat{\mathbf{P}} \widehat{\beta}}} = [2.990 \quad -1.85 \quad 0.205]$ . However, vector  $\widehat{\beta}'$  was not normalized.

To normalize  $\widehat{\beta}'$  we need to divide it by  $\widehat{\beta}' \widehat{\beta} = 3.314$ , but  $\widehat{\beta}' \widehat{\beta}$  should only affect  $\widehat{R}_E = 42.44$ , which should be divided by 3.314, that is,  $\widehat{R}_E = \frac{42.44}{3.314} = 12.806$ .

According to the theory of similar matrices (Harville 1997), the estimated maximized ESIM accuracy,  $\widehat{\lambda}_{E_1} = 0.6782$ , should not be affected by matrix  $\mathbf{F}$ .

We can compare ESIM efficiency versus LPSI efficiency to predict the net genetic merit using the ratio of the estimated ESIM accuracy  $\widehat{\lambda}_{E_1} = 0.6782$  to LPSI accuracy  $\widehat{\rho}_S = 0.362$ , i.e.,  $\frac{\widehat{\lambda}_{E_1}}{\widehat{\rho}_S} = \frac{0.6782}{0.362} = 1.873$ , or in percentage terms,

$$\widehat{p}_E = 100(1.873 - 1) = 87.3$$

According to the latter result, the ESIM is a better predictor of the net genetic merit and its efficiency is 87.3% higher than that of the LPSI for this data set.

Now, we compare ESIM efficiency versus LPSI efficiency using the data set described in Sect. 2.8.1 of Chapter. From this data set, we ran five phenotypic selection cycles, each with four traits ( $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ ), 500 genotypes, and four replicates for each genotype. The economic weights for  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$  were 1, -1, 1, and 1 respectively. In this case, matrix  $\mathbf{F}$  is an identity matrix of size  $4 \times 4$  for all five selection cycles.

Table 2.1 presents the estimated LPSI, the restricted LPSI (RLPSI), and the predetermined proportional gain LPSI (PPG-LPSI) selection response (the latter two for one, two, and three restrictions) for five simulated selection cycles when their vectors of coefficients are normalized. Table 2.1 also presents the estimated ESIM, the RESIM and the PPG-ESIM selection response for one, two, and three restrictions for five simulated selection cycles. The selection intensity was 10% ( $k_I = 1.755$ ) for all five selection cycles. In this subsection, we compare only LPSI results versus ESIM results. The estimated LPSI selection response when the vector of coefficients was not normalized was described in Chapter. (Table 2.4). The averages of the estimated LPSI and ESIM selection responses were 4.70 and 6.31 respectively.

Table 2.2 presents the estimated ESIM expected genetic gain per trait, accuracy ( $\widehat{\rho}_E$ ), and the values  $\widehat{p}_E = 100(\widehat{\lambda}_E - 1)$ , where  $\widehat{\lambda}_E = \widehat{p}_E / \widehat{\rho}_S$  is the ratio of  $\widehat{p}_E$  to the estimated LPSI accuracy ( $\widehat{\rho}_S$ ), expressed as percentages. Table 2.2 also presents the accuracy of the PPG-ESIM and the estimated ratio ( $\widehat{p}_{PE}$ ) of the estimated PPG-ESIM accuracy to the estimated PPG-LPSI accuracy, expressed as percentages, for one, two, and three predetermined restrictions for five simulated selection cycles. In this subsection, we use only the estimated ESIM expected genetic gain per trait and  $\widehat{p}_E = 100(\widehat{\lambda}_E - 1)$  to compare ESIM efficiency versus LPSI efficiency.

The estimated LPSI expected genetic gains per trait were presented in Chapter. Table 2.4. According to the results shown in Table 2.4, the averages of the estimated



**Table 2.1** Estimated linear phenotypic selection index (LPSI), restricted null LPSI (RLPSI), and predetermined proportional gains LPSI (PPG-LPSI) selection responses when their vectors of coefficients are normalized; estimated eigen selection index method (ESIM), restricted null ESIM (RESIM), and predetermined proportional gain ESIM (PPG-ESIM) selection responses for one, two, and three restrictions for five simulated selection cycles

Cycle	LPSI response	RLPSI response for one, two, and three null restrictions			PPG-LPSI response for one, two, and three predetermined restrictions		
		1	2	3	1	2	3
1	4.78	4.79	4.44	5.06	4.78	5.41	3.18
2	4.84	4.51	4.39	5.15	4.84	5.19	3.35
3	4.59	4.51	4.39	5.26	4.59	4.83	3.53
4	4.80	4.15	4.06	4.71	4.80	4.96	2.64
5	4.48	4.19	4.22	4.41	4.48	4.14	2.99
Average	4.70	4.43	4.30	4.92	4.70	4.91	3.14
Cycle	ESIM response	RESIM response for one, two, and three null restrictions			PPG-ESIM response for one, two, and three predetermined restrictions		
		1	2	3	1	2	3
1	8.88	4.78	4.64	4.57	8.88	7.1	7.4
2	6.13	4.86	4.69	4.69	6.13	6.04	7.3
3	5.44	4.96	4.79	4.68	5.44	5.87	6.91
4	4.84	4.30	4.19	4.19	4.84	4.91	5.77
5	6.24	3.79	3.78	3.78	6.24	7.49	6.39
Average	6.31	4.54	4.42	4.38	6.31	6.28	6.75

LPSI expected genetic gain per trait T1, T2, T3, and T4 for five simulated selection cycles were 7.26, -3.52, 2.78, and -1.58, whereas according to the results of Table 2.2, the averages of the estimated ESIM expected genetic gains per trait were 5.67, -2.67, 1.81, and 2.9 respectively. This means that the estimated LPSI expected genetic gain for traits T1, T2, and T3 was higher than the estimated ESIM expected genetic gain for those traits.

The average of the  $\hat{p}_E = 100(\hat{\lambda}_E - 1)$  values was 9.76 for all five selection cycles (Table 2.2). The latter result is not in accordance with the LPSI and ESIM expected genetic gain per trait; however, note that the  $\hat{p}_E$  values are associated with the estimated LPSI and ESIM selection responses (Table 2.1), not with the expected genetic gain per trait, because  $\hat{\lambda}_E = \frac{\hat{p}_E}{\hat{p}_S} \approx \frac{\hat{R}_E}{\hat{R}_S}$ , where  $\hat{R}_E$  and  $\hat{R}_S$  are the estimated ESIM and LPSI selection responses respectively. Thus, the  $\hat{p}_E$  values indicate that the efficiency of the ESIM and that of the LPSI were very similar because the former was only 9.76% higher than the latter for this data set.

The equality  $\frac{\hat{p}_E}{\hat{p}_S} = \frac{\hat{R}_E}{\hat{R}_S}$  is true only when the denominators of both estimated correlations are the same, as in the linear selection indices described in Chapter.

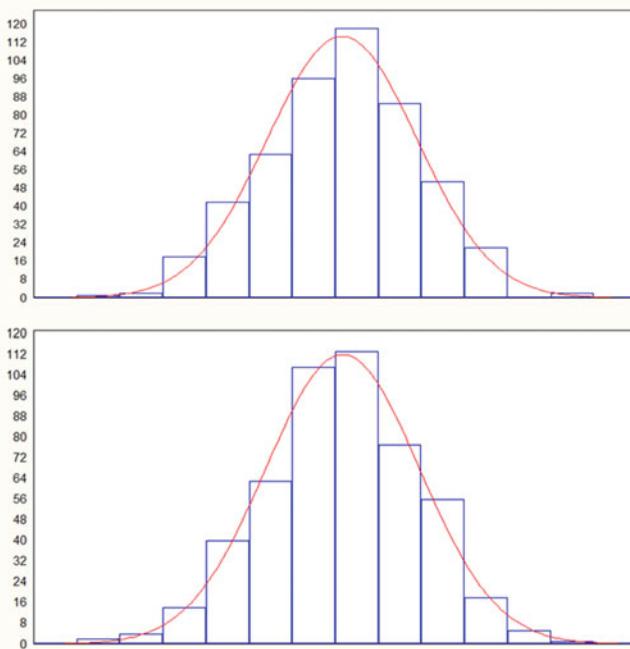
**Table 2.2** Estimated eigen selection index method (ESIM) expected genetic gain per trait, accuracy ( $\hat{\rho}_E$ ), and ratio of  $\hat{\rho}_E$  to the estimated LPSI (data not presented) accuracy ( $\hat{\rho}_S$ ), expressed in percentage terms,  $\hat{\rho}_E = 100(\hat{\lambda}_E - 1)$  (where  $\hat{\lambda}_E = \hat{\rho}_E/\hat{\rho}_S$ )

Cycle	ESIM expected genetic gain per trait				ESIM accuracy	$\hat{\rho}_E$ values (in %)
	T1	T2	T3	T4		
1	7.81	-4.62	3.11	2.21	0.98	8.11
2	5.15	-2.98	2.31	3.48	0.96	9.34
3	4.74	-1.15	0.66	3.79	0.97	10.94
4	3.94	-2.44	0.74	3.34	0.95	10.04
5	6.68	-2.15	2.24	2.05	0.95	10.35
Average	5.67	-2.67	1.81	2.97	0.96	9.76
PPG-ESIM accuracies for one, two, and three predetermined restrictions				$\hat{\rho}_P$ values (in %) for one, two, and three predetermined restrictions		
Cycle	1	2	3	1	2	3
1	0.98	0.96	0.99	9.34	8.90	20.99
2	0.96	0.96	0.98	10.94	12.46	25.20
3	0.97	0.97	1.00	10.04	9.71	41.43
4	0.95	0.94	0.99	10.35	13.98	28.95
5	0.98	0.96	0.99	9.34	8.90	20.99
Average	0.96	0.96	0.99	9.76	11.71	29.03

Estimated PPG-ESIM accuracy ( $\hat{\rho}_P$ ) and estimated ratio ( $\hat{\rho}_P$ ) of the  $\hat{\rho}_P$  to the estimated accuracy of the PPG-LPSI (data not presented), expressed in percentages (%), for one, two, and three predetermined restrictions for five simulated selection cycles

Note that  $\hat{\rho}_S = \frac{\sqrt{\mathbf{b}'_S \widehat{\mathbf{P}} \mathbf{b}_S}}{\sqrt{\mathbf{w}' \widehat{\mathbf{C}} \mathbf{w}}}$  and  $\hat{\rho}_E = \frac{\sqrt{\mathbf{b}'_E \widehat{\mathbf{P}} \mathbf{b}_E}}{\sqrt{\mathbf{w}'_E \widehat{\mathbf{C}} \mathbf{w}_E}}$ , whereas  $\widehat{R}_S = \sqrt{\mathbf{b}'_S \widehat{\mathbf{P}} \mathbf{b}_S}$  and  $\widehat{R}_E = \sqrt{\mathbf{b}'_E \widehat{\mathbf{P}} \mathbf{b}_E}$ ; this means that if  $\sqrt{\mathbf{w}'_E \widehat{\mathbf{C}} \mathbf{w}_E} \neq \sqrt{\mathbf{w}' \widehat{\mathbf{C}} \mathbf{w}}$ ,  $\frac{\hat{\rho}_E}{\hat{\rho}_S} \neq \frac{\widehat{R}_E}{\widehat{R}_S}$ . For the Akbar et al. (1984) data,  $\widehat{R}_E = 9.54$  and  $\widehat{R}_S = 4.75$ , then  $\frac{\widehat{R}_E}{\widehat{R}_S} = 2.0$  but  $\frac{\hat{\lambda}_{E_1}}{\hat{\rho}_S} = 1.873$ ; that is,  $\frac{\hat{\rho}_E}{\hat{\rho}_S} \approx \frac{\widehat{R}_E}{\widehat{R}_S}$ , where “ $\approx$ ” indicates an approximation.

Figure 2.1 presents the frequency distribution of 500 estimated ESIM values for cycle 2 (Fig. 2.1a) and cycle 5 (Fig. 2.1b), obtained from one selection cycle for 500 genotypes and four traits simulated in one environment. Figure 2.1a, b indicates that the frequency distribution of the estimated ESIM values approaches normal distribution.



**a** ESIM values,  
cycle 2.

**b** ESIM values,  
Cycle 5.

**Fig. 2.1** Frequency distribution of 500 estimated eigen selection index method (ESIM) values for (a) cycle 2 and (b) cycle 5, obtained from one selection cycle for 500 genotypes and four traits simulated in one environment

## 2.2 The Linear Phenotypic Restricted Eigen Selection Index Method

Similar to the RLPSI (see Chapter.), the objective of the RESIM is to fix  $r$  of  $t$  ( $r < t$ ) traits by predicting only the genetic gains of  $(t - r)$  of them. Let  $H = \mathbf{w}'\mathbf{g}$  be the net genetic merit and  $I = \mathbf{b}'\mathbf{y}$  the ESIM index. In Chap. 2, we showed that  $\text{Cov}(I, \mathbf{g}) = \mathbf{C}\mathbf{b}$  is the covariance between the breeding value vector ( $\mathbf{g}$ ) and  $I = \mathbf{b}'\mathbf{y}$ . Thus, to fix  $r$  of  $t$  traits, we need  $r$  covariances between the linear combinations of  $\mathbf{g}$  ( $\mathbf{U}'\mathbf{g}$ ) and  $I = \mathbf{b}'\mathbf{y}$  to be zero, i.e.,  $\text{Cov}(I, \mathbf{U}'\mathbf{g}) = \mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}$ , where  $\mathbf{U}'$  is a matrix with 1s and 0s (1 indicates that the trait is restricted and 0 that the trait has no restrictions). In the RESIM, it is possible to solve this problem by maximizing

$$\rho_{HI}^2 = \frac{(\mathbf{w}'\mathbf{C}\mathbf{b})^2}{(\mathbf{w}'\mathbf{C}\mathbf{w})(\mathbf{b}'\mathbf{P}\mathbf{b})} \quad \text{with respect to vectors } \mathbf{b} \text{ and } \mathbf{w} \text{ under the restrictions}$$

$\mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}$ ,  $\mathbf{b}'\mathbf{b} = 1$ ,  $\mathbf{w}'\mathbf{C}\mathbf{w} = 1$ , and  $\mathbf{b}'\mathbf{P}\mathbf{b} = 1$ , where  $\mathbf{w}'\mathbf{C}\mathbf{w}$  is the variance of  $H = \mathbf{w}'\mathbf{g}$  and  $\mathbf{b}'\mathbf{P}\mathbf{b}$  is the variance of  $I = \mathbf{b}'\mathbf{y}$ . Also, the RESIM problem can be solved by maximizing  $\frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}}$  (Eq. 2.12) with respect to vectors  $\mathbf{b}$  only under the restrictions

$\mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}$  and  $\mathbf{b}'\mathbf{b} = 1$ , as we did to obtain Eq. (2.13). Both approaches give the same result, but it is easier to work with the second approach than with the first one.



## 2.2.1 The RESIM Parameters

To obtain the RESIM vector of coefficients that maximizes the RESIM selection response and the expected genetic gain per trait, we need to maximize the function

$$f(\mathbf{b}, \mathbf{v}') = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}} - \mathbf{v}'\mathbf{U}'\mathbf{C}\mathbf{b} \quad (2.28a)$$

with respect to  $\mathbf{b}$  and  $\mathbf{v}'$ , where  $\mathbf{v}' = [v_1 \ v_2 \ \cdots \ v_{r-1}]$  is a vector of Lagrange multipliers. The derivatives of Eq. (2.28a) with respect to  $\mathbf{b}$  and  $\mathbf{v}'$  can be written as

$$2(\mathbf{b}'\mathbf{P}\mathbf{b})^{1/2}\mathbf{C}\mathbf{b} - (\mathbf{b}'\mathbf{P}\mathbf{b})^{-1/2}(\mathbf{b}'\mathbf{C}\mathbf{b})\mathbf{P}\mathbf{b} - \mathbf{C}\mathbf{U}\mathbf{v} = \mathbf{0} \quad (2.28b)$$

and

$$\mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}, \quad (2.29)$$

respectively, where Eq. (2.29) denotes the restriction imposed for maximizing Eq. (2.28a). Using algebraic methods on Eq. (2.28b) similar to those used to obtain Eqs. (2.10) and (2.13), we get

$$\left( \mathbf{K}\mathbf{P}^{-1}\mathbf{C} - h_{I_R}^2 \mathbf{I}_t \right) \mathbf{b}_R = \mathbf{0}, \quad (2.30)$$

where  $\mathbf{K} = [\mathbf{I}_t - \mathbf{Q}_R]$ ,  $\mathbf{I}_t$  is an identity matrix of size  $t \times t$ ,  $\mathbf{Q}_R = \mathbf{P}^{-1}\mathbf{C}\mathbf{U}(\mathbf{U}'\mathbf{C}\mathbf{P}^{-1}\mathbf{C}\mathbf{U})^{-1}\mathbf{U}'\mathbf{C}$ , and  $h_{I_R}^2 = \frac{\mathbf{b}_R'\mathbf{C}\mathbf{b}_R}{\mathbf{b}_R'\mathbf{P}\mathbf{b}_R}$  is the maximized RESIM heritability obtained under the restriction  $\mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}$ ;  $h_{I_R}^2$  is also the square of the maximized correlation between the net genetic merit and  $I_R = \mathbf{b}_R'\mathbf{y}$ , that is,  $h_{I_R}^2 = \lambda_R^2$ . This means that Eq. (2.30) can be written as

$$(\mathbf{K}\mathbf{P}^{-1}\mathbf{C} - \lambda_R^2 \mathbf{I}_t) \mathbf{b}_R = \mathbf{0}. \quad (2.31)$$

Thus, the optimized RESIM index is  $I = \mathbf{b}_R'\mathbf{y}$ . The only difference between Eqs. (2.31) and (2.13) is matrix  $\mathbf{K}$ . Equation (2.31) was obtained by Cerón-Rojas et al. (2008) by maximizing  $\rho_{HI}^2$  (Eq. 2.1) with respect to vectors  $\mathbf{b}$  and  $\mathbf{w}$  under the restriction  $\mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}$ ,  $\mathbf{b}'\mathbf{b} = 1$ ,  $\mathbf{w}'\mathbf{C}\mathbf{w} = 1$  and  $\mathbf{b}'\mathbf{P}\mathbf{b} = 1$  in a similar manner to the canonical correlation theory. The RESIM expected genetic gain per trait uses the first eigenvector ( $\mathbf{b}_R$ ) of matrix  $\mathbf{K}\mathbf{P}^{-1}\mathbf{C}$ , whereas the RESIM selection response uses  $\mathbf{b}_R$  and the first eigenvalue ( $\lambda_R^2$ ) of matrix  $\mathbf{K}\mathbf{P}^{-1}\mathbf{C}$ . When  $\mathbf{U}'$  is a null matrix,  $\mathbf{b}_R = \mathbf{b}_E$  (the vector of the ESIM coefficients); thus, the RESIM is more general than the ESIM and includes the ESIM as a particular case.

In the RESIM context, vector  $\mathbf{w}$  can be obtained (Cerón-Rojas et al. 2008) as

$$\mathbf{w}_R = \mathbf{C}^{-1} [\lambda_R \mathbf{P} \mathbf{b}_R + \Psi \mathbf{v}], \quad (2.32)$$

where  $\lambda_R$  and  $\mathbf{b}_R$  are the square roots of the first eigenvalue ( $\lambda_R^2$ ) and the first eigenvector of matrix  $\mathbf{K} \mathbf{P}^{-1} \mathbf{C}$  respectively;  $\Psi = \mathbf{C} \mathbf{U}$  and  $\mathbf{v} = \lambda_R^{-1} (\Psi' \mathbf{P}^{-1} \Psi)^{-1} \Psi' \mathbf{P}^{-1} \mathbf{C} \mathbf{b}_R$ . Let  $H_R = \mathbf{w}'_R \mathbf{g}$  be the net genetic merit in the RESIM context; then, because the correlation between  $I_R = \mathbf{b}'_R \mathbf{y}$  and  $H_R = \mathbf{w}'_R \mathbf{g}$  is not affected by scale change,  $\lambda_R$  and  $\lambda_R^{-1}$  can be considered proportional constants and then  $\Psi \mathbf{v}$  can be written as  $\Psi \mathbf{v} = \Psi (\Psi' \mathbf{P}^{-1} \Psi)^{-1} \Psi' \mathbf{P}^{-1} \mathbf{C} \mathbf{b}_R = \mathbf{Q}'_R \mathbf{C} \mathbf{b}_R$ , where  $\mathbf{Q}'_R$  is the transpose of matrix  $\mathbf{Q}_R$  described in Eq. (2.30). Thus, another way of writing Eq. (2.32) is

$$\mathbf{w}_R = \mathbf{C}^{-1} [\mathbf{P} + \mathbf{Q}'_R \mathbf{C}] \mathbf{b}_R. \quad (2.33)$$

By Eq. (2.33) and the restriction  $\mathbf{b}' \Psi = \mathbf{0}$ , the covariance between  $I_R = \mathbf{b}'_R \mathbf{y}$  and  $H_R = \mathbf{w}'_R \mathbf{g}$  ( $\sigma_{H_R I_R}$ ) can be written as

$$\sigma_{H_R I_R} = \mathbf{w}'_R \mathbf{C} \mathbf{b}_R = \mathbf{b}'_R \mathbf{P} \mathbf{b}_R + \mathbf{b}'_R \mathbf{Q}'_R \mathbf{C} \mathbf{b}_R = \mathbf{b}'_R \mathbf{P} \mathbf{b}_R, \quad (2.34)$$

where  $\mathbf{b}'_R \mathbf{Q}'_R \mathbf{C} \mathbf{b}_R = 0$  according to the restriction  $\mathbf{b}' \Psi = \mathbf{0}$ . Equation (2.34) indicates that the covariance between  $I_R$  and  $H_R$  ( $\sigma_{H_R I_R}$ ) is equal to the variance of  $I_R$  ( $\sigma_{I_R}^2 = \mathbf{b}'_R \mathbf{P} \mathbf{b}_R$ ).

The maximized correlation between  $I_R$  and  $H_R$  (or RESIM accuracy) can be written as

$$\rho_{H_R I_R} = \frac{\sqrt{\mathbf{b}'_R \mathbf{P} \mathbf{b}_R}}{\sqrt{\mathbf{w}'_R \mathbf{C} \mathbf{w}_R}}, \quad (2.35)$$

where  $\mathbf{w}'_R \mathbf{C} \mathbf{w}_R = \sigma_{H_R}^2$  is the variance of  $H_R$ ,  $\mathbf{w}_R = \mathbf{C}^{-1} [\mathbf{P} + \mathbf{Q}'_R \mathbf{C}] \mathbf{b}_R$ ,  $\mathbf{Q}'_R = \Psi (\Psi' \mathbf{P}^{-1} \Psi)^{-1} \Psi' \mathbf{P}^{-1}$ , and  $\Psi = \mathbf{C} \mathbf{U}$ . When  $\mathbf{U}'$  is a null matrix,  $\mathbf{w}'_R \mathbf{C} \mathbf{w}_R = \mathbf{b}'_E \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_E = \mathbf{w}'_E \mathbf{C} \mathbf{w}_E$ , the variance of  $H_E$ , and  $\sigma_{I_E}^2 = \mathbf{b}'_R \mathbf{P} \mathbf{b}_R = \mathbf{b}'_E \mathbf{P} \mathbf{b}_E = \sigma_{I_E}^2$ , the variance of  $I_E$ . Hereafter, to simplify the notation, we write Eq. (2.35) as  $\rho_R$  or  $\lambda_R$ .

The maximized selection response ( $R_R$ ) and expected genetic gain per trait ( $\mathbf{E}_R$ ) of the RESIM can be written as

$$R_R = k_I \sqrt{\mathbf{b}'_R \mathbf{P} \mathbf{b}_R} \quad (2.36)$$

and

$$\mathbf{E}_R = k_I \frac{\mathbf{C} \mathbf{b}_R}{\sqrt{\mathbf{b}'_R \mathbf{P} \mathbf{b}_R}}, \quad (2.37)$$



respectively, where  $\sqrt{\mathbf{b}'_R \mathbf{P} \mathbf{b}_R} = \sigma_{I_R}$  is the standard deviation of the variance of  $I_R = \mathbf{b}'_R \mathbf{y}$ . If vector  $\mathbf{b}_R$  is transformed as  $\boldsymbol{\beta}_R = \mathbf{F} \mathbf{b}_R$ , where matrix  $\mathbf{F}$  was defined earlier, vector  $\mathbf{b}_R$  should be changed by  $\boldsymbol{\beta}_R$  in Eqs. (2.36) and (2.37), and in  $I_R = \mathbf{b}'_R \mathbf{y}$ .

Equation (2.36) can also be written as  $R_R = k_I \sigma_{H_R} \lambda_R$ , where  $\sigma_{H_R} = \sqrt{\mathbf{b}'_R \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_R + \mathbf{b}'_R \mathbf{P} \mathbf{C}^{-1} \mathbf{Q}'_R \mathbf{C} \mathbf{b}_R}$  is the standard deviation of the variance of  $H_R$ , and  $\lambda_R = \rho_{H_R I_R}$  is the first canonical correlation between  $H_R = \mathbf{w}'_R \mathbf{g}$  and  $I_R = \mathbf{b}'_R \mathbf{y}$ . When  $\sigma_{H_R} = 1$ ,  $\lambda_R$  is the covariance between  $H_R = \mathbf{w}'_R \mathbf{g}$  and  $I_R = \mathbf{b}'_R \mathbf{y}$ , and then Eq. (2.36) can be written as  $R_R = k_I \lambda_R$ . This last result was presented by Cerón-Rojas et al. (2008) in their original paper.

The ratio of the index accuracies and the VPE are also valid in the RESIM context. In Eq. (2.34) we showed that the covariance between  $I_R = \mathbf{b}'_R \mathbf{y}$  and  $H_R = \mathbf{w}'_R \mathbf{g}$  ( $\sigma_{H_R I_R}$ ) is equal to the variance of  $I_R = \mathbf{b}'_R \mathbf{y}$  ( $\sigma_{I_R}^2$ ). This means that the VPE of the RESIM can be written as

$$E[(H_R - I_R)^2] = \sigma_{H_R}^2 + \sigma_{I_R}^2 - 2\sigma_{H_R I_R} = \sigma_{H_R}^2 - \sigma_{I_R}^2 = (1 - \rho_R^2)\sigma_{H_R}^2. \quad (2.38)$$

Statistical properties associated with the ESIM and described in Sect. 2.1.2 are also valid for the RESIM.

## 2.2.2 Estimating the RESIM Parameters

We can estimate the RESIM parameters in a similar manner to the ESIM parameters in Sect. 2.1.4. With matrices  $\widehat{\mathbf{C}}$  and  $\widehat{\mathbf{P}}$ , we constructed matrix  $\widehat{\mathbf{S}}_R = \widehat{\mathbf{K}} \widehat{\mathbf{P}}^{-1} \widehat{\mathbf{C}}$  and equation

$$(\widehat{\mathbf{S}}_R \widehat{\mathbf{S}}'_R - \widehat{\mu}_{Rj} \mathbf{I}_t) \widehat{\mathbf{b}}_{R_j} = \mathbf{0}, \quad (2.39)$$

where  $\widehat{\mu}_{Rj} = \widehat{\lambda}_{Rj}^4$ ,  $j = 1, 2, \dots, t$ . The estimated RESIM index ( $I_R = \mathbf{b}'_R \mathbf{y}$ ) is  $\widehat{I}_R = \widehat{\mathbf{b}}'_{R_1} \mathbf{y}$  and the estimator of the maximized RESIM selection response and its expected genetic gain per trait can be denoted as  $\widehat{R}_R = k_I \sqrt{\widehat{\mathbf{b}}'_{R_1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{R_1}}$  and  $\widehat{E}_R = k_I \frac{\widehat{\mathbf{C}} \widehat{\mathbf{b}}_{R_1}}{\sqrt{\widehat{\mathbf{b}}'_{R_1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{R_1}}}$  respectively, whereas the estimator of the maximized RESIM accuracy is  $\widehat{\lambda}_{R_1}$ .

## 2.2.3 Numerical Examples

We compare the RLPSI results with those of the RESIM using the Akbar et al. (1984) data described in Sect. 2.1.5. We restrict the trait RL (number of eggs) in both indices. In Chapter, we indicated how to construct matrix  $\mathbf{U}$  and, in Sect.



of the same chapter, we described how to obtain matrix  $\mathbf{K} = \widehat{\mathbf{I}} - [\mathbf{Q}]$  for one and two restrictions. Matrix  $\widehat{\mathbf{K}}$  is the same for the RLPSI and the RESIM. Thus, in this subsection we omit the steps needed to construct matrices  $\mathbf{U}'$  and  $\widehat{\mathbf{K}}$ .

First, we estimate the RLPSI parameters. Assume a selection intensity of 10% ( $k_t = 1.755$ ) and a vector of economic weights  $\mathbf{w}' = [19.54 \quad -3.56 \quad 17.01]$ . The estimated RLPSI vector of coefficients for one restriction was  $\widehat{\mathbf{b}}' = [0.29 \quad -0.84 \quad 5.78]$ , and the estimated selection response, expected genetic gain per trait, accuracy, and heritability of the RLPSI were  $\widehat{R} = 1.755\sqrt{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}} = 53.01$ ,  $\widehat{\mathbf{E}}' = 1.755\frac{\widehat{\mathbf{b}}'\widehat{\mathbf{C}}}{\sqrt{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}}} = [0 \quad -0.71 \quad 2.96]$ ,  $\widehat{\rho} = \frac{\sqrt{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}}}{\sqrt{\mathbf{w}'\widehat{\mathbf{C}}\mathbf{w}}} = 0.26$ , and  $\widehat{h}^2 = \frac{\widehat{\mathbf{b}}'\widehat{\mathbf{C}}\widehat{\mathbf{b}}}{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}} = 0.33$  respectively. In this case,  $\widehat{\mathbf{b}}'\widehat{\mathbf{b}} = 34.25$ ; then, the estimated RLPSI selection response using the normalized RLPSI vector of coefficients was  $\widehat{R} = \frac{53.01}{34.25} = 1.55$ , and the rest of the estimated RLPSI parameters were the same.

In the RESIM, matrix  $\mathbf{F}$  was an identity matrix of size  $3 \times 3$ ; that is, we did not use matrix  $\mathbf{F}$  to transform the RESIM vector of coefficients. In Sect. 2.1.5 we

obtained matrix  $\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}} = \begin{bmatrix} 0.1102 & -0.0405 & -0.0280 \\ -0.0390 & 0.0864 & -0.0184 \\ -0.1833 & 0.0517 & 0.4115 \end{bmatrix}$ , and we have indi-

cated that matrix  $\widehat{\mathbf{K}}$  is the same for the RLPSI and the RESIM. In the RESIM, we need matrix  $\widehat{\mathbf{S}}_R = \widehat{\mathbf{K}}\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}}$  to solve equation  $(\widehat{\mathbf{S}}_R\widehat{\mathbf{S}}'_R - \widehat{\mu}_{Rj}\mathbf{I}_t)\widehat{\mathbf{b}}_{Rj} = \mathbf{0}$ , where  $\widehat{\mu}_{Rj} = \widehat{\lambda}_{Rj}^4$ , whence we shall obtain the eigenvalues and eigenvectors that form matrices  $\mathbf{L}_R^{1/2}$ ,  $\mathbf{V}_{R1}$ , and  $\widehat{\mathbf{S}}_R = \mathbf{V}_{R1}\mathbf{L}_R^{1/2}\mathbf{V}'_{R2}$ .

For one null restriction, matrix  $\widehat{\mathbf{S}}_R = \widehat{\mathbf{K}}\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}} = \begin{bmatrix} 0 & 0.0285 & 0.0232 \\ 0 & 0.0620 & -0.0365 \\ 0 & -0.0630 & 0.3263 \end{bmatrix}$ .

This means that  $\widehat{\mathbf{S}}_R$  reflects the trait restrictions imposed on the covariance between the RESIM and the vector of genotypic values; thus, if  $r$  traits are restricted,  $r$  columns of  $\widehat{\mathbf{S}}_R$  are equal to zero. Matrix  $\widehat{\mathbf{S}}_R\widehat{\mathbf{S}}'_R = \begin{bmatrix} 0.0013 & 0.0009 & 0.0058 \\ 0.0009 & 0.0052 & -0.0158 \\ 0.0058 & -0.0158 & 0.1104 \end{bmatrix}$  and  $\mathbf{V}_{R1} = \begin{bmatrix} 0.0500 & 0.5216 & -0.8517 \\ -0.1446 & 0.8476 & 0.5106 \\ 0.9882 & 0.0976 & 0.1178 \end{bmatrix}$ , whereas the  $\widehat{\mu}_{Rj} = \widehat{\lambda}_{Rj}^4$  values were 0.1130, 0.0039, and 0.0, whence  $\mathbf{L}_R^{1/2} = \begin{bmatrix} 0.3362 & 0 & 0 \\ 0 & 0.0626 & 0 \\ 0 & 0 & 0.0 \end{bmatrix}$ . Thus,  $\widehat{\mu}_{R1} = \widehat{\lambda}_{R1}^4 = 0.1130$ ,  $\widehat{\lambda}_{R1}^2 = 0.3362$ , and the

estimated RESIM accuracy was  $\widehat{\lambda}_{E1} = 0.5798$ . The estimated RESIM eigenvector, index, the selection response, and expected genetic gain per trait were  $\widehat{\mathbf{b}}'_{R1} = [0.0500 \quad -0.1446 \quad 0.9882]$ ,  $\widehat{I}_R = 0.0500\text{RL} - 0.1446\text{SM} + 0.9882\text{EW}$ ,



$$\hat{R}_R = 1.755 \sqrt{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{P}} \hat{\mathbf{b}}_{R_1}} = 9.06, \text{ and } \hat{\mathbf{E}}'_R = 1.755 \frac{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{C}}}{\sqrt{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{P}} \hat{\mathbf{b}}_{R_1}}} = [0 \ -0.72 \ 2.96]$$

respectively.

The estimated RLPSI selection response was  $\hat{R} = \frac{53.01}{34.25} = 1.55$ ; thus, the estimated RESIM selection response was higher than the estimated RLPSI response. In addition, the estimated RLPSI expected genetic gain per trait was  $\hat{\mathbf{E}}' = [0 \ -0.71 \ 2.96]$ , which is the same as the estimated RESIM expected genetic gain per trait.

We can compare RESIM efficiency versus RLPSI efficiency to predict the net genetic merit using the ratio of the estimated RESIM accuracy  $\hat{\lambda}_{E_1} = 0.5798$  to the RLPSI accuracy  $\hat{\rho} = 0.26$ , i.e.,  $\frac{\hat{\lambda}_{R_1}}{\hat{\rho}_S} = \frac{0.5798}{0.26} = 2.23$ , or in percentage terms,  $\hat{\rho}_E = 100(2.23 - 1) = 123$  (see Chapter.) That is, the RESIM is a better predictor of the net genetic merit and its efficiency was 123% higher than the RLPSI efficiency for this data set.

Now, we compare RESIM efficiency versus RLPSI efficiency using the simulated data set described in Sect. 2.8.1 of Chap. 2 for five phenotypic selection cycles, each with four traits ( $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ ), 500 genotypes, and four replicates for 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, matrix  $\mathbf{F}$  was equal to an identity matrix of size  $4 \times 4$  for all five selection cycles.

The first and second parts of columns 3, 4, and 5 of Table 2.1 present the estimated RLPSI and RESIM selection responses respectively for one, two, and three null restrictions for five simulated selection cycles, where the selection intensity was 10% ( $k_I = 1.755$ ) for all five selection cycles. The averages of the estimated RLPSI selection response for each null restriction were 4.43, 4.30, and 4.92, whereas the averages of the estimated RESIM selection response were 4.54, 4.42, and 4.38 respectively. These results indicate that the estimated RLPSI selection response was greater than the estimated RESIM selection response only for three null restrictions.

The first part of Table 2.3 presents the estimated RESIM expected genetic gain per trait for one, two, and three restrictions for five simulated selection cycles. The estimated RLPSI expected genetic gains per trait for one, two, and three restrictions are given in Chap. According to the results shown in the averages of the estimated RLPSI expected genetic gains per trait for five simulated selection cycles were -2.52, 2.25, and 2.26 for one restriction; 2.84 and 2.65 for two restrictions; and 3.90 for three restrictions. According to the results shown in Table 2.3, the averages of the estimated RESIM expected genetic gains per trait for five simulated selection cycles were -0.43, -0.75, and 3.90 for one restriction; -0.59 and 3.89 for two restrictions; and 3.90 for three restrictions. This means that the RESIM and RLPSI were the same only for three restrictions, whereas for one and two restrictions, the average of the estimated RESIM expected



**Table 2.3** Estimated RESIM and PPG-ESIM expected genetic gain per trait for one, two, and three restrictions for five simulated selection cycles

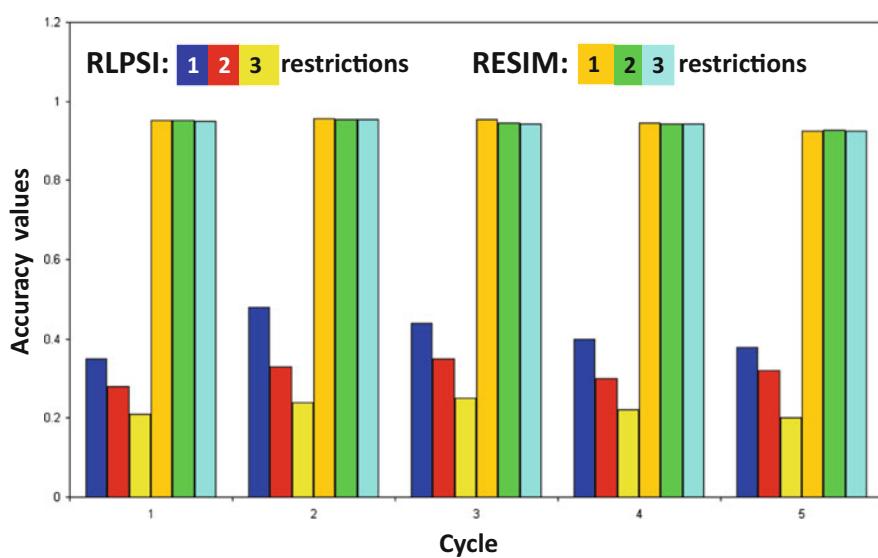
Cycle	Estimated RESIM expected genetic gain per trait											
	One null restriction				Two null restrictions				Three null restrictions			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
1	0	-0.86	-1.22	4.14	0	0	-0.96	4.12	0	0	0	4.13
2	0	-1.38	-0.004	4.31	0	0	-0.07	4.27	0	0	0	4.27
3	0	1.36	-1.74	4.07	0	0	-1.39	4.09	0	0	0	4.16
4	0	-1.13	-0.34	3.73	0	0	-0.08	3.72	0	0	0	3.72
5	0	-0.14	-0.43	3.22	0	0	-0.43	3.22	0	0	0	3.24
Average	0	-0.43	-0.75	3.90	0	0	-0.59	3.89	0	0	0	3.90
Cycle	Estimated PPG-ESIM expected genetic gain per trait											
	One predetermined restriction				Two predetermined restrictions				Three predetermined restrictions			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
1	7.81	-4.62	3.11	2.21	7.09	-3.04	3.12	2.76	6.62	-2.84	4.73	0.83
2	5.15	-2.98	2.31	3.48	5.41	-2.32	2.41	3.48	6.14	-2.63	4.39	0.92
3	4.74	-1.15	0.66	3.79	5.45	-2.34	1.24	3.26	5.52	-2.37	3.94	1.35
4	3.94	-2.44	0.74	3.34	4.57	-1.96	1.17	3.24	5.03	-2.15	3.59	0.30
5	6.68	-2.15	2.24	2.05	6.93	-2.97	2.25	1.4	5.25	-2.25	3.75	0.72
Average	5.67	-2.67	1.81	2.97	5.89	-2.52	2.04	2.83	5.71	-2.45	4.08	0.82

The selection intensity was 10% ( $k_I = 1.755$ ) and the vectors of the PPG for each predetermined restriction were  $\mathbf{d}'_1 = 7$ ,  $\mathbf{d}'_2 = [7 \ -3]$  and  $\mathbf{d}'_3 = [7 \ -3 \ 5]$  respectively

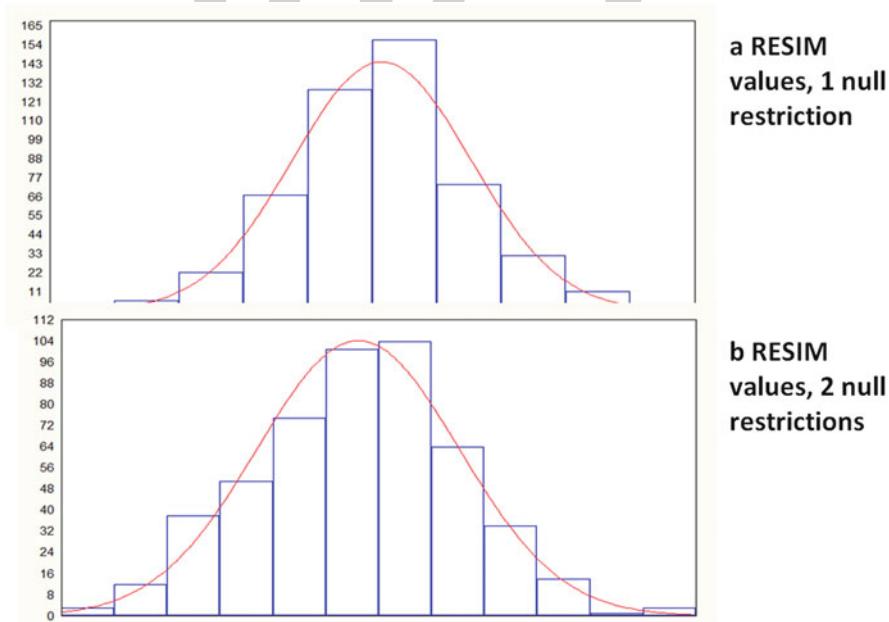
genetic gains per trait was higher than that of the estimated RLPSI expected genetic gains per trait only for trait 4.

Figure 7.2 presents the estimated accuracy of the RLPSI and the RESIM for one, two, and three null restrictions for five simulated selection cycles. In all five selection cycles, the estimated RESIM accuracy was greater than the RLPSI accuracy. This means that the RESIM is a better predictor of the net genetic merit than the RLPSI. Additional results associated with the frequency distribution of the estimated RESIM values are presented in Fig. 2.3. Figure 2.3a presents the frequency distribution of the estimated RESIM values with one null restriction for cycle 2, whereas Fig. 2.3b presents the frequency distribution of the estimated RESIM values with two null restrictions for cycle 5; both figures indicate that the estimated RESIM values approach normal distribution.

Finally, in Chapter 5 we present the results of comparing the ESIM with the LPSI and the RESIM with the RLPSI for many selection cycles. Such results are similar to those obtained in this chapter.



**Fig. 2.2** Estimated correlation values between the restricted linear phenotypic selection index (RLPSI) and the net genetic merit ( $H = \mathbf{w} \cdot \mathbf{g}$ ); estimated correlation values between the restricted eigen selection index method (RESIM) and  $H$  for one, two and three null restrictions for four traits and 500 genotypes in one environment simulated for five selection cycles



**Fig. 2.3** Frequency distribution of 500 estimated RESIM values for (a) cycle 2 and (b) cycle 5, obtained from one selection cycle for 500 genotypes and four traits simulated in one environment



## 2.3 The Linear Phenotypic Predetermined Proportional Gain Eigen Selection Index Method

In a similar manner to the PPG-LPSI (see Chapter.), in the PPG-ESIM the breeder pre-sets optimal levels (predetermined proportional gains) on certain traits before the selection is carried out. Let  $\mathbf{d}' = [d_1 \ d_2 \ \dots \ d_r]$  be the vector of the PPGs (predetermined proportional gains) imposed by the breeder on  $r$  traits and assume that  $\mu_q$  is the population mean of the  $q$ th trait before selection. The objective of the PPG-ESIM is to change  $\mu_q$  to  $\mu_q + d_q$ , where  $d_q$  is a predetermined change in  $\mu_q$  (in the RESIM,  $d_q = 0$ ,  $q = 1, 2, \dots, r$ , where  $r$  is the number of PPGs). That is, the PPG-ESIM attempts to make some traits change their expected genetic gain values based on a predetermined level, whereas the rest of the traits remain without restrictions.

The simplest way to solve the foregoing problem is by maximizing the PPG-ESIM heritability under the restriction  $\mathbf{D}'\mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}$ , where

$$\mathbf{D}' = \begin{bmatrix} d_r & 0 & \cdots & 0 & -d_1 \\ 0 & d_r & \cdots & 0 & -d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_r & -d_{r-1} \end{bmatrix} \quad (\text{see Chap. 3 for details})$$

(see Chap. 3 for details)

(see Chap. 3 for details) is a matrix  $(r - 1) \times r$ ,

$r$  is the number of PPGs,  $d_q$  ( $q = 1, 2, \dots, r$ ) is the  $q$ th element of vector  $\mathbf{d}'$ ,  $\mathbf{U}'$  is the RLPSI matrix of restrictions of 1s and 0s, and  $\mathbf{C}$  is the covariance matrix of genotypic values. Matrix  $\mathbf{D}'$  is a Mallard (1972) matrix of PPGs used to impose predetermined restrictions.

The Mallard (1972) matrix of predetermined restrictions can be written as  $\mathbf{M}' = \mathbf{D}'\Psi'$ , where  $\Psi' = \mathbf{U}'\mathbf{C}$  and  $\mathbf{U}'$  is the Kempthorne and Nordskog (1959) matrix of restrictions of 1s and 0s (1 indicates that the trait is restricted, i.e.,  $d_q = 0$ , and 0 that the trait has no restrictions).

To find the PPG-ESIM vector of coefficients that maximizes the PPG-ESIM selection response and expected genetic gain per trait, we can maximize  $\rho_{HI}^2 = \frac{(\mathbf{w}'\mathbf{C}\mathbf{b})^2}{(\mathbf{w}'\mathbf{C}\mathbf{w})(\mathbf{b}'\mathbf{P}\mathbf{b})}$  with respect to vectors  $\mathbf{b}$  and  $\mathbf{w}$  under the restrictions  $\mathbf{M}'\mathbf{b} = \mathbf{0}$ ,  $\mathbf{b}'\mathbf{b} = 1$ ,  $\mathbf{w}'\mathbf{C}\mathbf{w} = 1$ , and  $\mathbf{b}'\mathbf{P}\mathbf{b} = 1$ , where  $\mathbf{w}'\mathbf{C}\mathbf{w}$  is the variance of  $H = \mathbf{w}'\mathbf{g}$  and  $\mathbf{b}'\mathbf{P}\mathbf{b}$  is the variance of  $I = \mathbf{b}'\mathbf{y}$ , as did Cerón-Rojas et al. (2016) according to the canonical correlation theory, or we can solve this problem by maximizing  $\frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}}$  (Eq. 2.12) only with respect to vectors  $\mathbf{b}$  under the restriction  $\mathbf{M}'\mathbf{b} = \mathbf{0}$  and  $\mathbf{b}'\mathbf{b} = 1$ , as we did to obtain the RESIM vector of coefficients. Both approaches give the same result, but we use the latter approach because it is easier to work with.



### 2.3.1 The PPG-ESIM Parameters

To obtain the PPG-ESIM vector of coefficients, we need to maximize the function

$$f(\mathbf{b}, \mathbf{v}') = \frac{\mathbf{b}' \mathbf{C} \mathbf{b}}{\sqrt{\mathbf{b}' \mathbf{P} \mathbf{b}}} - \mathbf{v}' \mathbf{M}' \mathbf{b} \quad (2.40)$$

with respect to vectors  $\mathbf{b}$  and  $\mathbf{v}'$ , where  $\mathbf{v}' = [v_1 \ v_2 \ \dots \ v_{r-1}]$  is a vector of Lagrange multipliers. The derivatives of Eq. (2.40) with respect to  $\mathbf{b}$  and  $\mathbf{v}'$  were:

$$2(\mathbf{b}' \mathbf{P} \mathbf{b})^{1/2} \mathbf{C} \mathbf{b} - (\mathbf{b}' \mathbf{P} \mathbf{b})^{-1/2} (\mathbf{b}' \mathbf{C} \mathbf{b}) \mathbf{P} \mathbf{b} - \mathbf{M} \mathbf{v} = \mathbf{0} \quad (2.41)$$

and

$$\mathbf{M}' \mathbf{b} = \mathbf{0}, \quad (2.42)$$

respectively, where Eq. (2.42) denotes the restriction imposed for maximizing Eq. (2.40). By using algebraic methods on Eq. (2.41) similar to those used to obtain Eq. (2.10) we get

$$(\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C} - \lambda_P^2 \mathbf{I}_t) \mathbf{b}_P = \mathbf{0}, \quad (2.43)$$

where  $\mathbf{K}_P = [\mathbf{I}_t - \mathbf{Q}_P]$ ,  $\mathbf{Q}_P = \mathbf{P}^{-1} \Psi \mathbf{D} (\mathbf{D}' \Psi' \mathbf{P}^{-1} \Psi \mathbf{D})^{-1} \mathbf{D}' \Psi'$ ,  $\Psi' = \mathbf{U}' \mathbf{C}$ ,  $\mathbf{I}_t$  is an identity matrix  $t \times t$ ,  $\lambda_P^2 = h_{I_P}^2$ , and  $\mathbf{b}_P$  are the first eigenvalue and the first eigenvector of matrix  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C}$  respectively. Note that  $h_{I_P}^2$  is PPG-ESIM heritability and  $\lambda_P$  is the maximum correlation between  $I_P = \mathbf{b}'_P \mathbf{y}$  and  $H = \mathbf{w}' \mathbf{g}$ . When  $\mathbf{D}' = \mathbf{U}'$ ,  $\mathbf{b}_P = \mathbf{b}_R$  (the vector of coefficients of the RESIM), and when  $\mathbf{U}'$  is a null matrix,  $\mathbf{b}_P = \mathbf{b}_E$  (the vector of coefficients of the ESIM). That is, the PPG-ESIM is more general than the RESIM and the ESIM and includes the latter two indices as particular cases. Matrices  $\mathbf{K}_P = [\mathbf{I}_t - \mathbf{Q}_P]$  and  $\mathbf{Q}_P = \mathbf{P}^{-1} \Psi \mathbf{D} (\mathbf{D}' \Psi' \mathbf{P}^{-1} \Psi \mathbf{D})^{-1} \mathbf{D}' \Psi'$  are the same as those obtained in the PPG-LPSI (see Chapter.). Also, vector  $\mathbf{b}_P$  can be transformed as  $\mathbf{b}_P = \mathbf{F} \mathbf{b}_P$ ; matrix  $\mathbf{F}$  was defined earlier.

Let  $\mathbf{S}_P = \Psi' \mathbf{P}^{-1} \Psi$ ; then, under the assumption  $\mathbf{D}' \mathbf{d} = \mathbf{0}$ , it is possible to show that  $\mathbf{D} (\mathbf{D}' \mathbf{S}_P \mathbf{D})^{-1} \mathbf{D}' = \mathbf{S}_P^{-1} - \mathbf{S}_P^{-1} \mathbf{d} (\mathbf{d}' \mathbf{S}_P^{-1} \mathbf{d})^{-1} \mathbf{d}' \mathbf{S}_P^{-1}$  (see Chapter). whence by substituting  $\mathbf{S}_P^{-1} - \mathbf{S}_P^{-1} \mathbf{d} (\mathbf{d}' \mathbf{S}_P^{-1} \mathbf{d})^{-1} \mathbf{d}' \mathbf{S}_P^{-1}$  for  $\mathbf{D} (\mathbf{D}' \mathbf{S}_P \mathbf{D})^{-1} \mathbf{D}'$  in matrix  $\mathbf{Q}_P = \mathbf{P}^{-1} \Psi \mathbf{D} (\mathbf{D}' \Psi' \mathbf{P}^{-1} \Psi \mathbf{D})^{-1} \mathbf{D}' \Psi'$ , matrix  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C}$  can be written as

$$\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C} = [\mathbf{I}_t - \mathbf{P}^{-1} \Psi \mathbf{S}^{-1} \Psi'] \mathbf{P}^{-1} \mathbf{C} + \mathbf{A}_P, \quad (2.44)$$

where  $\Psi' = \mathbf{U}' \mathbf{C}$ ,  $\mathbf{A}_P = \boldsymbol{\delta} \boldsymbol{\alpha}'$ ,  $\boldsymbol{\delta} = \mathbf{P}^{-1} \Psi (\Psi' \mathbf{P}^{-1} \Psi)^{-1} \mathbf{d}$ , and  $\boldsymbol{\alpha}' = \frac{\mathbf{d}' \mathbf{S}^{-1} \Psi' \mathbf{P}^{-1} \mathbf{C}}{\mathbf{d}' \mathbf{S}^{-1} \mathbf{d}}$ . When  $\mathbf{A}_P$  is a null matrix,  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C} = \mathbf{K} \mathbf{P}^{-1} \mathbf{C}$  (matrix of the RESIM), and if  $\mathbf{U}'$  is a null matrix,  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C} = \mathbf{P}^{-1} \mathbf{C}$  (matrix of the ESIM), this means that Eq. (2.44) is a mathematical equivalent form of matrix  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C}$  and that Eq. (2.44) does not require matrix  $\mathbf{D}'$ . The easiest way to obtain  $\mathbf{b}_P$  and  $\lambda_P$  is to use matrix  $[\mathbf{I}_t - \mathbf{P}^{-1} \Psi \mathbf{S}^{-1} \Psi'] \mathbf{P}^{-1} \mathbf{C} + \mathbf{A}_P$  in Eq. (2.43) instead of matrix  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C}$ .

In the PPG-ESIM context, vector  $\mathbf{w}$  can be obtained as

$$\mathbf{w}_P = \mathbf{C}^{-1} [\lambda_P \mathbf{P} \mathbf{b}_P + \mathbf{M} \mathbf{v}_P], \quad (2.45)$$

whence  $H = \mathbf{w}'\mathbf{g}$  can be written as  $H_P = \mathbf{w}'_P \mathbf{g}$ . In Eq. (2.45),  $\lambda_P$  is the maximum correlation between  $I_P = \mathbf{b}'_P \mathbf{y}$  and  $H_P = \mathbf{w}'_P \mathbf{g}$ ,  $\mathbf{b}_P$  is the first eigenvector of matrix  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C}$ ,  $\mathbf{v}_P = \lambda_P^{-1} (\mathbf{M}' \mathbf{P}^{-1} \mathbf{M})^{-1} \mathbf{M}' \mathbf{P}^{-1} \mathbf{C} \mathbf{b}_P$ ,  $\mathbf{M}' = \mathbf{D}' \Psi'$ , and  $\Psi' = \mathbf{U}' \mathbf{C}$ . In a similar manner to the RESIM context, we can assume that  $\lambda_P$  and  $\lambda_P^{-1}$  are proportionality constants and it can be shown that the covariance between  $I_P = \mathbf{b}'_P \mathbf{y}$  and  $H_P = \mathbf{w}'_P \mathbf{g}$  ( $\sigma_{H_P I_P}$ ) is equal to the variance of  $I_P = \mathbf{b}'_P \mathbf{y}$  ( $\sigma_{I_P}^2 = \mathbf{b}'_P \mathbf{P} \mathbf{b}_P$ ), that is,  $\sigma_{H_P I_P} = \mathbf{w}'_P \mathbf{C} \mathbf{b}_P = \mathbf{b}'_P \mathbf{P} \mathbf{b}_P$ .

The accuracy of the PPG-ESIM can also be written as

$$\rho_{H_P I_P} = \frac{\sqrt{\mathbf{b}'_P \mathbf{P} \mathbf{b}_P}}{\sqrt{\mathbf{w}'_P \mathbf{C} \mathbf{w}_P}}, \quad (2.46)$$

where  $\sigma_{H_P}^2 = \mathbf{w}'_P \mathbf{C} \mathbf{w}_P = \mathbf{b}'_P \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_P + \mathbf{b}'_P \mathbf{P} \mathbf{C}^{-1} \mathbf{Q}'_P \mathbf{C} \mathbf{b}_P$  is the variance of  $H_P$ . When  $\mathbf{D}' = \mathbf{U}'$ ,  $\mathbf{w}'_P \mathbf{C} \mathbf{w}_P = \mathbf{w}'_R \mathbf{C} \mathbf{w}_R$  (the variance of  $H_R$ ), and when  $\mathbf{U}'$  is a null matrix,  $\mathbf{w}'_P \mathbf{C} \mathbf{w}_P = \mathbf{w}'_E \mathbf{C} \mathbf{w}_E$  (the variance of  $H_E$ ). Hereafter, to simplify the notation, we write Eq. (2.46) as  $\rho_P$  or  $\lambda_P$ .

Let  $\boldsymbol{\beta}_P = \mathbf{F} \mathbf{b}_P$  be the PPG-ESIM transformed vector of coefficients by matrix  $\mathbf{F}$ . By Eqs. (2.1) and (2.46), the maximized selection response ( $R_P$ ) and expected genetic gain per trait ( $\mathbf{E}_P$ ) of the PPG-ESIM can be written as

$$R_P = k_I \sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P} \quad (2.47)$$

and

$$\mathbf{E}_P = k_I \frac{\mathbf{C} \boldsymbol{\beta}_P}{\sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P}}, \quad (2.48)$$

respectively, where  $\sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P} = \sigma_{I_P}$  is the standard deviation of the variance of  $I_P = \boldsymbol{\beta}'_P \mathbf{y}$ . Equations (2.47) and (2.48) do not require economic weights. When  $\mathbf{F}$  is an identity matrix,  $\boldsymbol{\beta}_P = \mathbf{b}_P$ ,  $I_P = \boldsymbol{\beta}'_P \mathbf{y}$ ,  $R_P = k_I \sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P}$ , and  $\mathbf{E}_P = k_I \frac{\mathbf{C} \mathbf{b}_P}{\sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P}}$ .

Equation (2.47) can also be written as  $R_P = k_I \sigma_{H_P} \lambda_P$ , where  $\sigma_{H_P} = \sqrt{\mathbf{b}'_P \mathbf{P} \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_P + \mathbf{b}'_P \mathbf{P} \mathbf{C}^{-1} \mathbf{Q}'_P \mathbf{C} \mathbf{b}_P}$  is the standard deviation of the variance of  $H_P$ , and  $\lambda_P$  is the canonical correlation between  $H_P$  and  $I_P = \boldsymbol{\beta}'_P \mathbf{y}$ . When  $\sigma_{H_P} = 1$ , Eq. (2.47) can be written as  $R_P = k_I \lambda_P$ , where  $\lambda_P$  is the covariance between  $I_P = \boldsymbol{\beta}'_P \mathbf{y}$  and  $H = \mathbf{w}'_P \mathbf{g}$ .

The prediction efficiency of the PPG-ESIM can be obtained in a similar manner to the ESIM and RESIM. The accuracy of the PPG-ESIM (Eq. 2.46) can be used to construct the ratio of index accuracies. The PPG-ESIM mean square error or the VPE can be obtained as



$$E[(H_P - I_P)^2] = \sigma_{H_P}^2 + \sigma_{I_P}^2 - 2\sigma_{H_P I_P} = \sigma_{H_P}^2 - \sigma_{I_P}^2 = (1 - \rho_P^2)\sigma_{H_P}^2. \quad (2.49)$$

Additional properties associated with the ESIM are also valid for the PPG-ESIM.

### 2.3.2 Estimating PPG-ESIM Parameters

The procedure used to estimate PPG-ESIM parameters is the same as that described for RESIM. Let  $\hat{\mathbf{C}}$  and  $\hat{\mathbf{P}}$  be the estimated matrices of  $\mathbf{C}$  and  $\mathbf{P}$ . In the PPG-ESIM context, we use matrix  $\hat{\mathbf{S}} = \hat{\mathbf{K}}_P \hat{\mathbf{P}}^{-1} \hat{\mathbf{C}}$  to obtain the estimated eigenvalues and eigenvectors of equation

$$(\hat{\mathbf{S}} - \hat{\lambda}_{Pj}^2 \mathbf{I}_t) \hat{\mathbf{b}}_{Pj} = \mathbf{0}, \quad (2.50)$$

$j = 1, 2, \dots, t$ , where  $t$  is the number of traits in the PPG-ESIM index,  $\hat{\mathbf{K}}_P = [\mathbf{I}_t - \hat{\mathbf{Q}}_P]$ ,  $\mathbf{I}_t$  is an identity matrix of size  $t \times t$  and  $\hat{\mathbf{Q}}_P = \hat{\mathbf{P}}^{-1} \hat{\mathbf{\Psi}} \mathbf{D} (\mathbf{D}' \hat{\mathbf{\Psi}}' \hat{\mathbf{P}}^{-1} \hat{\mathbf{\Psi}} \mathbf{D})^{-1} \mathbf{D}' \hat{\mathbf{\Psi}}'$ . As  $\hat{\mathbf{S}}$  is an asymmetric matrix, the values of  $\hat{\mathbf{b}}_{Pj}$  and  $\hat{\lambda}_{Pj}^2$  should be obtained using SVD (singular value decomposition).

According to SVD, we need to solve equation

$$(\hat{\mathbf{S}} \hat{\mathbf{S}}' - \hat{\mu}_{Pj} \mathbf{I}_t) \hat{\mathbf{b}}_{Pj} = \mathbf{0}, \quad (2.51)$$

where  $\hat{\mu}_{Pj} = \hat{\lambda}_{Pj}^4$  ( $j = 1, 2, \dots, t$ ). By Eq. (2.51), the estimated PPG-ESIM index ( $I_P = \mathbf{b}'_P \mathbf{y}$ ) is  $\hat{I}_P = \hat{\mathbf{b}}'_P \mathbf{y}$ . The estimator of the maximized PPG-ESIM selection response, and its expected genetic gain per trait, can be denoted as  $\hat{R}_P = k_I \sqrt{\hat{\mathbf{b}}'_P \hat{\mathbf{P}} \hat{\mathbf{b}}_P}$  and  $\hat{E}_P = k_I \frac{\hat{\mathbf{C}} \hat{\mathbf{b}}_P}{\sqrt{\hat{\mathbf{b}}'_P \hat{\mathbf{P}} \hat{\mathbf{b}}_P}}$  respectively, whereas the estimator of the maximized accuracy of the PPG-ESIM is  $\hat{\lambda}_{Pj}$ .

### 2.3.3 Numerical Examples

We compare the results of the PPG-LPSI and the PPG-ESIM using the Akbar et al. (1984) data described earlier. We restrict traits RL and SM, on both indices using the PPG vector  $\mathbf{d}' = [3 \ -1]$ . In Chapter. we indicated how to construct matrix  $\mathbf{U}'$  and, in of the same chapter, we described how to obtain matrix  $\hat{\mathbf{K}}_P$  for one and two restrictions. Matrix  $\hat{\mathbf{K}}_P$  is the same for the PPG-LPSI and the PPG-ESIM. Thus, we omit the steps for constructing matrices  $\mathbf{U}'$  and  $\hat{\mathbf{K}}_P$ .

Assume a selection intensity of 10% ( $k_I = 1.755$ ) and that the vector of economic weights is  $\mathbf{w}' = [19.54 \ -3.56 \ 17.01]$ . The estimated PPG-LPSI vector of coefficients for two predetermined restrictions was  $\hat{\mathbf{b}}' = [1.70 \ 1.04 \ 2.93]$ , and its

estimated selection response, expected genetic gain per trait, accuracy, and heritability were  $\hat{R} = 1.755\sqrt{\hat{\mathbf{b}}'\hat{\mathbf{P}}\hat{\mathbf{b}}} = 49.02$ ,  $\hat{\mathbf{E}}' = 1.755\frac{\hat{\mathbf{b}}'\hat{\mathbf{C}}}{\sqrt{\hat{\mathbf{b}}'\hat{\mathbf{P}}\hat{\mathbf{b}}}} = [1.25 \quad -0.42 \quad 1.36]$ ,  $\hat{\rho} = \frac{\sqrt{\hat{\mathbf{b}}'\hat{\mathbf{P}}\hat{\mathbf{b}}}}{\sqrt{\mathbf{w}'\hat{\mathbf{C}}\mathbf{w}}} = 0.24$ , and  $\hat{h}^2 = \frac{\hat{\mathbf{b}}'\hat{\mathbf{C}}\hat{\mathbf{b}}}{\hat{\mathbf{b}}'\hat{\mathbf{P}}\hat{\mathbf{b}}} = 0.12$  respectively. In this case,  $\hat{\mathbf{b}}'\hat{\mathbf{b}} = 12.57$ ; then, the estimated PPG-LPSI selection response using the normalized PPG-LPSI vector of coefficients was  $\hat{R} = \frac{49.02}{12.57} = 3.90$ , whereas the rest of the estimated PPG-LPSI parameters were the same.

In the PPG-ESIM, we need matrix  $\hat{\mathbf{S}} = \hat{\mathbf{K}}_P\hat{\mathbf{P}}^{-1}\hat{\mathbf{C}}$  to obtain the eigenvalues and eigenvectors of  $(\hat{\mathbf{S}}\hat{\mathbf{S}}' - \hat{\mu}_{P_j}\mathbf{I}_t)\hat{\mathbf{b}}_{P_j} = \mathbf{0}$  that make up matrices  $\mathbf{L}_P^{1/2}$ ,  $\mathbf{V}_{P_1}$ , and  $\hat{\mathbf{S}} = \mathbf{V}_{P_1}\mathbf{L}_P^{1/2}\mathbf{V}_{P_2}'$ , where  $\hat{\mu}_{P_j} = \hat{\lambda}_{P_j}^4$ . It can be shown that  $\hat{\mathbf{S}} = \hat{\mathbf{K}}_P\hat{\mathbf{P}}^{-1}\hat{\mathbf{C}} = \begin{bmatrix} 0.1047 & -0.0349 & -0.0279 \\ 0.0678 & -0.0226 & -0.0213 \\ -0.1970 & 0.0657 & 0.4119 \end{bmatrix}$ ,  $\hat{\mathbf{S}}\hat{\mathbf{S}}' = \begin{bmatrix} 0.0130 & 0.0085 & -0.0344 \\ 0.0085 & 0.0056 & -0.0236 \\ -0.0344 & -0.0236 & 0.2118 \end{bmatrix}$ , and  $\mathbf{V}_{P_1} = \begin{bmatrix} -0.1663 & 0.8292 & 0.5336 \\ -0.1138 & 0.5214 & -0.8457 \\ 0.9795 & 0.2014 & -0.0076 \end{bmatrix}$ , whereas the  $\hat{\mu}_{P_j} = \hat{\lambda}_{P_j}^4$  values were 0.2214, 0.0099, and 0.0, whence  $\mathbf{L}_P^{1/2} = \begin{bmatrix} 0.4705 & 0 & 0 \\ 0 & 0.0997 & 0 \\ 0 & 0 & 0.0 \end{bmatrix}$ . Thus,  $\hat{\mu}_{P_1} = \hat{\lambda}_{P_1}^4 = 0.2214$ ,  $\hat{\lambda}_{P_1}^2 = 0.4705$ , and the estimated maximized PPG-ESIM accuracy was  $\hat{\lambda}_{P_1} = 0.6859$ .

We transformed the first eigenvector  $\hat{\mathbf{b}}'_{P_1} = [-0.1663 \quad -0.1138 \quad 0.9795]$  using matrix  $\mathbf{F} = \begin{bmatrix} -9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$  to obtain vector  $\hat{\beta}_P = \hat{\mathbf{b}}'_{P_1}\mathbf{F} = [1.4968 \quad -0.1138 \quad 0.9795]$  and  $\hat{\beta}'_P\hat{\beta}_P = 3.21$ , whence the estimates of the index, the selection response, and expected genetic gain per trait of the PPG-ESIM were  $\hat{I}_P = 1.4968\text{RL} - 0.1138\text{SM} + 0.9795\text{EW}$ ,  $\hat{R}_P = \frac{1.755\sqrt{\hat{\beta}'_P\hat{\mathbf{P}}\hat{\beta}_P}}{\hat{\beta}'_P\hat{\beta}_P} = \frac{43.01}{3.21} = 13.39$ , and  $\hat{\mathbf{E}}'_P = 1.755\frac{\hat{\beta}'_P\hat{\mathbf{C}}}{\sqrt{\hat{\beta}'_P\hat{\mathbf{P}}\hat{\beta}_P}} = [3.05 \quad -1.96 \quad 0.19]$  respectively. The estimated PPG-LPSI selection response was  $\hat{R} = \frac{49.02}{12.57} = 3.90$ , which means that the estimated PPG-ESIM selection response was greater than the estimated PPG-LPSI response.

We compared PPG-ESIM efficiency versus LPSI efficiency to predict the net genetic merit using the ratio of the estimated PPG-ESIM accuracy ( $\hat{\lambda}_{P_1} = 0.6859$ ) to PPG-LPSI accuracy ( $\hat{\rho} = 0.24$ ), i.e.,  $\frac{\hat{\lambda}_{P_1}}{\hat{\rho}} = \frac{0.6859}{0.24} = 2.858$  or, in percentage terms,  $\hat{p}_P = 100(2.858 - 1) = 185.80$ . Then, the PPG-ESIM was a better predictor of the net genetic merit and its efficiency was 185.80% higher than that of the PPG-LPSI for this data set.



Now, we compare PPG-ESIM efficiency versus PPG-LPSI efficiency using the data set described in Sect. 2.8.1 of Chap. 2 for five phenotypic selection cycles, each with four traits ( $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ ), 500 genotypes, and four replicates for 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, matrix  $\mathbf{F}$  was an identity matrix of size  $4 \times 4$  for all five selection cycles.

The first and second parts of columns 6, 7, and 8 in Table 2.1 present the estimated PPG-LPSI and PPG-ESIM selection responses for one, two, and three predetermined restrictions for five simulated selection cycles. The selection intensity was 10% ( $k_I = 1.755$ ) and the vectors of PPG for each predetermined restriction were  $\mathbf{d}'_1 = 7$ ,  $\mathbf{d}'_2 = [7 \ -3]$ , and  $\mathbf{d}'_3 = [7 \ -3 \ 5]$  respectively, for all five selection cycles. The estimated PPG-LPSI selection response when the vector of coefficients was not normalized was presented in Chap. 3 (Table 3.5). The averages of the estimated PPG-LPSI selection response for each predetermined restriction were 4.70, 4.91, and 3.14, whereas the averages of the estimated PPG-ESIM selection response were 6.31, 6.28, and 6.75 respectively. These results indicate that the estimated PPG-ESIM selection response was greater than the estimated PPG-LPSI selection response for all predetermined restrictions.

The second part of Table 2.2 presents the estimated PPG-ESIM accuracy ( $\hat{\rho}_P$ ) and the ratio of  $\hat{\rho}_P$  to the estimated PPG-LPSI accuracy ( $\hat{\rho}$ ), expressed in percentage terms,  $\hat{\rho}_P = 100(\hat{\lambda}_P - 1)$ , where  $\hat{\lambda}_P = \hat{\rho}_P/\hat{\rho}$ , for one, two, and three predetermined restrictions for five simulated selection cycles. The estimated PPG-LPSI accuracies were presented in Chapter. The average estimated PPG-ESIM efficiency for each restriction was 9.76%, 11.71%, and 29.03% greater than the PPG-LPSI efficiency for this data set in all five selection cycles.

The second part of Table 2.3 presents the estimated PPG-ESIM expected genetic gain per trait for one, two, and three predetermined restrictions for five simulated selection cycles. The estimated PPG-LPSI expected genetic gains per trait for one, two, and three predetermined restrictions were presented in Chapter.

where it can be seen that the averages of the estimated PPG-LPSI expected genetic gains per trait for five simulated selection cycles were 6.85, -3.25, 2.62 and 1.48 for one restriction; 6.93, -2.97, 2.65 and 1.45 for two restrictions; and 5.20, -2.23, 3.72 and 1.43 for three restrictions, whereas for the same set of restrictions, the averages of the estimated PPG-ESIM expected genetic gain per trait were 5.67, -2.67, 1.81, and 2.97 for one restriction; 5.89, -2.52, 2.04, and 2.83 for two restrictions; and 5.71, -2.45, 4.08, and 0.82 for three restrictions (Table 2.3). Because the vectors of predetermined proportional gains for each predetermined restriction were  $\mathbf{d}'_1 = 7$ ,  $\mathbf{d}'_2 = [7 \ -3]$ , and  $\mathbf{d}'_3 = [7 \ -3 \ 5]$ , the averages of the estimated PPG-LPSI expected genetic gains per trait were closer than those of the estimated PPG-ESIM expected genetic gains per trait for one and two predetermined restrictions, whereas for three restrictions, the results of both selection indices were similar.



## References

- Akbar MK, Lin CY, Gyles NR, Gavora JS, Brown CJ (1984) Some aspects of selection indices with constraints. *Poul Sci* 63:1899–1905
- Anderson TW (1999) Asymptotic theory for canonical correlation analysis. *J Multivar Anal* 70:1–29
- Anderson TW (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, New Jersey
- Bilodeau M, Brenner D (1999) Theory of multivariate statistics. Springer, New York
- Cerón-Rojas JJ, Crossa J, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A (2006) A selection index method based on eigen analysis. *Crop Sci* 46:1711–1721
- Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Crossa J (2008) A restricted selection index method based on eigenanalysis. *J Agric Biol Environ Stat* 13 (4):421–438
- Cerón-Rojas JJ, Crossa J, Toledo FH, Sahagún-Castellanos J (2016) A predetermined proportional gains eigen selection index method. *Crop Sci* 56:2436–2447
- Gentle JE (2007) Matrix algebra theory, computations, and applications in statistics. Springer, New York
- Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, New York
- Hotelling H (1935) The most predictable criterion. *J Educ Psychol* 26:139–142
- Hotelling H (1936) Relations between two sets of variables. *Biometrika* 28:321–377
- Kempthorne O, Nordskog AW (1959) Restricted selection indices. *Biometrics* 15:10–19
- Mallard J (1972) The theory and computation of selection indices with constraints: a critical synthesis. *Biometrics* 28:713–735
- Mardia KV, Kent JT, Bibby JM (1982) Multivariate analysis. Academic, New York
- Muirhead RJ (2005) Aspects of multivariate statistical theory. Wiley, Hoboken
- Okamoto M (1973) Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann Stat* 1(4):763–765
- Rao CR (2002) Linear statistical inference and its applications, 2nd edn. Wiley, New York
- Watkins DS (2002) Fundamentals of matrix computations, 2nd edn. Wiley, New York
- Wilms I, Croux C (2016) Robust sparse canonical correlation analysis. *BMC Syst Biol* 10:72

## The Molecular Eigen Selection Index Method: A Conceptual Approach



**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.

### 3.1 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. 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 Chapter 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.

### 3.1.1 The MESIM Parameters

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

$$H = \mathbf{w}'_1 \mathbf{g} + \mathbf{w}'_2 \mathbf{s} = [\mathbf{w}'_1 \ \mathbf{w}'_2] \begin{bmatrix} \mathbf{g} \\ \mathbf{s} \end{bmatrix} = \mathbf{w}' \mathbf{a}, \quad (3.1)$$

where  $\mathbf{g}' = [g_1 \ \dots \ g_t]$  is the vector of true breeding values,  $t$  is the number of traits,  $\mathbf{w}'_1 = [w_1 \ \dots \ w_t]$  is a vector of unknown economic weights associated with  $\mathbf{g}$ ,  $\mathbf{w}'_2 = [0_1 \ \dots \ 0_t]$  is a null vector associated with the vector of marker score values  $\mathbf{s}' = [s_1 \ s_2 \ \dots \ s_t]$ ,  $\mathbf{w}' = [\mathbf{w}'_1 \ \mathbf{w}'_2]$  and  $\mathbf{a}' = [\mathbf{g}' \ \mathbf{s}']$

The MESIM index can be written as

$$I = \boldsymbol{\beta}'_{\mathbf{y}} \mathbf{y} + \boldsymbol{\beta}'_{\mathbf{s}} \mathbf{s} = [\boldsymbol{\beta}'_{\mathbf{y}} \ \boldsymbol{\beta}'_{\mathbf{s}}] \begin{bmatrix} \mathbf{y} \\ \mathbf{s} \end{bmatrix} = \boldsymbol{\beta}' \mathbf{t}, \quad (3.2)$$

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

Let  $Var(H) = \mathbf{w}' \Psi_M \mathbf{w} = \sigma_H^2$  be the variance of  $H$ ,  $Var(I) = \boldsymbol{\beta}' \mathbf{T}_M \boldsymbol{\beta} = \sigma_I^2$  the variance of  $I$ , and  $Cov(H, I) = \mathbf{w}' \Psi_M \boldsymbol{\beta}$  the covariance between  $H$  and  $I$ , where  $\Psi_M = Var \begin{bmatrix} \mathbf{g} \\ \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{S}_M \\ \mathbf{S}_M & \mathbf{S}_M \end{bmatrix}$  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  $\rho_{HI} = \frac{\mathbf{w}' \Psi_M \boldsymbol{\beta}}{\sqrt{\mathbf{w}' \Psi_M \mathbf{w}} \sqrt{\boldsymbol{\beta}' \mathbf{T}_M \boldsymbol{\beta}}}$  and  $h_I^2 = \frac{\boldsymbol{\beta}' \Psi_M \boldsymbol{\beta}}{\boldsymbol{\beta}' \mathbf{T}_M \boldsymbol{\beta}}$  be the correlation between  $H$  and  $I$ , and the heritability of  $I$  respectively; then, the MESIM selection response can be written as

$$R = k_I \sigma_H \rho_{HI} \quad (3.3)$$

and

$$R = k_I \sigma_I h_I^2, \quad (3.4)$$

where  $k_I$  is the standardized selection differential (or selection intensity) associated with MESIM;  $\sigma_H = \sqrt{\mathbf{w}' \Psi_M \mathbf{w}}$  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  $\boldsymbol{\Psi}_M$  are known; therefore, we can maximize  $R$  by maximizing  $\rho_{HI}$  (Eq. 3.3) with respect to vectors  $\mathbf{w}$  and  $\boldsymbol{\beta}$ , or by maximizing  $h_I^2$  (Eq. 3.4) only with respect to vector  $\boldsymbol{\beta}$ .

Maximizing  $h_I^2$  only with respect to  $\boldsymbol{\beta}$  is simpler than maximizing  $\rho_{HI}$  with respect to  $\mathbf{w}$  and  $\boldsymbol{\beta}$ ; however, in the latter case the maximization process of  $\rho_{HI}$  gives more information associated with MESIM parameters than when  $h_I^2$  is maximized only with respect to  $\boldsymbol{\beta}$  (see Chap. 2, Eq. 2.13, for details). In this subsection, we maximize  $\rho_{HI}$  with respect to vectors  $\mathbf{w}$  and  $\boldsymbol{\beta}$  similar to the ESIM in Chap. 2, Sect. 2.1.1. Thus, we omit the steps and details of the maximization process of  $\rho_{HI}$ .

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

$$f(\boldsymbol{\beta}, \mathbf{w}, \mu, \phi) = \mathbf{w}'\boldsymbol{\Psi}\boldsymbol{\beta} - 0.5\mu(\boldsymbol{\beta}'\mathbf{T}\boldsymbol{\beta} - \sigma_I^2) - 0.5\phi(\mathbf{w}'\boldsymbol{\Psi}\mathbf{w} - \sigma_H^2) \quad (3.5)$$

with respect to  $\boldsymbol{\beta}$ ,  $\mathbf{w}$ ,  $\mu$ , and  $\phi$ , where  $\mu$  and  $\phi$  are Lagrange multipliers. The derivatives of Eq. (3.5) with respect to  $\boldsymbol{\beta}$ ,  $\mathbf{w}$ ,  $\mu$ , and  $\phi$  are:

$$\boldsymbol{\Psi}\mathbf{w} - \mu\mathbf{T}\boldsymbol{\beta} = \mathbf{0}, \quad (3.6)$$

$$\boldsymbol{\Psi}\boldsymbol{\beta} - \phi\boldsymbol{\Psi}\mathbf{w} = \mathbf{0}, \quad (3.7)$$

$$\boldsymbol{\beta}'\mathbf{T}\boldsymbol{\beta} = \sigma_I^2 \quad \text{and} \quad \mathbf{w}'\boldsymbol{\Psi}\mathbf{w} = \sigma_H^2, \quad (3.8)$$

respectively, where Eq. (3.8) denotes the restrictions imposed for maximizing  $\rho_{HI}$ . It can be shown (see Chap. 2) that vector  $\mathbf{w}$  can be obtained as

$$\mathbf{w}_M = \boldsymbol{\Psi}_M^{-1}\mathbf{T}_M\boldsymbol{\beta} \quad (3.9)$$

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  $\rho_{H_M I} = \frac{\sqrt{\boldsymbol{\beta}'\mathbf{T}\boldsymbol{\beta}}}{\sqrt{\boldsymbol{\beta}'\mathbf{T}\boldsymbol{\Psi}^{-1}\mathbf{T}\boldsymbol{\beta}}}$  and the MESIM vector of coefficients ( $\boldsymbol{\beta}$ ) that maximizes  $\rho_{H_M I}$  can be obtained from equation

$$(\mathbf{T}^{-1}\boldsymbol{\Psi} - \lambda_M^2 \mathbf{I}_{2t})\boldsymbol{\beta}_M = \mathbf{0}, \quad (3.10)$$

where  $\mathbf{I}_{2t}$  is an identity matrix of size  $2t \times 2t$  ( $t$  is the number of traits), and  $\lambda_M^2$  and  $\boldsymbol{\beta}_M$  are the *eigenvalue* and *eigenvector* of matrix  $\mathbf{T}_M^{-1}\boldsymbol{\Psi}_M$ . 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  $\lambda_M^2$  ( $\lambda_M$ ) is the canonical correlation between  $H_M = \mathbf{w}'_M\mathbf{a}$  and  $I_M = \boldsymbol{\beta}'_M\mathbf{t}$ , and the optimized MESIM index can be written as  $I_M = \boldsymbol{\beta}'_M\mathbf{t}$ . Using a similar procedure to that described .

be show that vector  $\beta_M$  can be transformed into  $\beta_C = F\beta_M$ , where  $F$  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  $I_M = \beta'_M \mathbf{t}$ , or MESIM accuracy, is

$$\rho_{H_M I_M} = \frac{\sqrt{\beta'_M \mathbf{T}_M \beta_M}}{\sqrt{\beta'_M \mathbf{T}_M \Psi_M^{-1} \mathbf{T}_M \beta_M}} = \frac{\sigma_{I_M}}{\sigma_{H_M}}, \quad (3.11)$$

where  $\sigma_{I_M} = \sqrt{\beta'_M \mathbf{T}_M \beta_M}$  is the standard deviation of  $I_M = \beta'_M \mathbf{t}$ , and  $\sigma_{H_M} = \sqrt{\beta'_M \mathbf{T}_M \Psi_M^{-1} \mathbf{T}_M \beta_M}$  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

$$R_M = k_I \sqrt{\beta'_{M_1} \mathbf{T}_M \beta_{M_1}} \quad (3.12)$$

and

$$\mathbf{E}_M = k_I \frac{\Psi_M \beta_{M_1}}{\sqrt{\beta'_{M_1} \mathbf{T}_M \beta_{M_1}}}, \quad (3.13)$$

respectively, where  $\beta_{M_1}$  is the first eigenvector of matrix  $\mathbf{T}_M^{-1} \Psi_M$ . If vector  $\beta_{M_1}$  is multiplied by matrix  $F$ , we obtain  $\beta_{C_1} = F\beta_{M_1}$ ; in this case, we can replace  $\beta_{M_1}$  with  $\beta_{C_1} = F\beta_{M_1}$  in Eqs. (3.12) and (3.13), and the optimized MESIM index should be written as  $I_M = \beta'_{C_1} \mathbf{y}$ .

### 3.1.2 Estimating MESIM Parameters

We estimate the MESIM parameters using the same procedure described in Chap. 2 (Sect. 2.1.4) 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  $\widehat{\Psi}_M = \begin{bmatrix} \widehat{\mathbf{C}} & \widehat{\mathbf{S}}_M \\ \widehat{\mathbf{S}}_M & \widehat{\mathbf{S}}_M \end{bmatrix}$  the estimated block matrices (Chap. 2) and  $\widehat{\mathbf{W}} = \widehat{\mathbf{T}}_M^{-1} \widehat{\Psi}_M$ ; then, to find the estimators  $\widehat{\beta}_{M_1}$  and  $\widehat{\lambda}_{M_1}^2$  of the first eigenvector ( $\beta_{M_1}$ ) and the first eigenvalue ( $\lambda_{M_1}^2$ ) respectively, we need to solve the equation

$$(\widehat{\mathbf{W}} \widehat{\mathbf{W}}' - \widehat{\mu}_j \mathbf{I}) \widehat{\beta}_{M_j} = \mathbf{0}, \quad (3.14)$$

where  $\widehat{\mu}_j = \widehat{\lambda}_{M_1}^4$ ,  $j = 1, 2, \dots, 2t$ . For additional details, see Eqs. (3.22) and (3.23), and Sect. 2.1.5 of Chap. 2. The result of Equation (3.14) allow the MESIM index

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

$$\hat{R}_M = k_I \sqrt{\hat{\beta}'_{M_1} \hat{\mathbf{T}}_M \hat{\beta}_{M_1}} \text{ and } \hat{\mathbf{E}}_M = k_I \frac{\hat{\Psi}_M \hat{\beta}_{M_1}}{\sqrt{\hat{\beta}'_{M_1} \hat{\mathbf{T}}_M \hat{\beta}_{M_1}}}, \quad (3.15)$$

respectively.

### 3.1.3 Numerical Examples

To validate the MESIM theoretical results, we use a real maize (*Zea mays*) F<sub>2</sub> 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  $\hat{\mathbf{P}} = \begin{bmatrix} 191.81 & 106.89 \\ 106.89 & 167.93 \end{bmatrix}$ ,  $\hat{\mathbf{C}} = \begin{bmatrix} 83.00 & 57.44 \\ 57.44 & 59.80 \end{bmatrix}$ , and  $\hat{\mathbf{S}}_M = \begin{bmatrix} 15.750 & 0.983 \\ 0.983 & 28.083 \end{bmatrix}$  respectively, and the vector of economic weights was  $\mathbf{a}' = [\mathbf{w}' \ 0']$ , where  $\mathbf{w}' = [-1 \ -1]$  and  $0' = [0 \ 0]$ . Details of how to estimate the marker scores and their variance were given in Chap. .

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

Vector  $\hat{\beta}'_{M_1} = [0.089 \ -0.061 \ -0.536 \ 0.837]$  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  $\hat{\beta}'_{M_1}$  was transformed as  $\hat{\beta}'_{C_1} = \hat{\beta}'_{M_1} \mathbf{F} = [-0.009 \ 0.006 \ -0.402 \ 0.628]$  and then the estimated MESIM index was  $\hat{I}_M = -0.009 \text{ PHT} + 0.006 \text{ EHT} - 0.402 \text{ S}_{\text{PHT}} + 0.628 \text{ S}_{\text{EHT}}$ , where  $S_{\text{PHT}}$  and  $S_{\text{EHT}}$  denote the marker scores associated with PHT and EHT respectively. The estimated MESIM expected

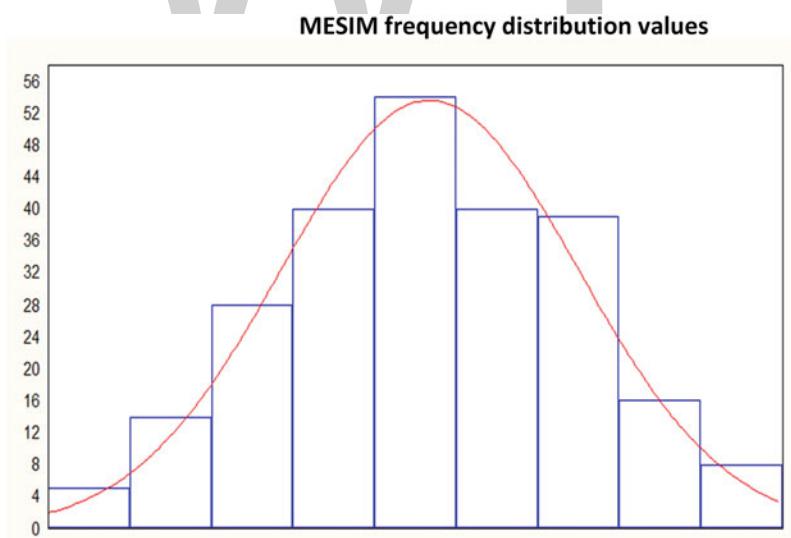


genetic gain, selection response, and accuracy were  $\widehat{\mathbf{E}}'_M = k_I \frac{\widehat{\boldsymbol{\beta}}'_{C_1} \widehat{\Psi}_M}{\sqrt{\widehat{\boldsymbol{\beta}}'_{C_1} \widehat{\mathbf{T}}_M \widehat{\boldsymbol{\beta}}_{C_1}}} = [-3.438 \quad -8.516 \quad -3.319 \quad -8.372]$ ,  $\widehat{R}_M = k_I \sqrt{\widehat{\boldsymbol{\beta}}'_{C_1} \widehat{\mathbf{T}}_M \widehat{\boldsymbol{\beta}}_{C_1}} = 6.573$  and  $\widehat{\rho}_{H_M I_M} = \frac{\widehat{\sigma}_{I_M}}{\widehat{\sigma}_{H_M}} = 0.99$  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. 3.1 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. 3.1** 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<sub>2</sub> population with 195 molecular markers and two traits, plant height (PHT, cm) and ear height (EHT, cm), and their associated marker scores S<sub>PHT</sub> and S<sub>EHT</sub> respectively



for four simulated selection cycles when their vectors of coefficients were normalized, are presented in Table 3.1. 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. 3.2) 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 3.2 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 3.1** 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 3.2** 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% ( $k_I = 1.755$ )

Cycle	Estimated LMSI expected genetic gain							
	Traits				Marker scores			
	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
Cycle	Estimated MESIM expected genetic gain							
	Traits				Marker scores			
	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.

## 3.2 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. 3.1 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.

### 3.2.1 The GESIM Parameters

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

$$H = \mathbf{w}'_1 \mathbf{g} + \mathbf{w}'_2 \boldsymbol{\gamma} = [\mathbf{w}'_1 \quad \mathbf{w}'_2] \begin{bmatrix} \mathbf{g} \\ \boldsymbol{\gamma} \end{bmatrix} = \mathbf{w}' \boldsymbol{\alpha}, \quad (3.16)$$

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

$$I = \boldsymbol{\beta}'_{\mathbf{y}} \mathbf{y} + \boldsymbol{\beta}'_{\boldsymbol{\gamma}} \boldsymbol{\gamma} = [\boldsymbol{\beta}'_{\mathbf{y}} \quad \boldsymbol{\beta}'_{\boldsymbol{\gamma}}] \begin{bmatrix} \mathbf{y} \\ \boldsymbol{\gamma} \end{bmatrix} = \boldsymbol{\beta}' \mathbf{f}, \quad (3.17)$$



where  $\mathbf{y}' = [y_1 \cdots y_t]$  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;  $\boldsymbol{\beta}' = [\boldsymbol{\beta}'_{\mathbf{y}} \quad \boldsymbol{\beta}'_{\boldsymbol{\gamma}}]$  and  $\mathbf{f}' = [\mathbf{y}' \quad \boldsymbol{\gamma}']$ .

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

$$R = k_I \sigma_H \rho_{HI}, \quad (3.19)$$

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  $\rho_{HI}$  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

$$\mathbf{w}_G = \mathbf{A}^{-1}\boldsymbol{\Phi}\boldsymbol{\beta} \quad (3.19)$$

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  $\rho_{HGI} = \frac{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta}}}{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}\boldsymbol{\beta}}}$  and the GESIM index vector of coefficients that maximizes  $\rho_{HGI}$  can be obtained from the equation

$$(\boldsymbol{\Phi}^{-1}\mathbf{A} - \lambda_G^2 \mathbf{I}_{2t})\boldsymbol{\beta}_G = \mathbf{0}, \quad (3.20)$$

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. (3.19) and (3.20), GESIM accuracy can be written as

$$\rho_{HGI_G} = \frac{\sigma_{I_G}}{\sigma_{H_G}}, \quad (3.21)$$

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. (3.20),  $\lambda_G^2 = \rho_{HGI_G}^2$  is the square of the canonical correlation between  $H_G$  and  $I_G$ , and  $\boldsymbol{\beta}_G$  is the canonical vector associated with  $\lambda_G^2 = \rho_{HGI_G}^2$ .

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

$$R_G = k_I \sqrt{\beta'_G \Phi \beta_G} \quad (3.22)$$

and

$$E_G = k_I \frac{A \beta_G}{\sqrt{\beta'_G \Phi \beta_G}}, \quad (3.23)$$

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

### 3.2.2 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<sup>-1</sup>), ear height (EHT, cm), and plant height (PHT, cm). We estimated matrices  $P$  and  $C$  using Eqs. (2.22) to (2.24) described in Chap. 2, whence the estimated matrices were

$$\hat{P} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix} \text{ and } \hat{C} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}.$$

In a similar manner, we estimated matrix  $\Gamma$  by applying Eqs. (5.21) to (5.23) described in Chap. 5 using phenotypic and marker information jointly; the estimated matrix

$$\text{was } \hat{\Gamma} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}.$$

The selection intensity for making a selection cycle was 10% ( $k_I = 1.755$ ) and the vector of economic weights was  $w' = [5 \ -0.1 \ -0.1 \ 0 \ 0 \ 0]$ . To obtain the estimated vector of coefficient of CLGSI ( $\hat{\beta} = \hat{\Phi}^{-1} \hat{A} w$ ) and GESIM (Eq. 3.20), it is necessary to construct matrices

$$\hat{A} = \begin{bmatrix} \hat{C} & \hat{\Gamma} \\ \hat{\Gamma} & \hat{\Gamma} \end{bmatrix} \text{ and } \hat{\Phi} = \begin{bmatrix} \hat{P} & \hat{\Gamma} \\ \hat{\Gamma} & \hat{\Gamma} \end{bmatrix}.$$

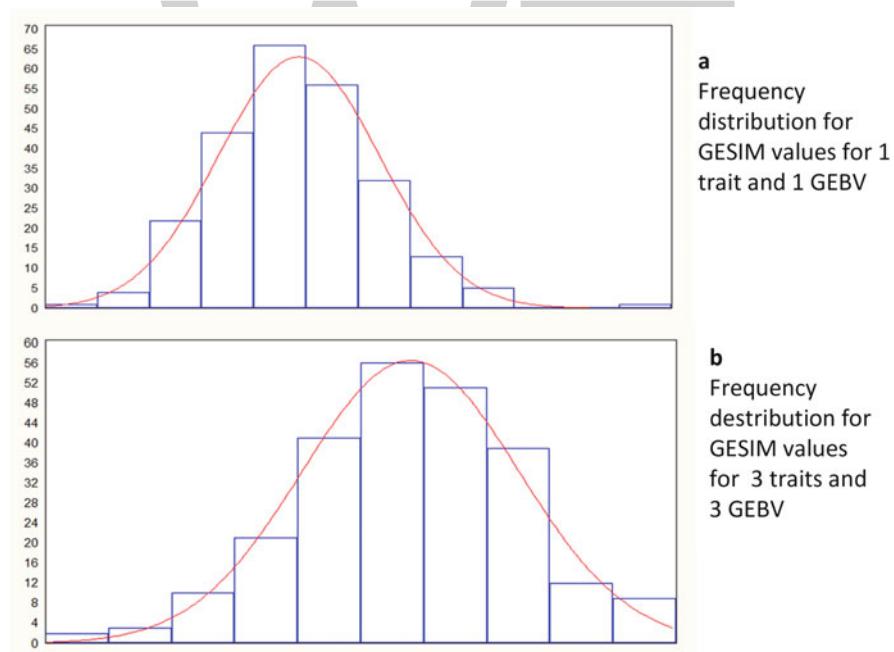
The estimated CLGSI vector of coefficients for the traits GY, EHT, and PHT and their associated GEBVs (GEBV<sub>GY</sub>, GEBV<sub>EHT</sub>, and GEBV<sub>PHT</sub> respectively) was  $\hat{\beta}' = [0.08 \ -0.02 \ -0.01 \ 4.92 \ -0.08 \ -0.09]$ , whereas the estimated CLGSI selection response, accuracy, and expected genetic gain per trait were  $\hat{R} = k_I \sqrt{\hat{\beta}' \hat{\Phi} \hat{\beta}} = 1.54$ ,  $\hat{\rho}_{HI} = \frac{\hat{\sigma}_I}{\hat{\sigma}_H} = 0.814$ , and  $\hat{E}' = k_I \frac{\hat{\beta}' \hat{A}}{\sqrt{\hat{\beta}' \hat{\Phi} \hat{\beta}}} =$

$[0.36 \ 1.04 \ 1.70 \ 0.36 \ 1.53 \ 2.38]$  respectively. Finally,  $\hat{I} = 0.08GY - 0.02EHT - 0.01PHT + 4.92GEBV_{GY} - 0.08GEBV_{EHT} - 0.09GEBV_{PHT}$  was the estimated CLGSI.

The estimated GESIM vector of coefficients, selection response, accuracy, and expected genetic gain per trait were  $\hat{\beta}'_{G_1} = [-0.207 \ 0.029 \ 0.041 \ 0.820 \ 0.337 \ 0.411]$ ,  $\hat{R}_G = k_1 \sqrt{\hat{\beta}'_{G_1} \hat{\Phi} \hat{\beta}_{G_1}} = 6.288$ ,  $\hat{\rho}_{HGI_G} = \frac{\sqrt{\hat{\beta}'_{G_1} \hat{\Phi} \hat{\beta}_{G_1}}}{\sqrt{\hat{\beta}'_{G_1} \hat{\Phi} \hat{\mathbf{A}}^{-1} \hat{\Phi} \hat{\beta}_{G_1}}} = 0.9056$ , and  $\hat{\mathbf{E}}'_G = k_1 \frac{\hat{\beta}'_{G_1} \hat{\mathbf{A}}}{\sqrt{\hat{\beta}'_{G_1} \hat{\Phi} \hat{\beta}_{G_1}}} = [0.369 \ 5.528 \ 9.186 \ 0.370 \ 5.250 \ 8.702]$  respectively.

Fig. 3.2 presents the frequency distribution of the 244 estimated GESIM index values for one (Fig. 3.2a) and three traits (Fig. 3.2b) 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. 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. 3.2** 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 3.3** 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% ( $k_I = 1.755$ )

Cycle	Estimated CLGSI expected genetic gain							
	Traits				Genomic estimated breeding value			
	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
Cycle	Estimated GESIM expected genetic gain							
	Traits				Genomic estimated breeding value			
	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  $\mathbf{F}$  was an identity matrix of size  $8 \times 8$  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 3.1) respectively. The estimated CLGSI selection response was lower than the estimated GESIM selection response. Table 3.3 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 2.18, -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.

### 3.3 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 (Chapter. ), the GW-ESIM vector of coefficients is estimated according to the singular value decomposition (SVD) described in Chap. 2.

### 3.3.1 The GW-ESIM Parameters

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

$$H = \mathbf{w}'_1 \mathbf{g} + \mathbf{w}'_2 \mathbf{m} = [\mathbf{w}'_1 \quad \mathbf{w}'_2] \begin{bmatrix} \mathbf{g} \\ \mathbf{m} \end{bmatrix} = \mathbf{w}' \mathbf{x}, \quad (3.24)$$

where  $\mathbf{g}' = [g_1 \quad \dots \quad g_t]$  is the vector of true breeding values,  $t$  is the number of traits,  $\mathbf{w}'_1 = [w_1 \quad \dots \quad w_t]$  is the vector of unknown economic weights associated with the breeding values;  $\mathbf{w}'_2 = [0_1 \quad \dots \quad 0_N]$  is a null vector associated with the vector of marker code values  $\mathbf{m}' = [m_1 \quad \dots \quad m_N]$ , where  $m_j$  ( $j = 1, 2, \dots, N$  = number of markers) is the  $j$ th marker in the training population;  $\mathbf{w}' = [\mathbf{w}'_1 \quad \mathbf{w}'_2]$  and  $\mathbf{x} = [\mathbf{g}' \quad \mathbf{m}']$ . The GW-ESIM ( $I$ ) index combines the phenotypic value and all the marker information of individuals to predict Eq. (3.24) values in each selection cycle and can be written as

$$I = \boldsymbol{\beta}'_y \mathbf{y} + \boldsymbol{\beta}'_m \mathbf{m} = [\boldsymbol{\beta}'_y \quad \boldsymbol{\beta}'_m] \begin{bmatrix} \mathbf{y} \\ \mathbf{m} \end{bmatrix} = \boldsymbol{\beta}' \mathbf{q}, \quad (3.25)$$

where  $\boldsymbol{\beta}'_y$  and  $\boldsymbol{\beta}_m$  are vectors of phenotypic and marker weights respectively;  $\mathbf{y}' = [y_1 \quad \dots \quad y_t]$  is the vector of phenotypic values;  $\mathbf{m}$  was defined in Eq. (3.24);  $\boldsymbol{\beta}' = [\boldsymbol{\beta}'_y \quad \boldsymbol{\beta}'_m]$  and  $\mathbf{q}' = [\mathbf{y}' \quad \mathbf{m}']$ .

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  $\mathbf{Q} = \text{Var} \begin{bmatrix} \mathbf{y} \\ \mathbf{m} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{G}'_M \\ \mathbf{G}_M & \mathbf{M} \end{bmatrix}$  and  $\mathbf{X} = \text{Var} \begin{bmatrix} \mathbf{g} \\ \mathbf{m} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{G}'_M \\ \mathbf{G}_M & \mathbf{M} \end{bmatrix}$  are block matrices of size  $(t + N) \times (t + N)$  ( $t$  is the number of traits and  $N$  is the number of markers) where  $\mathbf{P} = \text{Var}(\mathbf{y})$ ,  $\mathbf{M} = \text{Var}(\mathbf{m})$ ,  $\mathbf{C} = \text{Var}(\mathbf{g})$ , and  $\mathbf{G}_M = \text{cov}(\mathbf{y}, \mathbf{m}) = \text{cov}(\mathbf{g}, \mathbf{m})$  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  $\mathbf{m}$ ;  $\mathbf{w}$  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.

we described matrix  $\mathbf{M}$  as

$$\mathbf{M} = \begin{bmatrix} 1 & (1 - 2\theta_{11}) & \dots & (1 - 2\theta_{1N}) \\ (1 - 2\theta_{21}) & 1 & \dots & (1 - 2\theta_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ (1 - 2\theta_{N1}) & (1 - 2\theta_{N2}) & \dots & 1 \end{bmatrix}, \quad (3.26)$$

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

$$\mathbf{G}_M = \begin{bmatrix} (1 - 2r_{11})\alpha_{11} & (1 - 2r_{11})\alpha_{12} & \dots & (1 - 2r_{1N_Q})\alpha_{1N_Q} \\ (1 - 2r_{21})\alpha_{21} & (1 - 2r_{22})\alpha_{22} & \dots & (1 - 2r_{2N_Q})\alpha_{2N_Q} \\ \vdots & \vdots & \ddots & \vdots \\ (1 - 2r_{t1})\alpha_{t1} & (1 - 2r_{t2})\alpha_{t2} & \dots & (1 - 2r_{tN_Q})\alpha_{tN_Q} \end{bmatrix}, \quad (3.27)$$

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  $q$ th trait and the  $i$ th marker;  $r_{ik}$  is the recombination frequency between the  $i$ th and  $k$ th QTL, and  $\alpha_{qk}$  is the effect of the  $k$ th QTL over the  $q$ th trait.

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

$$R = k_I \sigma_H \rho_{HI}, \quad (3.28)$$

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. 3.28) by maximizing  $\rho_{HI}$  with respect to vectors  $\mathbf{w}'$  and  $\beta$  under the restrictions  $\sigma_H^2 = \mathbf{w}'\mathbf{X}\mathbf{w}$ ,  $\sigma_I^2 = \beta'\mathbf{Q}\beta$ , 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

$$\mathbf{w}_W = \mathbf{X}^{-1}\mathbf{Q}\beta \quad (3.29)$$

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  $I = \beta'\mathbf{q}$  is  $\rho_{H_W I} = \frac{\sqrt{\beta'\mathbf{Q}\beta}}{\sqrt{\beta'\mathbf{Q}\mathbf{X}^{-1}\mathbf{Q}\beta}}$  and the GW-ESIM vector of coefficients ( $\beta$ ) that maximizes  $\rho_{H_W I}$  can be obtained from equation

$$(\mathbf{Q}^{-1}\mathbf{Z} - \lambda_W^2 \mathbf{I}_{(t+N)})\beta_W = \mathbf{0}, \quad (3.30)$$



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

$$\rho_{H_W I_W} = \frac{\sqrt{\beta'_W \mathbf{Q} \beta_W}}{\sqrt{\beta'_W \mathbf{Q} \mathbf{X}^{-1} \mathbf{Q} \beta_W}} = \frac{\sigma_{I_W}}{\sigma_{H_W}}, \quad (3.31)$$

where  $\sigma_{I_W} = \sqrt{\beta'_W \mathbf{Q} \beta_W}$  is the standard deviation of  $I_W = \beta'_W \mathbf{q}$ , and  $\sigma_{H_W} = \sqrt{\beta'_W \mathbf{Q} \mathbf{X}^{-1} \mathbf{Q} \beta_W}$  is the standard deviation of  $H_W = \mathbf{w}'_W \mathbf{x}$ . In Eq. (3.30)  $\lambda_W^2 = \rho_{H_W I_W}^2$  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

$$R_W = k_I \sqrt{\beta'_W \mathbf{Q} \beta_W} \quad (3.32)$$

and

$$\mathbf{E}_W = k_I \frac{\mathbf{X} \beta_W}{\sqrt{\beta'_W \mathbf{Q} \beta_W}}, \quad (3.33)$$

respectively, where  $\beta_W$  is the first eigenvector of Eq. (3.30).

### 3.3.2 Estimating GW-ESIM Parameters

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

$$(\hat{\mathbf{Q}}^{-1} \hat{\mathbf{X}} - \hat{\lambda}_{Wj}^2 \mathbf{I}) \hat{\beta}_{Wj} = \mathbf{0}, \quad (3.34)$$

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

$$(\hat{\mathbf{E}} \hat{\mathbf{E}}' - \hat{\mu}_j \mathbf{I}) \hat{\beta}_{W_j} = \mathbf{0}, \quad (3.35)$$

where  $\widehat{\mathbf{E}} = \widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{X}}$  and  $\widehat{\mu}_j = \widehat{\lambda}_{W_j}^4$ . These results allow the GW-ESIM index selection response and its expected genetic gain per trait to be estimated as  $\widehat{I}_W = \widehat{\beta}'_{W_1}\widehat{\mathbf{q}}$ ,  $\widehat{R}_W = k_I \sqrt{\widehat{\beta}'_{W_1}\widehat{\mathbf{Q}}\widehat{\beta}'_{W_1}}$  and  $\widehat{\mathbf{E}}_w = k_I \frac{\widehat{\mathbf{X}}\widehat{\beta}'_{W_1}}{\sqrt{\widehat{\beta}'_{W_1}\widehat{\mathbf{Q}}\widehat{\beta}'_{W_1}}}$  respectively, whereas the estimator of GW-ESIM accuracy is  $\widehat{\lambda}_{W_1}$ .

### 3.3.3 Numerical Examples

We compare the estimated GW-LMSI and GW-ESIM selection responses using the simulated data set described in with a selection intensity of 10% ( $k_I = 1.755$ ). Table 3.1 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 3.1 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.

## 3.4 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. 2. 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. 2 for details). This is the main difference between the RGESIM and the RESIM.

It can be shown that  $Cov(I, \alpha) = A\beta$  is the covariance between the breeding value vector ( $\alpha' = [\mathbf{g}' \quad \gamma']$ ) and the GESIM index ( $I = \beta'\mathbf{f}$ ). In the RGESIM, we want some covariances between the linear combinations of  $\alpha$  ( $\mathbf{U}'_G\alpha$ ) and  $I = \beta'\mathbf{f}$  to be zero, i.e.,  $Cov(I_G, \mathbf{U}'_G\alpha) = \mathbf{U}'_G A \beta = \mathbf{0}$ , 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  $\frac{\beta'A\beta}{\sqrt{\beta'\Phi\beta}}$  with respect to vector  $\beta$  under the restriction  $\mathbf{U}'_G A \beta = \mathbf{0}$  and  $\beta'\beta = 1$  similar to the RESIM, or by maximizing the correlation between  $H = \mathbf{w}'\alpha$  and



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

### 3.4.1 The RGESIM Parameters

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

$$f(\beta, v') = \frac{\beta' A \beta}{\sqrt{\beta' \Phi \beta}} - v' U'_G A \beta \quad (3.36)$$

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

$$2(\beta' \Phi \beta)^{1/2} A \beta - (\beta' \Phi \beta)^{-1/2} (\beta' A \beta) \Phi \beta - A U_G v = \mathbf{0}, \quad (3.37)$$

$$U'_G A \beta = \mathbf{0}, \quad (3.38)$$

respectively, where Eq. (3.38) denotes the restriction imposed for maximizing Eq. (3.36). Using algebraic methods on Eq. (3.37), we get

$$(K_{RG} \Phi^{-1} A - \lambda_{RG}^2 I_{2t}) \beta_{RG} = \mathbf{0}, \quad (3.39)$$

where  $\lambda_{RG}^2 = h_{I_{RG}}^2$ ,  $h_{I_{RG}}^2$  is the RGESIM heritability obtained under the restriction  $U'_G A \beta = \mathbf{0}$ ;  $K_{RG} = [I_{2t} - Q_{RG}]$ ,  $I_{2t}$  is an identity matrix of size  $2t \times 2t$ , and  $Q_{RG} = \Phi^{-1} A U_G (U'_G A \Phi^{-1} A U_G)^{-1} U'_G A$ . When  $U'_G$  is a null matrix,  $\beta'_{RG} = \beta'_{RG}$  (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  $I_{RG} = \beta'_{RG} y$  and its selection response and expected genetic gain per trait use the first eigenvector of matrix  $K_{RG} \Phi^{-1} A$ . It can be shown that the vector of coefficients of  $H = w'_{RG} \alpha$  in the RGESIM can be written as

$$w_{RG} = A^{-1} [\Phi + Q'_{RG} A] \beta_{RG}, \quad (3.40)$$

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

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

$$\sigma_{H_{RG} I_{RG}} = w'_{RG} A \beta'_{RG} = \beta'_{RG} \Phi \beta_{RG} + \beta'_{RG} Q'_{RG} C \beta_{RG} = \beta'_{RG} \Phi \beta_{RG}. \quad (3.41)$$

Equation (3.41) indicates that  $\sigma_{H_{RG}I_{RG}}$  is equal to the variance of  $I_{RG} = \beta'_{RG}\mathbf{f}$  ( $\sigma_{I_{RG}}^2 = \beta'_{RG}\Phi\beta_{RG}$ ); therefore, the maximized correlation between  $I_{RG}$  and  $H_{RG}$  or RGESIM accuracy can be written as

$$\rho_{H_{RG}I_{RG}} = \frac{\sqrt{\beta'_{RG}\Phi\beta_{RG}}}{\sqrt{\mathbf{w}'_{RG}\mathbf{A}\mathbf{w}_{RG}}}, \quad (3.42)$$

where  $\mathbf{w}'_{RG}\mathbf{A}\mathbf{w}_{RG}$  is the variance of  $H_{RG}$ . Hereafter, to simplify the notation, we write Eq. (3.42) as  $\lambda_{RG}$ .

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

$$R_{RG} = k_I \sqrt{\beta'_{RG}\Phi\beta_{RG}} \quad (3.43)$$

and

$$\mathbf{E}_{RG} = k_I \frac{\mathbf{A}\beta_{RG}}{\sqrt{\beta'_{RG}\Phi\beta_{RG}}}, \quad (3.44)$$

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

### 3.4.2 Estimating RGESIM Parameters

In Sect. 3.2, we indicated how to estimate matrices  $\mathbf{P}$ ,  $\mathbf{\Gamma}$ , and  $\mathbf{C}$  using phenotypic and genomic information, whence we can estimate matrices  $\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{\Gamma} \\ \mathbf{\Gamma} & \mathbf{\Gamma} \end{bmatrix}$  and  $\Phi = \begin{bmatrix} \mathbf{P} & \mathbf{\Gamma} \\ \mathbf{\Gamma} & \mathbf{\Gamma} \end{bmatrix}$ . 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.

### 3.4.3 Numerical Examples

With a selection intensity of 10% ( $k_I = 1.755$ ), we compare the CRLGSI (for details see Chap. 1) 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. 3.2.2, 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  $\widehat{\Gamma} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}$  were the estimated matrices of  $\mathbf{P}$ ,  $\mathbf{C}$ , and  $\Gamma$  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 ( $\text{ton ha}^{-1}$ ) and its associated GEBV<sub>GY</sub>; 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<sub>GY</sub> and GEBV<sub>EHT</sub>, 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  $\widehat{\mathbf{K}}_{RG} = [\mathbf{L}_{2t} - \widehat{\mathbf{Q}}_{RG}]$  and  $\widehat{\mathbf{Q}}_{RG} = \widehat{\Phi}^{-1}\widehat{\mathbf{A}}\mathbf{U}_G(\mathbf{U}'_G\widehat{\mathbf{A}}\widehat{\Phi}^{-1}\widehat{\mathbf{A}}\mathbf{U}_G)^{-1}\mathbf{U}'_G$  was described in Chap. 1, and is also valid for estimating RGESIM parameters.

The estimated CRLGSI vector of coefficients is  $\widehat{\beta}_{CR} = \widehat{\mathbf{K}}_{RG}\widehat{\beta}$ , where  $\widehat{\beta} = \widehat{\Phi}^{-1}\widehat{\mathbf{A}}$  is the estimated CLGSI vector of coefficients (Chap. 1). Let  $\mathbf{w}' = [5 \quad -0.1 \quad -0.1 \quad 0 \quad 0 \quad 0]$  be the vector of economic weights and suppose that we restrict trait GY and its associated GEBV<sub>GY</sub>; 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  $\widehat{\Gamma}$  described earlier,  $\widehat{\beta}'_{CR} = [0.076 \quad -0.004 \quad -0.018 \quad 2.353 \quad -0.096 \quad -0.082]$  was the estimated CRLGSI vector of coefficients and the estimated CRLGSI was

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

where GEBV<sub>GY</sub>, GEBV<sub>EHT</sub>, and GEBV<sub>PHT</sub> 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  $\widehat{R}_{CR} = k_I \sqrt{\widehat{\beta}'_{CR}\widehat{\Phi}\widehat{\beta}_{CR}} = 0.96$  and  $\widehat{E}'_{CR} = k_I = \frac{\widehat{\beta}'_{CR}\widehat{\mathbf{A}}}{\sqrt{\widehat{\beta}'_{CR}\widehat{\Phi}\widehat{\beta}_{CR}}} = [0 \quad -3.53 \quad -6.03 \quad 0 \quad -2.93 \quad -4.87]$  respectively, whereas the estimated CRLGSI accuracy was  $\widehat{\rho}_{Hl_{CR}} = \frac{\widehat{\sigma}_{I_{CR}}}{\widehat{\sigma}_H} = 0.51$ . Note that in  $\widehat{\mathbf{E}}'_{CR}$ , the trait GY and its associated GEBV<sub>GY</sub> have null values, as we would expect.

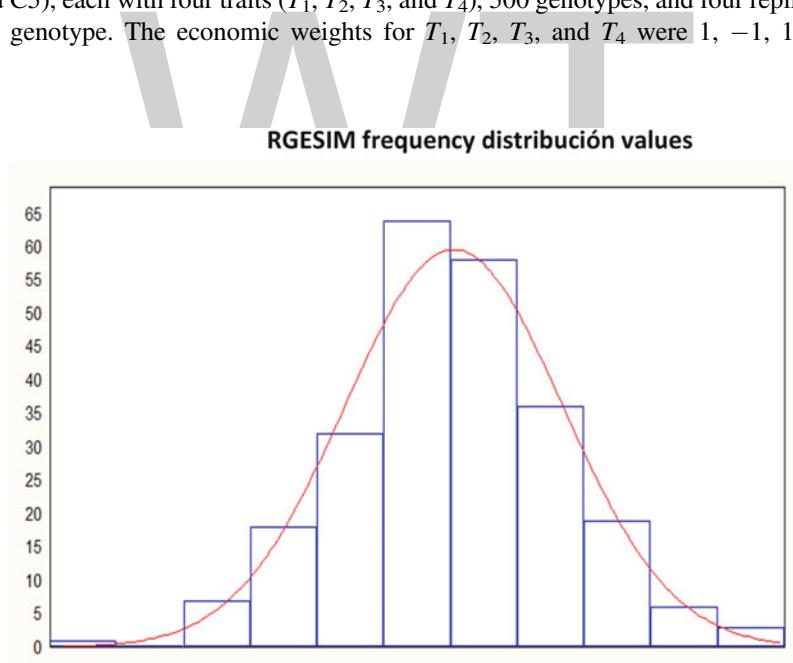
The estimated RGESIM vector of coefficients was  $\widehat{\beta}'_{CR} = [0.015 \quad -0.001 \quad -0.004 \quad 0.998 \quad -0.029 \quad -0.045]$ , and the estimated RGESIM index was  $\widehat{I}_{RG} = 0.015\text{GY} - 0.001\text{EHT} - 0.004\text{PHT} + 0.998\text{GEBV}_{GY} - 0.029\text{GEBV}_{EHT}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

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  $\hat{R}_{RG} = k_I \sqrt{\hat{\beta}'_{RG} \hat{\Phi} \hat{\beta}_{RG}} = 0.37$  and  $\hat{E}'_{RG} = k_I = \frac{\hat{\beta}'_{RG} \mathbf{A}}{\sqrt{\hat{\beta}'_{RG} \hat{\Phi} \hat{\beta}_{RG}}}$  [0 -3.28 -6.03 0 -2.93 -5.40] respectively, whereas the estimated RGESIM accuracy was  $\hat{\rho}_{H_{RG} I_{RG}} = \frac{\hat{\sigma}_{I_{RG}}}{\hat{\sigma}_{H_{RG}}} = 0.86$ .

Fig. 3.3 presents the frequency distribution of the 244 estimated RGESIM index values for two null restrictions on traits GY and EHT and their associated GEBV<sub>GY</sub> and GEBV<sub>EHT</sub>, for one selection cycle in an environment for a real maize (*Zea mays*) F<sub>2</sub> 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. 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



**Fig. 3.3** 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<sub>GY</sub> and GEBV<sub>EHT</sub> respectively, for one selection cycle in an environment for a real maize (*Zea mays*) F<sub>2</sub> 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  $8 \times 8$  for all four selection cycles.

Columns 2, 3, and 4 (from left to right) of Table 3.4 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 3.5 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 3.5 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 3.4** 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

Cycle	CRLGSI response for one, two and three null restrictions			CPPG-LGSI response for one, two and three predetermined restrictions		
	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
Cycle	RGESIM response for one, two, and three null restrictions			PPG-GESIM response for one, two, and three predetermined restrictions		
	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% ( $k_f = 1.755$ )



**Table 3.5** 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% ( $k_I = 1.755$ )

Cycle	CRLGSI expected genetic gains for one, two and three null restrictions											
	One restriction <sup>a</sup>						Two restrictions <sup>b</sup>				Three restrictions <sup>c</sup>	
	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												
Cycle	One restriction <sup>a</sup>						Two restrictions <sup>b</sup>				Three restrictions <sup>c</sup>	
	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

<sup>a</sup>All T1 and G1 expected genetic gains were null

<sup>b</sup>All T1, T2, G1, and G2 expected genetic gains were null

<sup>c</sup>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.

### 3.5 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  $\mathbf{d}' = [d_1 \ d_2 \ \dots \ d_r]$ . 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,  $\mathbf{d}'_{PG} = [d_1 \ d_2 \ \cdots \ d_r \ d_{r+1} \ d_{r+2} \ \cdots \ d_{2r}]$ , 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. 2 and the PPG-GESIM.

### 3.5.1 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

$$(\mathbf{T}_{PG} - \lambda_{PG}^2 \mathbf{I}_{2t}) \boldsymbol{\beta}_{PG} = \mathbf{0}, \quad (3.45)$$

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

When  $\mathbf{B}$  is a null matrix,  $\mathbf{T}_{PG} = \mathbf{K}_{RG}\Phi^{-1}\mathbf{A}$  (matrix of the RGESIM), and when  $\mathbf{U}'_G$  is a null matrix,  $\mathbf{T}_{PG} = \Phi^{-1}\mathbf{A}$  (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

$$\mathbf{w}_{PG} = \mathbf{A}^{-1}[\Phi + \mathbf{Q}'_{PG}\mathbf{A}] \boldsymbol{\beta}_{PG}, \quad (3.46)$$

where  $\mathbf{Q}'_{PG} = \mathbf{A}\mathbf{U}_G\mathbf{D}_G(\mathbf{D}'_G\mathbf{U}'_G\mathbf{A}\Phi^{-1}\mathbf{A}\mathbf{U}_G\mathbf{D}_G)^{-1}\mathbf{D}'_G\mathbf{U}'_G\mathbf{A}\Phi^{-1}$ , 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  $I_{PG} = \boldsymbol{\beta}'_{PG}\mathbf{f}$  ( $\sigma_{I_{PG}}^2 = \boldsymbol{\beta}'_{PG}\Phi\boldsymbol{\beta}_{PG}$ ), that is,  $\sigma_{H_{PG}I_{PG}} = \mathbf{w}'_{PG}\mathbf{A}\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}'_{PG}\Phi\boldsymbol{\beta}_{PG} = \sigma_{I_{PG}}^2$ .

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

$$\rho_{H_{PG}I_{PG}} = \frac{\sqrt{\beta'_{PG}\Phi\beta_{PG}}}{\sqrt{\mathbf{w}'_{PG}\mathbf{A}\mathbf{w}_{PG}}} \quad (3.47)$$

where  $\mathbf{w}'_{PG}\mathbf{A}\mathbf{w}_{PG}$  is the variance of  $H_{PG}$ . Hereafter, to simplify the notation, we write Eq. (3.47) as  $\lambda_{PG}$ .

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

$$R_{PG} = k_I \sqrt{\beta'_{PG}\Phi\beta_{PG}} \quad (3.48)$$

and

$$\mathbf{E}_{PG} = k_I \frac{\mathbf{A}\beta_{PG}}{\sqrt{\beta'_{PG}\Phi\beta_{PG}}}, \quad (3.49)$$

respectively, where  $\beta_{PG}$  is the first eigenvector of Eq. (3.45).

### 3.5.2 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<sub>2</sub> population with 244 genotypes, 233 molecular markers, and three traits—GY (ton ha<sup>-1</sup>), EHT (cm), and PHT (cm)—where

$$\widehat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}, \quad \widehat{\mathbf{G}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix} \text{ and}$$

$$\widehat{\mathbf{\Gamma}} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix} \text{ are the estimated matrices of } \mathbf{P}, \mathbf{G}, \text{ and } \mathbf{\Gamma} \text{ respectively, whereas } \mathbf{w}' = [5 \quad -0.1 \quad -0.1 \quad 0 \quad 0 \quad 0] \text{ was the vector of economic weights.}$$

The estimated CPPG-LGSI vector of coefficients was  $\widehat{\beta}_{CP} = \widehat{\beta}_{CG} + \widehat{\beta}_{CP}\widehat{\delta}$  (see Chap. 1 for additional details). Let  $\widehat{\mathbf{A}} = \begin{bmatrix} \widehat{\mathbf{G}} & \widehat{\mathbf{\Gamma}} \\ \widehat{\mathbf{\Gamma}} & \widehat{\mathbf{\Gamma}} \end{bmatrix}$  and  $\widehat{\Phi} = \begin{bmatrix} \widehat{\mathbf{P}} & \widehat{\mathbf{\Gamma}} \\ \widehat{\mathbf{\Gamma}} & \widehat{\mathbf{\Gamma}} \end{bmatrix}$  be the estimated block matrices and  $\mathbf{d}'_{PG} = [7 \quad -3 \quad 3.5 \quad -1.5]$  the vector of PPG imposed by the breeder on the traits GY and EHT, and their associated genomic estimated breeding values (GEBV<sub>GY</sub> and GEBV<sub>EHT</sub>), 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  $\hat{\theta}_{CP} = -0.00009$  is the estimated value of the proportionality constant,  $\hat{\delta}' = [-112.92 \ -72.16 \ 61.35 \ 231.79 \ 64.75 \ -61.35]$ ,  $\hat{\beta}'_{CP} = [-0.01 \ 0.01 \ -0.01 \ 0.59 \ 0.09 \ -0.09]$  is the estimated CPPG-LGSI vector of coefficients, and the estimated CPPG-LGSI can be written as

$$\begin{aligned}\hat{I}_{CP} = & -0.01GY + 0.01EHT - 0.01PHT + 0.59GEBV_{GY} + 0.09GEBV_{EHT} \\ & - 0.09GEBV_{PHT}\end{aligned}$$

where GEBV<sub>GY</sub>, GEBV<sub>EHT</sub>, and GEBV<sub>PHT</sub> 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  $\hat{R}_{CP} = k_I \sqrt{\hat{\beta}'_{CP} \hat{\Phi} \hat{\beta}_{CP}} = 0.443$  and  $\hat{E}'_{CP} = k_I \frac{\hat{\beta}'_{CP} \hat{\mathbf{A}}}{\sqrt{\hat{\beta}'_{CP} \hat{\Phi} \hat{\beta}_{CP}}} = [-0.004 \ 0.002 \ -4.639 \ -0.002 \ 0.001 \ -4.326]$

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

Because the estimated value of the proportionality constant was negative ( $\hat{\theta}_{CP} = -0.00009$ ), the expected genetic gains of the traits GY and EHT, and their associated genomic estimated breeding values (GEBV<sub>GY</sub> and GEBV<sub>EHT</sub>), which appeared in the  $\hat{E}'_{CP}$  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  $\hat{\theta}_{CP}$  to be positive (see Chaps. 3 and 6 for details).

In the PPG-GESIM, we need to find the solutions to equation  $(\hat{\mathbf{T}}_{PG} - \hat{\lambda}_{PG_j}^2 \mathbf{I}_{2t})$   $\hat{\beta}_{PG_j} = \mathbf{0}$ , for  $\hat{\lambda}_{PG_j}^2$  and  $\hat{\beta}_{PG_j}$  (see Eq. 3.45). The estimated PPG-GESIM vector of coefficients was  $\hat{\beta}'_{PG} = [0.001 \ -0.050 \ 0.029 \ 0.975 \ 0.154 \ -0.157]$ , 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  $\hat{\beta}'_{PG}$  values, we can estimate the PPG-GESIM index as

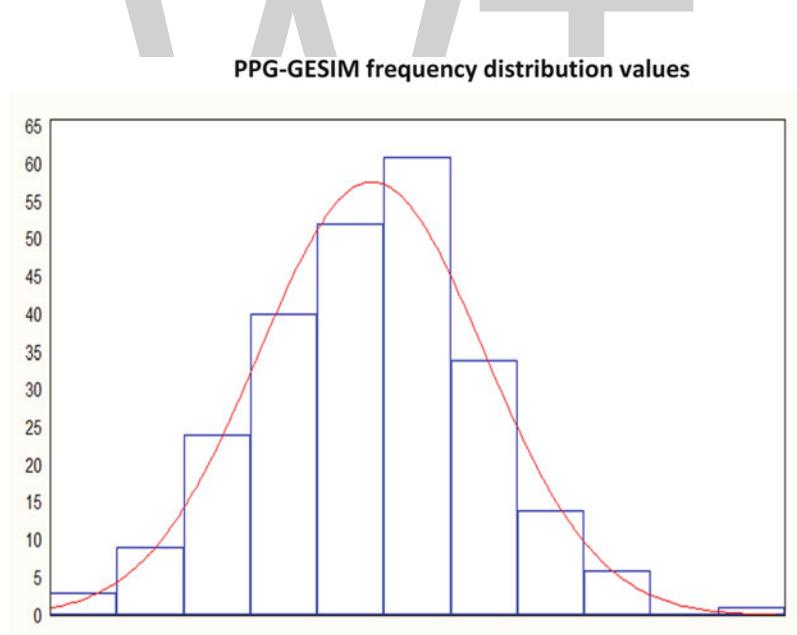


$$\widehat{I}_{PG} = 0.001GY - 0.05EHT + 0.029PHT + 0.975GEBV_{GY} + 0.154GEBV_{EHT} - 0.157GEBV_{PHT}$$

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  $\widehat{R}_{PG} = k_I \sqrt{\widehat{\beta}'_{PG} \widehat{\Phi} \widehat{\beta}_{PG}} = 0.696$ ,  $\widehat{\rho}_{H_{PG} I_{PG}} = \frac{\widehat{\sigma}_{I_{PG}}}{\widehat{\sigma}_{H_{PG}}} = 0.843$ , and  $\widehat{E}'_{PG} = k_I \frac{\widehat{\beta}'_{PG} \widehat{\mathbf{A}}}{\sqrt{\widehat{\beta}'_{PG} \widehat{\Phi} \widehat{\beta}_{PG}}} = [0.01 \ -1.00 \ -3.56 \ 0 \ -0.46 \ -3.98]$  respectively.

Fig. 3.4 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. 3.4** 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  $\mathbf{F}$  was an identity matrix of size  $8 \times 8$  for all four selection cycles.

Table 3.6 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 3.6** 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

Cycle	CPPG-LGSI expected genetic gain for one predetermined restriction							
	Traits				Genomic estimated breeding values			
	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								
Cycle	Traits				Genomic estimated breeding values			
	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
	8.06	-3.46	3.30	2.06	4.03	-1.73	1.72	0.98
1	8.17	-3.50	3.08	2.65	4.09	-1.75	1.79	0.98
2	8.88	-3.81	3.31	1.83	4.44	-1.90	1.72	0.90
3	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								
Cycle	Traits				Genomic estimated breeding values			
	T1	T2	T3	T4	GEBV1	GEBV2	GEBV3	GEBV4
	5.77	-2.47	4.12	2.28	2.88	-1.24	2.06	0.98
1	5.68	-2.43	4.06	2.76	2.84	-1.22	2.03	0.97
2	5.87	-2.52	4.20	1.98	2.94	-1.26	2.10	0.79
3	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  $\mathbf{d}'_{PG} = [7 \quad -3 \quad 5 \quad 3.5 \quad -1.5 \quad 2.5]$ . Trait T4 and its associated GEBV4 were not restricted



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

Tables 3.7 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 3.28, -4.12, 3.23, 2.23, 4.14, -2.26, 1.71, and 1.01 for one restriction; 3.43, -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 3.12, -3.27, 2.99, 1.13, 2.19, -1.15, 1.30, and 0.45 for three

**Table 3.7** 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

Cycle	PPG-GESIM expected genetic gain for one predetermined restriction							
	Traits				Genomic estimated breeding values			
	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
Cycle	PPG-GESIM expected genetic gain for two predetermined restrictions							
	Traits				Genomic estimated breeding values			
	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
Cycle	PPG-GESIM expected genetic gain for three predetermined restrictions							
	Traits				Genomic estimated breeding values			
	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



## Multistage Selection Indices for Maximum Genetic Gain



**Abstract** Multistage linear selection indices select individual traits available at different times or stages and are applied mainly in animals and tree breeding, where the traits under consideration become evident at different ages. The main indices are: the unrestricted, the restricted, and the predetermined proportional gain selection index. The restricted and predetermined proportional gain indices allow null and predetermined restrictions to be imposed on the trait expected genetic gain (or multi-trait selection response) values, whereas the rest of the traits remain changed without any restriction. The three indices can use phenotypic, genomic, or both sets of information to predict the unobservable net genetic merit values of the candidates for selection and all of them maximize the selection response, the expected genetic gain for each trait, have maximum accuracy, are the best predictor of the net genetic merit, and provide the breeder with an objective rule for evaluating and selecting several traits simultaneously. The theory of the foregoing indices is based on the independent culling method and on the linear phenotypic selection index, and is described in this chapter in the phenotypic and genomic selection context. Their theoretical results are validated in a two-stage breeding selection scheme using real and simulated data.

### 4.1 Multistage Linear Phenotypic Selection Index

In a similar manner to the linear phenotypic selection index (LPSI, Chap. 2), the objectives of the multistage linear phenotypic selection index (MLPSI) are:

1. To predict the net genetic merit  $H = \mathbf{w}'\mathbf{g}$ , where  $\mathbf{g}' = [g_1 \ g_2 \ \dots \ g_t]$  is the vector of true breeding values of an individual for  $t$  traits and  $\mathbf{w}' = [w_1 \ w_2 \ \dots \ w_t]$  is the vector of economic weights.
2. To select individuals with the highest  $H$  values at each stage as parents of the next generation.
3. To maximize the MLPSI selection response and its expected genetic gain per trait.



4. To provide the breeder with an objective rule for evaluating and selecting several traits simultaneously.

When selection is based on all the individual traits of interest jointly, the LPSI vector of coefficients that maximizes the selection response  $R = k\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}$  and the expected genetic gain per trait  $\mathbf{E} = k\frac{\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}}$  is  $\mathbf{b} = \mathbf{P}^{-1}\mathbf{C}\mathbf{w}$ , where  $\mathbf{C}$  and  $\mathbf{P}$  are the covariance matrices of the true breeding values ( $\mathbf{g}$ ) and trait phenotypic values ( $\mathbf{y}$ ) respectively, and  $k$  is the selection intensity. In MLPSI terminology, the LPSI is called a one-stage selection index. The MLPSI is an extension of the LPSI theory to the multistage selection context and, as we shall see, the MLPSI theoretical results are very similar to the LPSI theoretical results described.

#### 4.1.1 The MLPSI Parameters for Two Stages

Let  $\mathbf{y}' = [y_1 \ y_2 \ \dots \ y_t]$  be a vector with  $t$  traits of interest and suppose that we can select only  $n_i$  of them ( $n_i < t$ ) at stage  $i$  ( $i = 1, 2, \dots, N$ ), such that after  $N$  stages ( $N < t$ ),  $\sum_{i=1}^N n_i = t$ . Thus, for each stage we should have a selection index with a different

number of traits. For example, at stage  $i$  the index would be  $I_i = \sum_{j=1}^{n_i} b_{ij}y_{ij}$ , and at stage  $N$  the index would be  $I_N = \sum_{j=1}^{n_1} b_{1j}y_{1j} + \sum_{j=1}^{n_2} b_{2j}y_{2j} + \dots + \sum_{j=1}^{n_N} b_{Nj}y_{Nj} = \sum_{i=1}^N I_i$ ,

where the double subscript of  $y_{ij}$  indicates that the  $j$ th trait is measured at stage  $i$ , so that at each sub-index  $I_i$ , all the  $n_i$  traits are measured at the same age.

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

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

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

respectively. For the second stage, in addition to matrix  $\mathbf{P}_1$ , we need the phenotypic covariance matrix between  $\mathbf{x}_1$  and  $\mathbf{x}_2$  ( $\mathbf{P}_{12}$ ) and the phenotypic covariance matrix of  $\mathbf{x}_2$  ( $\mathbf{P}_2$ ); thus, the covariance matrix of phenotypic values at stage 2 is  $\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_2 \end{bmatrix}$ . In a similar manner, in addition to matrix  $\mathbf{G}_1$ , at stage 2 we need the covariance between  $\mathbf{x}_2$  and  $\mathbf{g}$  ( $\mathbf{G}_2$ ); that is, at stage 2 the covariance matrix



between phenotypic and breeding values can be written as  $\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$ . Matrices  $\mathbf{G}$  and  $\mathbf{C}$  are not exactly the same, because although  $\mathbf{C} = \text{Var}(\mathbf{g})$ ,  $\mathbf{G} = \begin{bmatrix} \text{Cov}(\mathbf{x}_1, \mathbf{g}) \\ \text{Cov}(\mathbf{x}_2, \mathbf{g}) \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$  and this latter matrix changes at each stage.

Let  $\mathbf{w}' = [w_1 \ w_2 \ w_3 \ w_4]$  be the vector of economic weights; then, at the first and second stages the MLPSI vectors of coefficients are  $\mathbf{b}'_1 = \mathbf{w}' \mathbf{G}'_1 \mathbf{P}_1^{-1} = [b_{11} \ b_{12}]$  and  $\mathbf{b}'_2 = \mathbf{w}' \mathbf{G}'_2 \mathbf{P}_2^{-1} = [b_{21} \ b_{22} \ b_{23} \ b_{24}]$  respectively. The selection indices at stages 1 and 2 can be written as  $I_1 = b_{11}y_1 + b_{12}y_2 = \mathbf{b}'_1 \mathbf{x}_1$  and  $I_2 = b_{21}y_1 + b_{22}y_2 + b_{23}y_3 + b_{24}y_4 = \mathbf{b}'_2 \mathbf{y}$ , which could be correlated and then numerical integration would be required to find optimal truncation points and selection intensities (Xu and Muir 1992; Hicks et al. 1998) before obtaining the maximized MLPSI selection response and expected genetic gain per trait.

The accuracy of the MLPSI at stages 1 and 2 can be written as

$$\rho_{HI_1} = \sqrt{\frac{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{HI_2} = \sqrt{\frac{\mathbf{b}'_2 \mathbf{P}^* \mathbf{b}_2}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}, \quad (4.1)$$

respectively. Let  $k_1$  and  $k_2$  be the selection intensities for stages 1 and 2; then, the maximized MLPSI expected genetic gains per trait can be written as

$$\mathbf{E}_1 = k_1 \frac{\mathbf{G}'_1 \mathbf{b}_1}{\sqrt{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}} \quad \text{and} \quad \mathbf{E}_2 = k_2 \frac{\mathbf{b}'_2 \mathbf{C}^*}{\sqrt{\mathbf{b}'_2 \mathbf{P}^* \mathbf{b}_2}}, \quad (4.2)$$

and the total expected genetic gain per trait for the two stages is equal to  $\mathbf{E}_1 + \mathbf{E}_2$ . In a similar manner, the maximized selection responses for both stages are

$$R_1 = k_1 \sqrt{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1} \quad \text{and} \quad R_2 = k_2 \sqrt{\mathbf{b}'_2 \mathbf{P}^* \mathbf{b}_2}, \quad (4.3)$$

and the total selection response for the two stages is  $R_1 + R_2$ . In Eqs. (4.1) to (4.3), matrices  $\mathbf{P}^*$  and  $\mathbf{C}^*$  are matrices  $\mathbf{P}$  and  $\mathbf{C}$  respectively, adjusted for previous selection on  $I_1 = \mathbf{b}'_1 \mathbf{x}_1$ . That is, the MLPSI accuracy, expected genetic gain per trait, and selection response at stage 2 are affected by previous selection on  $I_1$  (Saxton 1983) and it is necessary to adjust  $\mathbf{P}$  and  $\mathbf{C}$ .

One method for adjusting matrices  $\mathbf{P}$  and  $\mathbf{C}$  has been provided by Cochran (1951) and Cunningham (1975). Suppose that  $X$ ,  $Y$ , and  $W$  are three jointly normally distributed random variables and that the covariance among them is known, then the covariance between  $X$  and  $Y$  adjusted for the effects of selection on  $W$  can be obtained as

$$\text{Cov}(X, Y)^* = \text{Cov}(X, Y) - u \frac{\text{Cov}(X, W)\text{Cov}(Y, W)}{\text{Var}(W)}, \quad (4.4)$$

where  $u = k_1(k_1 - \tau)$ ,  $k_1$  is the selection intensity at stage 1 and  $\tau$  is the truncation point when  $I_1 = \mathbf{b}'_1 \mathbf{x}_1$  is applied. For example, if the selection intensity at the first stage is 5%,  $k_1 = 2.063$ ,  $\tau = 1.645$ , and  $u = 0.862$  (Falconer and Mackay 1996, Table A).

According to Dekkers (2014), with the result of Eq. (4.4), it is possible to obtain matrices  $\mathbf{P}^*$  and  $\mathbf{C}^*$  using the following two equations:

$$\begin{aligned}\mathbf{P}^* &= Var(\mathbf{y})^* = \mathbf{P} - u \frac{Cov(\mathbf{y}, \mathbf{x}_1) \mathbf{b}_1 \mathbf{b}'_1 Cov(\mathbf{x}_1, \mathbf{y})}{\mathbf{b}'_1 Var(\mathbf{x}_1) \mathbf{b}_1} \\ &= \mathbf{P} - u \frac{\begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_{21} \end{bmatrix} \mathbf{b}_1 \mathbf{b}'_1 [\mathbf{P}_1 \quad \mathbf{P}_{21}]}{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}\end{aligned}\quad (4.5)$$

and

$$\mathbf{C}^* = Var(\mathbf{g})^* = \mathbf{C} - u \frac{Cov(\mathbf{g}, \mathbf{x}_1) \mathbf{b}_1 \mathbf{b}'_1 Cov(\mathbf{x}_1, \mathbf{g})}{\mathbf{b}'_1 Var(\mathbf{x}_1) \mathbf{b}_1} = \mathbf{C} - u \frac{\mathbf{G}'_1 \mathbf{b}_1 \mathbf{b}'_1 \mathbf{G}_1}{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}. \quad (4.6)$$

With the Eq. (4.5) result, the correlation between  $I_1 = \mathbf{b}'_1 \mathbf{x}_1$  and  $I_2 = \mathbf{b}'_2 \mathbf{y}$  is

$$Corr(I_1, I_2) = \frac{\mathbf{b}'_1 [\mathbf{P}_1 \quad \mathbf{P}_{21}] \mathbf{b}_2}{\sqrt{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1} \sqrt{\mathbf{b}'_2 \mathbf{P}_2 \mathbf{b}_2}} = \rho_{12}, \quad (4.7)$$

where  $\sqrt{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}$  and  $\sqrt{\mathbf{b}'_2 \mathbf{P}_2 \mathbf{b}_2}$  are the standard deviations of the variances of  $I_1 = \mathbf{b}'_1 \mathbf{x}_1$  and  $I_2 = \mathbf{b}'_2 \mathbf{y}$  respectively.

#### 4.1.2 The Selection Intensities

Selection intensity  $k$  is related to the height of the ordinate of the normal curve ( $z$ ) and the proportion selected ( $p$ ) in the LPSI as  $k = z/p$ . In the multistage selection context, it is usual to fix the total proportion to be selected ( $p$ ) before selection is carried out and then to determine the unknown proportion  $q_i$  ( $i=1, 2, \dots, N$ ) for each stage under the restriction

$$p = \prod_{i=1}^N q_i, \quad (4.8)$$

where  $N$  is the number of stages. In the two-stage selection scheme, we would have  $p = q_1 q_2$ . Based on the fixed proportion  $p$  and the  $\rho_{12}$  value (Eq. 4.7), Young (1964) used the bivariate truncated normal distribution theory to obtain the selection intensity for two stages. A truncated distribution is a conditional distribution



resulting when the domain of the parent distribution is restricted to a smaller region (Hattaway 2010). In the multistage selection context, a truncation occurs when a sample of individuals from the parent distribution are selected as parents for the next selection cycle, thus creating a new population of individuals that follow a truncated normal distribution.

Suppose that  $I_1 = \mathbf{b}'\mathbf{x}_1$  and  $I_2 = \mathbf{b}'\mathbf{y}$  have joint normal distribution and let  $I_1$  and  $I_2$  be transformed as  $v_1 = \frac{I_1 - \mu_{I_1}}{\sigma_{I_1}}$  and  $v_2 = \frac{I_2 - \mu_{I_2}}{\sigma_{I_2}}$  with a mean of zero and a variance of 1, where  $\mu_{I_1}$  and  $\mu_{I_2}$  are the means, whereas  $\sigma_{I_1}$  and  $\sigma_{I_2}$  are the standard deviations of the variances of  $I_1$  and  $I_2$  respectively. In this case, the method of selection is to retain animals or plants with  $v_1 \geq c_1$  at stage 1 and  $v_1 + v_2 \geq c_2$  at stage 2, where  $c_1$  and  $c_2$  are truncation points for  $I_1$  and  $I_2$  respectively.

The selected population has bivariate left truncated normal distribution with a probability density function given by  $h(v_1, v_2) = \frac{f(v_1, v_2)}{p}$ , where  $f(v_1, v_2) = \frac{1}{2\pi\sqrt{1-\rho_{12}^2}} \exp\left\{-\frac{1}{2(1-\rho_{12}^2)}[v_1^2 + v_2^2 - 2\rho_{12}v_1v_2]\right\}$  and  $\rho_{12}$  is the correlation between  $v_1$  and  $v_2$ . The fixed total proportion ( $p$ ) before selection can be written as  $p = \int_{c_1}^{\infty} \int_{c_2-v_1}^{\infty} f(v_1, v_2) dv_2 dv_1$ , where  $c_1$  and  $c_2$  are truncation points for  $I_1$  and  $I_2$ , respectively. Then, as  $p$  is fixed, Young (1964) integrated by parts (Thomas 2014)

$$\int_{c_1}^{\infty} \int_{c_2-v_1}^{\infty} f(v_1, v_2) dv_2 dv_1 \quad (4.9)$$

and found the expectations of  $v_1$  and  $v_2$  in the selected population, writing the selection intensity values for stages 1 ( $k_1$ ) and 2 ( $k_2$ ) as

$$k_1 = \frac{z(c_1)Q(a)}{p} + \frac{z(c_3)Q(b)\sqrt{(1+\rho_{12})/2}}{p} \quad (4.10)$$

and

$$k_2 = \frac{\rho_{12}z(c_1)Q(a)}{p} + \frac{z(c_3)Q(b)\sqrt{(1+\rho_{12})/2}}{p} \quad (4.11)$$

respectively, where  $z(c_1) = \frac{\exp\{-0.5c_1^2\}}{\sqrt{2\pi}}$  and  $z(c_3) = \frac{\exp\{-0.5c_3^2\}}{\sqrt{2\pi}}$  are the heights of the ordinates of the standard normal distribution at the lowest value of  $c_1$  and  $c_3 = \frac{c_2}{\sqrt{2(1+\rho_{12})}}$  and  $p$  is the total proportion of the population of animal or plant lines selected;  $a = \frac{c_2 - c_1(1 + \rho_{12})}{\sqrt{1 - \rho_{12}^2}}$  and  $b = \frac{c_1 - c_2}{\sqrt{2(1 - \rho_{12})}}$ , whereas  $Q(a) = 1 - \Phi(a)$  and  $Q(b) = 1 - \Phi(b)$  are the complement of the standard normal distribution;  $\Phi(a) = \int_{-\infty}^a \frac{1}{\sqrt{2\pi}} \exp\{-0.5w^2\} dw$  and  $\Phi(b) = \int_{-\infty}^b \frac{1}{\sqrt{2\pi}} \exp\{-0.5t^2\} dt$  are



probabilities of the standard normal distribution, i.e.,  $\Phi(a) = P_r(W \leq a)$  and  $\Phi(b) = P_r(T \leq b)$ .

Young (1964) provided figures to obtain values of  $c_1$  and  $c_2$  when the  $\rho_{12}$  values are between -0.8 and 0.8, and the  $p$  values are between 0.05 and 0.8. For example, suppose that  $\rho_{12} = 0.8$  and  $p = 0.2$  (or 20%), then, according to Young (1964, Fig. 9),  $c_1 = 0.80$  and  $c_2 = 1.6$ , and to find the selection intensities for the first ( $k_1$ ) and second stages ( $k_2$ ) we need to solve Eqs. (4.10) and (4.11). That is, as  $c_1 = 0.80$ ,

$$c_2 = 1.6, \rho_{12} = 0.8, \text{ and } p = 0.2, \text{ then } z(c_1) = \frac{\exp\{-0.5(0.8)^2\}}{\sqrt{2\pi}} = 0.290, \\ z(c_3) = \frac{\exp\{-0.5[(1.6)^2/2(1.8)]\}}{\sqrt{2\pi}} = 0.28, \quad a = \frac{1.6-0.8(1.8)}{\sqrt{1-(0.8)^2}} = 0.27, \quad b = \frac{2(0.8)-1.6}{\sqrt{2(0.2)}} = 0,$$

$\Phi(a) = 0.6064$ ,  $\Phi(b) = 0.5$ ,  $Q(a) = 1 - \Phi(a) = 0.3936$ , and  $Q(b) = 1 - \Phi(b) = 0.5$ . Based on these results, the selection intensities for stages 1 and 2 are

$$k_1 = \frac{(0.29)(0.3936)}{0.2} + \frac{(0.28)(0.5)(0.9)}{0.2} = 0.744 \quad \text{and}$$

$$k_2 = \frac{(0.8)(0.29)(0.3936)}{0.2} + \frac{(0.28)(0.5)(0.9)}{0.2} = 0.721$$

respectively. Note that the values of  $\Phi(a) = 0.6064$  and  $\Phi(b) = 0.5$  can be obtained from any table with values showing the area under the curve of the standard normal distribution (e.g., Rausand and Høyland 2004, Table F.1).

One problem with Eqs. (4.10) and (4.11) is that they tend to overestimate the selection intensities values and also overestimate the selection response when the total proportion retained  $p$  is lower than 10%. Cochran (1951) have given two equations to obtain selection intensities in the two stages context but his equations also overestimate the selection intensities values when  $p$  is lower than 10%. Up to now, there is not an accurate method to estimate selection intensities for two or more stages in the MLPSI context. Mi et al. (2014) have developed an R package called *selectiongain* that enables calculation of the OMLPSI selection response for up to 20 selection stages. *Selectiongain* uses raw integration to obtain the first moment of a lower truncated multivariate standard normal distribution and then it estimates the OMLPSI selection response at each stage; however, this integral requires complex numerical algorithms with no convergence criteria (Arismendi 2013) and could also overestimate the selection intensity at each stage.

### 4.1.3 Numerical Example

To illustrate the two-stage selection theory, we use the poultry data of Xu and Muir (1992). This data set contains four traits: age at sexual maturity, defined as the age



(in days) at which the first trap-nested egg was laid ( $y_1$ ); rate of lay, defined as 100 times (total eggs in the laying period)/(total days in the laying period) ( $y_2$ ); body weight (in pounds) measured at 32 weeks of age ( $y_3$ ); and average egg weight (in ounces per dozen) of all the eggs laid up to 32 weeks of age ( $y_4$ ). The estimated phenotypic and

$$\text{genetic covariance matrices were } \hat{\mathbf{P}} = \begin{bmatrix} 137.178 & -90.957 & 0.136 & 0.564 \\ -90.957 & 201.558 & 1.103 & -1.231 \\ 0.136 & 1.103 & 0.202 & 0.104 \\ 0.564 & -1.231 & 0.104 & 2.874 \end{bmatrix}$$

$$\text{and } \hat{\mathbf{C}} = \begin{bmatrix} 14.634 & -18.356 & -0.109 & 1.233 \\ -18.356 & 32.029 & 0.103 & -2.574 \\ -0.109 & 0.103 & 0.089 & 0.023 \\ 1.233 & -2.574 & 0.023 & 1.225 \end{bmatrix} \text{ respectively, whereas}$$

the vector of economic weights for the four traits was  $\mathbf{w}' = [-3.555 \ 19.536 \ -113.746 \ 48.307]$ .

Suppose that at the first and second stages we select two traits ( $n_1 = n_2 = 2$ ); then,  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2]$ , where  $\mathbf{x}'_1 = [y_1 \ y_2]$  and  $\mathbf{x}'_2 = [y_3 \ y_4]$ . The estimated phenotypic ( $\hat{\mathbf{P}}_1$ ) and genetic ( $\hat{\mathbf{G}}_1$ ) covariance matrices for the first stage were

$$\hat{\mathbf{P}}_1 = \begin{bmatrix} 137.178 & -90.957 \\ -90.957 & 1.103 \end{bmatrix} \text{ and } \hat{\mathbf{G}}_1 = \begin{bmatrix} 14.634 & -18.356 & -0.109 & 1.233 \\ -18.356 & 32.029 & 0.103 & -2.574 \end{bmatrix}$$

respectively. For the first and second stages, the estimated MLPSI vector of coefficients were  $\hat{\mathbf{b}}'_1 = \mathbf{w}' \hat{\mathbf{G}}'_1 \hat{\mathbf{P}}_1 = [-0.918 \ 2.339]$  and  $\hat{\mathbf{b}}'_2 = \hat{\mathbf{w}}' \hat{\mathbf{C}} \hat{\mathbf{P}}^{-1} = [-0.59 \ 2.78 \ -49.45 \ 3.75]$  respectively.

The estimated correlation value between the estimated indices  $\hat{I}_1 = \hat{\mathbf{b}}'_1 \mathbf{x}_1$  and  $\hat{I}_2 = \hat{\mathbf{b}}'_2 \mathbf{y}$  was  $\hat{\rho}_{12} = \frac{\hat{\mathbf{b}}'_1 [\hat{\mathbf{P}}_1 \ \hat{\mathbf{P}}_{21}] \hat{\mathbf{b}}_2}{\sqrt{\hat{\mathbf{b}}'_1 \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_1} \sqrt{\hat{\mathbf{b}}'_2 \hat{\mathbf{P}}_2 \hat{\mathbf{b}}_2}} = 0.88$ , where  $\sqrt{\hat{\mathbf{b}}'_1 \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_1}$  and  $\sqrt{\hat{\mathbf{b}}'_2 \hat{\mathbf{P}}_2 \hat{\mathbf{b}}_2}$

were the estimated standard deviations of the variance of  $\hat{I}_1$  and  $\hat{I}_2$  respectively. Assuming that  $p = 0.2$  (or 20%), an approximate selection intensity for the first stage was  $k_1 = 0.744$ , whence the estimated MLPSI selection response, expected genetic gain per trait, and accuracy were  $\hat{R}_1 = k_1 \sqrt{\hat{\mathbf{b}}'_1 \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_1} = 29.85$ ,  $\hat{\mathbf{E}}'_1 = k_1 \frac{\hat{\mathbf{G}}'_1 \hat{\mathbf{b}}_1}{\sqrt{\hat{\mathbf{b}}'_1 \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_1}} = [-1.046 \ 1.702 \ 0.006 \ -0.133]$ , and  $\hat{\rho}_{HI_1} = \sqrt{\frac{\hat{\mathbf{b}}'_1 \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_1}{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}}} = 0.353$  respectively.

According to the  $k_1 = 0.744$  value, the approached value of  $u$  was  $u = 0.554$ , and by Eqs. (4.5) and (4.6), the estimated and adjusted phenotypic ( $\hat{\mathbf{P}}^*$ ) and genetic ( $\hat{\mathbf{C}}^*$ ) covariance matrices for the second stage were

$$\hat{\mathbf{P}}^* = \begin{bmatrix} 97.682 & -26.241 & 0.422 & 0.168 \\ -26.241 & 95.518 & 0.634 & -0.582 \\ 0.422 & 0.634 & 0.200 & 0.107 \\ 0.168 & -0.582 & 0.107 & 2.870 \end{bmatrix} \text{ and}$$

★ ★ ★

$$\hat{\mathbf{C}}^* = \begin{bmatrix} 13.540 & -16.575 & -0.102 & 1.094 \\ -16.575 & 29.129 & 0.092 & -2.348 \\ -0.102 & 0.092 & 0.089 & 0.024 \\ 1.094 & -2.384 & 0.024 & 1.207 \end{bmatrix}, \text{ respectively.}$$

For the second stage, the approximated selection intensity was  $k_2 = 0.721$ , whereas the estimated MLPSI selection response, expected genetic gain per trait and accuracy, were  $\hat{R}_2 = k_{I_2} \sqrt{\hat{\mathbf{b}}'_2 \hat{\mathbf{P}}_2^* \hat{\mathbf{b}}_2} = 24.84$ ,  $\hat{\mathbf{E}}'_2 = k_{I_2} \frac{\hat{\mathbf{C}}^{*'} \hat{\mathbf{b}}_2}{\sqrt{\hat{\mathbf{b}}'_2 \hat{\mathbf{P}}_2^* \hat{\mathbf{b}}_2}} = [-0.443 \quad 0.804 \quad -0.087 \quad -0.087]$ , and  $\hat{\rho}_{HI_2} = \sqrt{\frac{\hat{\mathbf{b}}'_2 \hat{\mathbf{P}}_2^* \hat{\mathbf{b}}_2}{\mathbf{w}' \hat{\mathbf{C}}^* \mathbf{w}}} = 0.314$  respectively. Finally, the total estimated MLPSI selection response and expected genetic gain per trait were  $\hat{R}_1 + \hat{R}_2 = 54.69$  and  $\hat{\mathbf{E}}'_1 + \hat{\mathbf{E}}'_2 = [-1.488 \quad 2.506 \quad -0.081 \quad -0.219]$ .

## 4.2 The Multistage Restricted Linear Phenotypic Selection Index

The multistage restricted linear phenotypic selection index (MRLPSI) is an extension of the null restricted linear phenotypic selection index (RLPSI) described in Chap. 3 to the multistage case; thus, the theoretical results of the MRLPSI are very similar to those of the RLPSI. The MRLPSI allows restrictions equal to zero to be imposed on the expected genetic gains of some traits, whereas other traits increase (or decrease) their expected genetic gains without any restrictions being imposed.

### 4.2.1 The MRLPSI Parameters for Two Stages

In Chap. 3, we indicated that vector  $\mathbf{b}_R = \mathbf{K}\mathbf{b}$  is a linear transformation of the LPSI vector of coefficients ( $\mathbf{b}$ ) made by the projector matrix  $\mathbf{K}$ , and that matrix  $\mathbf{K}$  is idempotent ( $\mathbf{K} = \mathbf{K}^2$ ) and projects  $\mathbf{b}$  into a space smaller than the original space of  $\mathbf{b}$ . The reduction of the space into which matrix  $\mathbf{K}$  projects  $\mathbf{b}$  is equal to the number of zeros that appears on the expected genetic gain per trait. Hence, the MRLPSI vector of coefficients for stages 1 and 2 should be a linear transformation of the MLPSI vector of coefficients at stages 1 ( $\mathbf{b}_1 = \mathbf{P}_1^{-1} \mathbf{G}_1 \mathbf{w}$ ) and 2 ( $\mathbf{b}_2 = \mathbf{P}_2^{-1} \mathbf{C} \mathbf{w}$ ) described in Sect. 4.1.1 of this chapter, and should be written as

$$\mathbf{b}_{R_1} = \mathbf{K}_1 \mathbf{b}_1 \quad (4.12)$$

and

$$\mathbf{b}_{R_2} = \mathbf{K}_2 \mathbf{b}_2, \quad (4.13)$$



respectively, where, at stage 1,  $\mathbf{K}_1 = [\mathbf{I}_1 - \mathbf{Q}_1]$ ,  $\mathbf{Q}_1 = \mathbf{P}_1^{-1} \boldsymbol{\Psi}_1 (\boldsymbol{\Psi}'_1 \mathbf{P}_1^{-1} \boldsymbol{\Psi}_1)^{-1} \boldsymbol{\Psi}'_1$ ,  $\boldsymbol{\Psi}'_1 = \mathbf{U}' \mathbf{G}'_1$ ,  $\mathbf{I}_1$  is an identity matrix of the same size as  $\mathbf{P}_1$ , and  $\mathbf{P}_1^{-1}$  is the inverse of matrix  $\mathbf{P}_1$ . At stage 2,  $\mathbf{K}_2 = [\mathbf{I}_2 - \mathbf{Q}_2]$ ,  $\mathbf{Q}_2 = \mathbf{P}^{-1} \boldsymbol{\Psi}_2 (\boldsymbol{\Psi}'_2 \mathbf{P}^{-1} \boldsymbol{\Psi}_2)^{-1} \boldsymbol{\Psi}'_2$ ,  $\boldsymbol{\Psi}'_2 = \mathbf{U}' \mathbf{C}$ ,  $\mathbf{I}_2$  is an identity matrix of the same size as  $\mathbf{P}$ , and  $\mathbf{P}^{-1}$  is the inverse of matrix  $\mathbf{P}$ . By Eqs. (4.12) and (4.13), the MRLPSI for stages 1 and 2 can be written as  $I_1 = \mathbf{b}'_{R_1} \mathbf{x}_1$  and  $I_2 = \mathbf{b}'_{R_2} \mathbf{y}$ , where  $\mathbf{y}' = [\mathbf{x}'_1 \quad \mathbf{x}'_2]$ ;  $\mathbf{x}'_1$  and  $\mathbf{x}'_2$  are the vectors of traits that become evident at the first and second stages respectively.

Let  $k_1$  and  $k_2$  be the selection intensities for stages 1 and 2 (Eqs. 4.10 and 4.11) respectively, and let  $\mathbf{P}^*$  and  $\mathbf{C}^*$  be the covariance matrices adjusted in the MRLPSI context according to Eqs. (4.5) and (4.5) respectively. The maximized MRLPSI selection response, expected genetic gain per trait, and accuracy at stages 1 and 2 can be written as

$$R_{R_1} = k_1 \sqrt{\mathbf{b}'_{R_1} \mathbf{P}_1 \mathbf{b}_{R_1}} \quad \text{and} \quad R_{R_1} = k_2 \sqrt{\mathbf{b}'_{R_2} \mathbf{P}^* \mathbf{b}_{R_2}}, \quad (4.14)$$

$$\mathbf{E}_{R_1} = k_1 \frac{\mathbf{G}'_1 \mathbf{b}_{R_1}}{\sqrt{\mathbf{b}'_{R_1} \mathbf{P}_1 \mathbf{b}_{R_1}}} \quad \text{and} \quad \mathbf{E}_{R_2} = k_2 \frac{\mathbf{b}'_{R_2} \mathbf{C}^*}{\sqrt{\mathbf{b}'_{R_2} \mathbf{P}^* \mathbf{b}_{R_2}}} \quad (4.15)$$

and

$$\rho_{R_1} = \sqrt{\frac{\mathbf{b}'_{R_1} \mathbf{P}_1 \mathbf{b}_{R_1}}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{R_2} = \sqrt{\frac{\mathbf{b}'_{R_2} \mathbf{P}^* \mathbf{b}_{R_2}}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}, \quad (4.16)$$

respectively, whereas the total MRLPSI selection response and expected genetic gain per trait for both stages are equal to  $R_{R_1} + R_{R_2}$  and  $\mathbf{E}_{R_1} + \mathbf{E}_{R_2}$ .

#### 4.2.2 Numerical Examples

To illustrate the MRLPSI theory for a two-stage selection breeding scheme, we use the real data set of the White Leghorn chickens of Hicks et al. (1998). This data set is conformed with six traits ( $y_1$  to  $y_6$ ) that correspond to records consisting of the number of eggs laid during different periods: from week 0 through 4 ( $y_1$ ), 4 through 8 ( $y_2$ ), 8 through 28 ( $y_3$ ), 28 through 32 ( $y_4$ ), 32 through 36 ( $y_5$ ), and 36 through 52 ( $y_6$ ) respectively. The estimated phenotypic and genotypic covariance matrices were

$$\widehat{\mathbf{P}} = \begin{bmatrix} 102 & 32 & 14 & 4 & 3 & -1 \\ 32 & 80 & 80 & 16 & 17 & 7 \\ 14 & 80 & 298 & 78 & 112 & 62 \\ 4 & 16 & 78 & 66 & 80 & 51 \\ 3 & 17 & 112 & 80 & 135 & 49 \\ -1 & 7 & 62 & 51 & 49 & 98 \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{C}} = \begin{bmatrix} 44 & 11 & -11 & -3 & -8 & -3 \\ 11 & 26 & 24 & 7 & 7 & 3 \\ -11 & 24 & 62 & 23 & 37 & 20 \\ -3 & 7 & 23 & 14 & 23 & 14 \\ -8 & 7 & 37 & 23 & 42 & 25 \\ -3 & 3 & 20 & 14 & 25 & 18 \end{bmatrix},$$

respectively, and  $\mathbf{w}' = [0.08 \ 0.08 \ 0.38 \ 0.08 \ 0.08 \ 0.31]$  was the vector of economic weights.

Let  $\mathbf{y}' = [y_1 \ y_2 \ y_3 \ y_4 \ y_5 \ y_6]$  and  $\mathbf{g}' = [g_1 \ g_2 \ g_3 \ g_4 \ g_5 \ g_6]$  be the vectors of observed phenotypic and unobserved genotypic values respectively, and suppose that at stage 1 we select four traits and at stage 2 we select two traits, then  $\mathbf{x}'_1 = [y_1 \ y_2 \ y_3 \ y_4]$  and  $\mathbf{x}'_2 = [y_5 \ y_6]$  are the vector of observations at stages 1 and 2 respectively, whereas  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2]$  is the vector of total observations at stage 2. We need to estimate vectors  $\mathbf{b}'_{R_1} = \mathbf{b}'_1 \mathbf{K}'_1$  and  $\mathbf{b}'_{R_2} = \mathbf{b}'_2 \mathbf{K}'_2$ , where  $\mathbf{b}'_1 = \mathbf{w}' \mathbf{G}'_1 \mathbf{P}_1^{-1}$  and  $\mathbf{b}'_2 = \mathbf{w}' \mathbf{G}' \mathbf{P}^{-1}$ . In Chap. 3, we described methods of estimating matrices  $\mathbf{K}_1 = [\mathbf{I}_1 - \mathbf{Q}_1]$ ,  $\mathbf{Q}_1 = \mathbf{P}_1^{-1} \Psi_1 (\Psi_1' \mathbf{P}_1' \Psi_1)^{-1} \Psi_1'$ ,  $\Psi_1' = \mathbf{U}' \mathbf{G}'_1$ ,  $\mathbf{K}_2 = [\mathbf{I}_2 - \mathbf{Q}_2]$ ,  $\mathbf{Q}_2 = \mathbf{P}^{-1} \Psi_2 (\Psi_2' \mathbf{P}^{-1} \Psi_2)^{-1} \Psi_2'$ , and  $\Psi_2' = \mathbf{U}' \mathbf{C}$ , which are used in this subsection.

At stage 1, the estimated phenotypic and genotypic covariance matrices were

$$\widehat{\mathbf{P}}_1 = \begin{bmatrix} 102 & 32 & 14 & 4 \\ 32 & 80 & 80 & 16 \\ 14 & 80 & 298 & 78 \\ 4 & 16 & 78 & 66 \end{bmatrix} \text{ and } \mathbf{G}_1 = \begin{bmatrix} 44 & 11 & -11 & -3 & -8 & -3 \\ 11 & 26 & 24 & 7 & 7 & 3 \\ -11 & 24 & 62 & 23 & 37 & 20 \\ -3 & 7 & 23 & 14 & 22 & 14 \end{bmatrix}$$

respectively. At both stages, traits  $y_1$  and  $y_2$  are restricted. Matrix  $\mathbf{U}$  can

be written as  $\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ , whence the estimated matrix of

restrictions was  $\widehat{\Psi}_1' = \mathbf{U} \widehat{\mathbf{G}}_1' = \begin{bmatrix} 44 & 11 & -11 & -3 \\ 11 & 26 & 24 & 7 \end{bmatrix}$ ; therefore, the estimated

matrices of  $\mathbf{Q}_1 = \mathbf{P}_1^{-1} \Psi_1 (\Psi_1' \mathbf{P}_1^{-1} \Psi_1)^{-1} \Psi_1'$  and  $\mathbf{K}_1 = [\mathbf{I}_4 - \mathbf{Q}_1]$  were

$$\widehat{\mathbf{Q}}_1 = \widehat{\mathbf{P}}_1^{-1} \widehat{\Psi}_1 (\widehat{\Psi}_1' \widehat{\mathbf{P}}_1^{-1} \widehat{\Psi}_1)^{-1} \widehat{\Psi}_1' = \begin{bmatrix} 0.923 & -0.013 & -0.511 & -0.144 \\ 0.164 & 1.026 & 1.093 & 0.317 \\ -0.145 & -0.069 & -0.001 & -0.001 \\ 0.010 & 0.159 & 0.178 & 0.052 \end{bmatrix} \text{ and}$$

$$\widehat{\mathbf{K}}_1 = [\mathbf{I}_4 - \widehat{\mathbf{Q}}_1] = \begin{bmatrix} 0.077 & 0.013 & 0.511 & 0.144 \\ 0.164 & -0.026 & -1.093 & -0.317 \\ 0.145 & 0.069 & 1.001 & 0.001 \\ -0.010 & -0.159 & -0.178 & 0.948 \end{bmatrix} \text{ respectively, where}$$

$\mathbf{I}_4$  is an identity matrix of size  $4 \times 4$ .

The estimated vector  $\mathbf{b}'_{R_1} = \mathbf{b}'_1 \mathbf{K}'_1$  was  $\widehat{\mathbf{b}}'_{R_1} = \widehat{\mathbf{b}}'_1 \widehat{\mathbf{K}}'_1 = [0.044 \ -0.095 \ 0.045 \ 0.131]$ , where  $\widehat{\mathbf{b}}'_1 = \mathbf{w}' \widehat{\mathbf{G}}_1' \widehat{\mathbf{P}}_1^{-1} = [-0.067 \ 0.125 \ 0.045 \ 0.167]$ , and  $\widehat{I}_{R_1} = \widehat{\mathbf{b}}'_{R_1} \mathbf{x}_1$  was the estimated MRLPSI at stage 1. The estimated MRLPSI vector of coefficients at stage 2 was  $\widehat{\mathbf{b}}'_{R_2} = \widehat{\mathbf{b}}'_2 \widehat{\mathbf{K}}'_2 = [0.045 \ -0.068 \ 0.028 \ -0.057 \ 0.099 \ 0.106]$  and  $\widehat{I}_{R_2} = \widehat{\mathbf{b}}'_{R_2} \mathbf{y}$  was the estimated MRLPSI at stage 2.

The estimated correlation value ( $\widehat{\rho}_{R_{12}}$ ) between  $\widehat{I}_{R_1} = \widehat{\mathbf{b}}'_{R_1} \mathbf{x}_1$  and  $\widehat{I}_{R_2} = \widehat{\mathbf{b}}'_{R_2} \mathbf{y}$  was  $\widehat{\rho}_{R_{12}} = \frac{\widehat{\mathbf{b}}'_{R_1} [\widehat{\mathbf{P}}_1 \ \widehat{\mathbf{P}}_{21}] \widehat{\mathbf{b}}_{R_2}}{\sqrt{\widehat{\mathbf{b}}'_{R_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{R_1}} \sqrt{\widehat{\mathbf{b}}'_{R_2} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{R_2}}} = 0.564$ , where  $\sqrt{\widehat{\mathbf{b}}'_{R_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{R_1}}$  and  $\sqrt{\widehat{\mathbf{b}}'_{R_2} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{R_2}}$  are

the estimated standard deviations of the variance of  $\widehat{I}_{R_1} = \widehat{\mathbf{b}}'_{R_1} \mathbf{x}_1$  and  $\widehat{I}_{R_2} = \widehat{\mathbf{b}}'_{R_2} \mathbf{y}$  respectively. According to Young (1964, Fig. 8), and Eqs. (4.10) and (4.11), the selection intensities for stages 1 and 2 were  $k_1 = 0.641$  and  $k_2 = 0.593$

respectively. The estimated selection responses and expected genetic gains per traits for both stages were  $\hat{R}_{R_1} = k_1 \sqrt{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_{R_1}} = 0.973$  and  $\hat{R}_{R_2} = k_2 \sqrt{\hat{\mathbf{b}}'_{R_2} \hat{\mathbf{P}}^* \hat{\mathbf{b}}_{R_2}} = 0.930$ ,  $\hat{\mathbf{E}}'_{R_1} = k_1 \frac{\hat{\mathbf{G}}'_1 \hat{\mathbf{b}}_{R_1}}{\sqrt{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_{R_1}}} = [0 \ 0 \ 1.271 \ 0.870 \ 1.482 \ 0.974]$  and  $\hat{\mathbf{E}}'_{R_2} = k_2 \frac{\hat{\mathbf{C}}^* \hat{\mathbf{b}}_{R_2}}{\sqrt{\hat{\mathbf{b}}'_{R_2} \hat{\mathbf{P}}^* \hat{\mathbf{b}}_{R_2}}} = [0 \ 0 \ 1.419 \ 1.014 \ 2.037 \ 1.349]$ , whereas  $\hat{R}_{R_1} + \hat{R}_{R_2} = 1.903$  and  $\hat{\mathbf{E}}'_{R_1} + \hat{\mathbf{E}}'_{R_2} = [0 \ 0 \ 2.691 \ 1.884 \ 3.519 \ 2.322]$  were the total estimated MRLPSI selection response and expected genetic gain per trait respectively.

Finally, the estimated MRLPSI accuracy at stage 1 was  $\hat{\rho}_{R_1} = \sqrt{\frac{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_{R_1}}{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}}} = 0.320$  and at stage 2 it was  $\hat{\rho}_{R_2} = \sqrt{\frac{\hat{\mathbf{b}}'_{R_2} \hat{\mathbf{P}}^* \hat{\mathbf{b}}_{R_2}}{\mathbf{w}' \hat{\mathbf{C}}^* \mathbf{w}}} = 0.334$ . In this case,  $\hat{\rho}_{R_2} > \hat{\rho}_{R_1}$ . We can explain these results considering that although  $\hat{\rho}_{R_2}$  was obtained with six traits,  $\hat{\rho}_{R_1}$  was obtained only with four traits, two of them restricted.

### 4.3 The Multistage Predetermined Proportional Gain Linear Phenotypic Selection Index

The main objectives of the multistage predetermined proportional gain linear phenotypic selection index (MPPG-LPSI) are the same as those of the predetermined proportional gain linear phenotypic selection index (PPG-LPSI) described in Chapter. i.e., to optimize, under some predetermined restrictions, the expected genetic gains per trait, to predict the net genetic merit, and to select the individual with the highest net genetic merit values as parents of the next generation under some predetermined restrictions. The MPPG-LPSI allows restrictions different from zero to be imposed on the expected genetic gains of some traits, whereas other traits increase (or decrease) their expected genetic gains without any restrictions being imposed.

#### 4.3.1 The MPPG-LPSI Parameters

In a similar manner to the MRLPSI, the MPPG-LPSI vector of coefficients for stages 1 and 2 should be a linear transformation of the MLPSI vector of coefficients at stages 1 ( $\mathbf{b}_1 = \mathbf{P}_1^{-1} \mathbf{G}_1 \mathbf{w}$ ) and 2 ( $\mathbf{b}_2 = \mathbf{P}_2^{-1} \mathbf{C} \mathbf{w}$ ), and should be written as



$$\mathbf{b}_{M_1} = \mathbf{K}_{M_1} \mathbf{b}_1 \quad (4.17)$$

and

$$\mathbf{b}_{M_2} = \mathbf{K}_{M_2} \mathbf{b}_2, \quad (4.18)$$

respectively, where, at stage 1,  $\mathbf{K}_{M_1} = [\mathbf{I}_1 - \mathbf{Q}_{M_1}]$ ,  $\mathbf{Q}_{M_1} = \mathbf{P}_1^{-1} \mathbf{M}_1 (\mathbf{M}'_1 \mathbf{P}_1^{-1} \mathbf{M}_1)^{-1} \mathbf{M}'_1$ ,  $\mathbf{M}'_1 = \mathbf{D}' \Psi'_1$ ,  $\Psi'_1 = \mathbf{U}' \mathbf{G}'_1$ ,  $\mathbf{I}_1$  is an identity matrix of the same size as  $\mathbf{P}_1$ , and  $\mathbf{P}_1^{-1}$  is the inverse of matrix  $\mathbf{P}_1$ . At stage 2,  $\mathbf{K}_M = [\mathbf{I} - \mathbf{Q}_M]$ ,  $\mathbf{Q}_M = \mathbf{P}^{-1} \mathbf{M} (\mathbf{M}' \mathbf{P}^{-1} \mathbf{M})^{-1} \mathbf{M}'$ ,  $\mathbf{M}' = \mathbf{D}' \Psi'$ ,  $\Psi' = \mathbf{U}' \mathbf{C}$ ,  $\mathbf{I}$  is an identity matrix of the same size as  $\mathbf{P}$ ,  $\mathbf{P}^{-1}$  is the inverse

of matrix  $\mathbf{P}$ , and  $\mathbf{D}' = \begin{bmatrix} d_r & 0 & \cdots & 0 & -d_1 \\ 0 & d_r & \cdots & 0 & -d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_r & -d_{r-1} \end{bmatrix}$ , where  $d_q$  ( $q = 1, 2, \dots, r$ ) is the  $q^{\text{th}}$

element of  $\mathbf{d}' = [d_1 \ d_2 \ \cdots \ d_r]$ , the vector PPG (predetermined proportional gains) imposed by the breeder (see Chap. 3 for details).

By Eqs. (4.17) and (4.18), the MPPG-LPSI for stages 1 and 2 can be written as  $I_{M_1} = \mathbf{b}_{M_1} \mathbf{x}_1$  and  $I_{M_2} = \mathbf{b}_{M_2} \mathbf{y}$  respectively, where, assuming that at stage 1 we select four traits and at stage 2 we select two traits,  $\mathbf{x}'_1 = [y_1 \ y_2 \ y_3 \ y_4]$  and  $\mathbf{x}'_2 = [y_5 \ y_6]$  are the vectors of phenotypic observations at stages 1 and 2 respectively, and  $\mathbf{y}' = [\mathbf{x}'_1 \ \mathbf{x}'_2]$  is the vector of total phenotypic observations at stage 2.

Let  $k_1$  and  $k_2$  be the selection intensities for stages 1 and 2 (Eqs. 4.10 and 4.11) respectively and let  $\mathbf{P}^*$  and  $\mathbf{C}^*$  be the adjusted matrices according to Eqs. (4.5) and (4.6) in the MPPG-LPSI context. Then, the MPPG-LPSI selection response and expected genetic gain per trait for both stages can be written as

$$R_{M_1} = k_1 \sqrt{\mathbf{b}'_{M_1} \mathbf{P}_1 \mathbf{b}_{M_1}} \quad \text{and} \quad R_{M_2} = k_2 \sqrt{\mathbf{b}'_{M_2} \mathbf{P}^* \mathbf{b}_{M_2}} \quad (4.19)$$

and

$$\mathbf{E}_{M_1} = k_1 \frac{\mathbf{G}'_1 \mathbf{b}_{M_1}}{\sqrt{\mathbf{b}'_{M_1} \mathbf{P}_1 \mathbf{b}_{M_1}}} \quad \text{and} \quad \mathbf{E}_{M_2} = k_2 \frac{\mathbf{b}'_{M_2} \mathbf{C}^*}{\sqrt{\mathbf{b}'_{M_2} \mathbf{P}^* \mathbf{b}_{M_2}}}, \quad (4.20)$$

respectively, whereas the total MPPG-LPSI selection response and expected genetic gain per trait for both stages are equal to  $R_{M_1} + R_{M_2}$  and  $\mathbf{E}_{M_1} + \mathbf{E}_{M_2}$ . In addition, the MPPG-LPSI accuracy for both stages can be written as

$$\rho_{M_1} = \sqrt{\frac{\mathbf{b}'_{M_1} \mathbf{P}_1 \mathbf{b}_{M_1}}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{M_2} = \sqrt{\frac{\mathbf{b}'_{M_2} \mathbf{P}^* \mathbf{b}_{M_2}}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}. \quad (4.21)$$



### 4.3.2 Numerical Examples

We use the real data set described in Sect. 4.2.2 to illustrate the theoretical results of the MPPG-LPSI in the same form as we did with those of the MRLPSI. We need to estimate vectors  $\mathbf{b}'_{M_1} = \mathbf{b}'_1 \mathbf{K}'_{M_1}$  and  $\mathbf{b}'_{M_2} = \mathbf{b}'_2 \mathbf{K}'_{M_2}$ , where  $\mathbf{b}'_1 = \mathbf{w}' \mathbf{G}'_1 \mathbf{P}_1^{-1}$  and  $\mathbf{b}'_2 = \mathbf{w}' \mathbf{G}' \mathbf{P}^{-1}$ . In Chap. 3 we have given methods to estimates  $\mathbf{K}_M = [\mathbf{I} - \mathbf{Q}_M]$ ,  $\mathbf{Q}_M = \mathbf{P}^{-1} \mathbf{M} (\mathbf{M}' \mathbf{P}^{-1} \mathbf{M})^{-1} \mathbf{M}'$ ,  $\mathbf{M}' = \mathbf{D}' \Psi'$ , and  $\Psi' = \mathbf{U}' \mathbf{C}$ , which will be used in this subsection.

The estimated phenotypic and genotypic covariance matrices at stage 1 were  $\widehat{\mathbf{P}}_1 = \begin{bmatrix} 102 & 32 & 14 & 4 \\ 32 & 80 & 80 & 16 \\ 14 & 80 & 298 & 78 \\ 4 & 16 & 78 & 66 \end{bmatrix}$  and  $\mathbf{G}_1 = \begin{bmatrix} 44 & 11 & -11 & -3 & -8 & -3 \\ 11 & 26 & 24 & 7 & 7 & 3 \\ -11 & 24 & 62 & 23 & 37 & 20 \\ -3 & 7 & 23 & 14 & 22 & 14 \end{bmatrix}$  respectively, whereas  $\mathbf{w}' = [0.08 \ 0.08 \ 0.38 \ 0.08 \ 0.08 \ 0.31]$  was the vector of economic weights. The traits restricted at both stages are  $y_1$ ,  $y_2$ , and  $y_3$ . The vector of PPG was  $\mathbf{d}' = [2 \ 3 \ 5]$ , whence  $\mathbf{D}' = \begin{bmatrix} 5 & 0 & -2 \\ 0 & 5 & -3 \end{bmatrix}$  and

$\mathbf{U}' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$  were matrices  $\mathbf{D}'$  and  $\mathbf{U}$ . The estimated matrices of  $\mathbf{M}'_1$  and  $\mathbf{K}_{M_1} = [\mathbf{I} - \mathbf{Q}_{M_1}]$  were  $\widehat{\mathbf{M}}'_1 = \mathbf{D}' \Psi'_1 = \begin{bmatrix} 242 & 7 & -178 & -61 \\ 88 & 58 & -66 & -34 \end{bmatrix}$  and  $\widehat{\mathbf{K}}_{M_1} = \begin{bmatrix} 0.176 & 0.205 & 0.606 & 0.159 \\ 0.031 & 0.032 & -0.007 & 0.199 \\ 0.195 & 0.235 & 0.852 & -0.098 \\ 0.130 & 0.130 & -0.098 & 0.940 \end{bmatrix}$  respectively, where  $\widehat{\Psi}'_1 = \mathbf{U}' \widehat{\mathbf{G}}'_1$ .

At stages 1 and 2, the estimated MPPG-LPSI vector of coefficients were  $\widehat{\mathbf{b}}'_{M_1} = \widehat{\mathbf{b}}'_1 \widehat{\mathbf{K}}'_{M_1} = [0.068 \ 0.035 \ 0.039 \ 0.160]$  and  $\widehat{\mathbf{b}}'_1 = \mathbf{w}' \widehat{\mathbf{G}}'_1 \widehat{\mathbf{P}}_1^{-1} = [-0.067 \ 0.125 \ 0.045 \ 0.167]$ , whence the estimated MPPG-LGSI were  $\widehat{I}_{M_1} = \widehat{\mathbf{b}}'_{M_1} \mathbf{x}_1$  and  $\widehat{I}_{M_2} = \widehat{\mathbf{b}}'_{M_2} \mathbf{y}$ . The estimated correlation value ( $\widehat{\rho}_{M_{12}}$ ) between  $\widehat{I}_{M_1} = \widehat{\mathbf{b}}'_{M_1} \mathbf{x}_1$  and  $\widehat{I}_{M_2} = \widehat{\mathbf{b}}'_{M_2} \mathbf{y}$  was  $\widehat{\rho}_{M_{12}} = \frac{\widehat{\mathbf{b}}'_{M_1} [\widehat{\mathbf{P}}_1 \ \widehat{\mathbf{P}}_{21}] \widehat{\mathbf{b}}_{M_2}}{\sqrt{\widehat{\mathbf{b}}'_{M_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{M_1}} \sqrt{\widehat{\mathbf{b}}'_{M_2} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{M_2}}}$  = 0.870, where  $\sqrt{\widehat{\mathbf{b}}'_{M_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{M_1}}$  and  $\sqrt{\widehat{\mathbf{b}}'_{M_2} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{M_2}}$  were the estimated standard deviations of variance of  $\widehat{I}_{M_1} = \widehat{\mathbf{b}}'_{M_1} \mathbf{x}_1$  and  $\widehat{I}_{M_2} = \widehat{\mathbf{b}}'_{M_2} \mathbf{y}$  respectively. According to Young (1964, Fig. 8), the selection intensities for stages 1 and 2 were  $k_1 = 0.744$  and  $k_2 = 0.721$  (Eqs. 4.10 and 2.11) respectively.

The estimated selection responses and expected genetic gains per traits for both stages were  $\widehat{R}_{M_1} = k_1 \sqrt{\widehat{\mathbf{b}}'_{M_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{M_1}} = 1.553$  and  $\widehat{R}_{M_2} = k_2 \sqrt{\widehat{\mathbf{b}}'_{M_2} \widehat{\mathbf{P}}^* \widehat{\mathbf{b}}_{M_2}} = 1.401$ ,  $\widehat{\mathbf{E}}'_{M_1} = k_1 \frac{\widehat{\mathbf{G}}'_1 \widehat{\mathbf{b}}_{M_1}}{\sqrt{\widehat{\mathbf{b}}'_{M_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{M_1}}} = [0.877 \ 1.316 \ 2.193 \ 1.128 \ 1.655 \ 1.037]$ , and



$\widehat{\mathbf{E}}'_{M_2} = k_2 \frac{\widehat{\mathbf{C}}^* \widehat{\mathbf{b}}_{M_2}}{\sqrt{\widehat{\mathbf{b}}'_{M_2} \widehat{\mathbf{P}}^* \widehat{\mathbf{b}}_{M_2}}} = [0.878 \ 1.346 \ 2.604 \ 1.433 \ 2.506 \ 1.602]$ , whereas  $\widehat{R}_{M_1} + \widehat{R}_{M_2} = 2.954$  and  $\widehat{\mathbf{E}}'_{M_1} + \widehat{\mathbf{E}}'_{M_2} = [1.755 \ 2.662 \ 4.797 \ 2.561 \ 4.161 \ 2.639]$  were the total estimated MPPGLPSI selection response and expected genetic gain per trait respectively. Note that the vector of predetermined restriction was  $\mathbf{d}' = [2 \ 3 \ 5]$ . This means that the MPPG-LPSI efficiency at predicting the total expected genetic gain per trait was high because the difference between each predetermined value (2, 3, and 5) and the total of each predicted value (1.755, 2.662, and 4.797) were 0.245, 0.338, and 0.203 respectively.

Finally, the estimated MPPG-LPSI accuracy at stage 1 was  $\widehat{\rho}_{M_1} = \sqrt{\frac{\widehat{\mathbf{b}}'_{M_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{M_1}}{\mathbf{w}' \widehat{\mathbf{C}} \mathbf{w}}}$  = 0.435, and at stage 2 it was  $\widehat{\rho}_{M_2} = \sqrt{\frac{\widehat{\mathbf{b}}'_{M_2} \widehat{\mathbf{P}}^* \widehat{\mathbf{b}}_{M_2}}{\mathbf{w}' \widehat{\mathbf{C}}^* \mathbf{w}}}$  = 0.428; that is, both were very similar.

## 4.4 The Multistage Linear Genomic Selection Index

We describe the multistage linear genomic selection indices (MLGSI) as an extension of the linear genomic selection index (LGSI, Chap. 5) theory to the multistage genomic selection context; thus, the theoretical results of the MLGSI are very similar to those of the LGSI. The MLGSI is a linear combination of genomic estimated breeding values (GEBVs) and is useful for predicting individual net genetic merit and for selecting individuals from a nonphenotyped testing population as parents of the next selection cycle.

### 4.4.1 The MLGSI Parameters

The objective of the MLGSI is to predict the net genetic merit  $H = \mathbf{w}'\mathbf{g}$ , where  $\mathbf{g}$  is a vector of true breeding values and  $\mathbf{w}'$  is the vector of economic weights, using only GEBVs. In Chap. 5, we indicated that the covariance between  $\gamma_i$  and  $\mathbf{g}_i$  is equal to the variance of  $\gamma_i$ , i.e.,  $Cov(\mathbf{g}_i, \gamma_i) = s_i^2$ , and that the GEBV associated with the  $i$ th trait is a predictor of the  $i$ th vector of genomic breeding values ( $\gamma_i$ ). In the testing population, the only observable information is  $\mathbf{w}'$  and the GEBV associated with the traits of interest. For this reason, in practice, we construct a linear combination of GEBVs, which should be a good predictor of  $H = \mathbf{w}'\mathbf{g}$ .

Suppose that the breeder is interested in four traits, and that  $\boldsymbol{\gamma}' = [\gamma_1 \ \gamma_2 \ \gamma_3 \ \gamma_4]$ ,  $\mathbf{g}' = [g_1 \ g_2 \ g_3 \ g_4]$ , and  $\mathbf{w}' = [w_1 \ w_2 \ w_3 \ w_4]$  are the vectors of genomic breeding values ( $\boldsymbol{\gamma}$ ), true breeding values ( $\mathbf{g}$ ), and



economic weights ( $\mathbf{w}$ ) respectively. Let  $\boldsymbol{\Gamma} = \text{Var}(\boldsymbol{\gamma}) = \begin{bmatrix} s_1^2 & s_{12} & s_{13} & s_{14} \\ s_{21} & s_2^2 & s_{23} & s_{24} \\ s_{31} & s_{32} & s_3^2 & s_{34} \\ s_{41} & s_{42} & s_{43} & s_4^2 \end{bmatrix}$  and

$\mathbf{C} = (\mathbf{g}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$  be the covariance matrix of  $\mathbf{g}$  and  $\boldsymbol{\gamma}$ . At a

two-stage selection breeding scheme,  $\boldsymbol{\gamma}' = [\gamma_1 \ \gamma_2 \ \gamma_3 \ \gamma_4]$  can be partitioned into  $\boldsymbol{\gamma}'_1 = [\gamma_1 \ \gamma_2]$  and  $\boldsymbol{\gamma}'_2 = [\gamma_3 \ \gamma_4]$ ; therefore, at stage 1,  $\boldsymbol{\Gamma}_1 = \text{Var}(\boldsymbol{\gamma}_1) = \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix}$  is the genomic covariance matrix of  $\boldsymbol{\gamma}'_1 = [\gamma_1 \ \gamma_2]$  and  $\text{Cov}(\boldsymbol{\gamma}_1, \mathbf{g}) = \begin{bmatrix} s_1^2 & s_{12} & s_{13} & s_{14} \\ s_{12} & s_2^2 & s_{23} & s_{24} \end{bmatrix} = \mathbf{A}_1$  is the covariance matrix of  $\boldsymbol{\gamma}'_1 = [\gamma_1 \ \gamma_2]$  with  $\mathbf{g}' = [g_1 \ g_2 \ g_3 \ g_4]$ . Matrix  $\mathbf{A}_1$  indicates that we are assuming that the covariance between  $\boldsymbol{\gamma}_i$  and  $\mathbf{g}_j$  ( $i, j = 1, 2, \dots, g$ ;  $g$  = number of genotypes) is equal to the covariance between  $\boldsymbol{\gamma}_i$  and  $\boldsymbol{\gamma}_j$ . This is because, in practice, in the testing population, we can only estimate matrix  $\boldsymbol{\Gamma}$ .

At stage 2,  $\boldsymbol{\Gamma} = \text{Var}(\boldsymbol{\gamma})$  is the covariance matrix of  $\boldsymbol{\gamma}$  and  $\mathbf{A} = \boldsymbol{\Gamma}$  is the covariance matrix of the vector of genomic breeding values  $\boldsymbol{\gamma}$  with the vector of breeding values  $\mathbf{g}$ . The MLGSI vector of coefficients at stages 1 and 2 are  $\boldsymbol{\beta}'_1 = \mathbf{w}'\mathbf{A}'_1\boldsymbol{\Gamma}_1^{-1} = [\beta_{11} \ \beta_{12}]$  and  $\boldsymbol{\beta}'_2 = \mathbf{w}'\mathbf{A}\boldsymbol{\Gamma}^{-1} = \mathbf{w}' = [w_1 \ w_2 \ w_3 \ w_4]$  respectively, and the MLGSI for both stages can be written as  $I_1 = \beta_{11}\gamma_1 + \beta_{12}\gamma_2 = \boldsymbol{\beta}'_1 \boldsymbol{\gamma}_1$  and  $I_2 = w_1\gamma_1 + w_2\gamma_2 + w_3\gamma_3 + w_4\gamma_4 = \mathbf{w}'\boldsymbol{\gamma}$ .

Let  $k_1$  and  $k_2$  be the MLGSI selection intensities for stages 1 and 2. For both stages, the MLGSI accuracies ( $\rho_{H1_1}$  and  $\rho_{H1_2}$ ), expected genetic gains per trait ( $\mathbf{E}_1$  and  $\mathbf{E}_2$ ) and selection responses ( $R_1$  and  $R_2$ ) can be written as

$$\rho_{H1_1} = \sqrt{\frac{\boldsymbol{\beta}'_1 \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_1}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{H1_2} = \sqrt{\frac{\mathbf{w}' \boldsymbol{\Gamma}^* \mathbf{w}}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}, \quad (4.22)$$

$$\mathbf{E}_1 = k_1 \frac{\mathbf{A}'_1 \boldsymbol{\beta}_1}{\sqrt{\boldsymbol{\beta}'_1 \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_1}} \quad \text{and} \quad \mathbf{E}_2 = k_2 \frac{\boldsymbol{\Gamma}^* \mathbf{w}}{\sqrt{\mathbf{w}' \boldsymbol{\Gamma}^* \mathbf{w}}} \quad (4.23)$$

and

$$R_1 = k_1 \sqrt{\boldsymbol{\beta}'_1 \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_1} \quad \text{and} \quad R_2 = k_2 \sqrt{\mathbf{w}' \boldsymbol{\Gamma}^* \mathbf{w}}. \quad (4.24)$$

The total MLGSI expected genetic gain per trait and selection response at both stages are equal to  $\mathbf{E}_1 + \mathbf{E}_2$  and  $R_1 + R_2$ . To simplify notation, in Eqs. (4.23) and (4.24), we have omitted the intervals between stages or selection cycles ( $L_G$ ). Matrices  $\mathbf{C}^*$  and  $\boldsymbol{\Gamma}^*$  in Eqs. (4.22) to (4.23) are matrices  $\boldsymbol{\Gamma}$  and  $\mathbf{C}$  adjusted for previous selection on  $I_1$ .

We adjust matrices  $\boldsymbol{\Gamma}$  and  $\mathbf{C}$  for previous selection on  $I_1$  as



$$\boldsymbol{\Gamma}^* = \boldsymbol{\Gamma} - u \frac{\mathbf{A}'_1 \boldsymbol{\beta}_1 \boldsymbol{\beta}'_1 \mathbf{A}_1}{\boldsymbol{\beta}'_1 \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_1} \quad (4.25)$$

and

$$\mathbf{C}^* = \mathbf{C} - u \frac{\mathbf{G}'_1 \mathbf{b}_1 \mathbf{b}'_1 \mathbf{G}_1}{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}, \quad (4.26)$$

respectively, where  $u = k_1(k_1 - \tau)$ ,  $k_1$  is the standardized selection differential, and  $\tau$  is the truncation point when  $I_1 = \boldsymbol{\beta}'_1 \boldsymbol{\gamma}_1$  is applied. All the terms in Eq. (4.26) were defined in Eq. (4.6).

The correlation between  $I_1 = \boldsymbol{\beta}'_1 \boldsymbol{\gamma}_1$  and  $I_2 = \mathbf{w}' \boldsymbol{\gamma}$  can be written as

$$\text{Corr}(I_1, I_2) = \frac{\boldsymbol{\beta}'_1 \mathbf{A}_1 \mathbf{w}}{\sqrt{\boldsymbol{\beta}'_1 \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_1} \sqrt{\mathbf{w}' \boldsymbol{\Gamma} \mathbf{w}}} = \rho_{I_1 I_2}, \quad (4.27)$$

where  $\sqrt{\boldsymbol{\beta}'_1 \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_1}$  and  $\sqrt{\mathbf{w}' \boldsymbol{\Gamma} \mathbf{w}}$  are the standard deviations of the variances of  $I_1 = \boldsymbol{\beta}'_1 \boldsymbol{\gamma}_1$  and  $I_2 = \mathbf{w}' \boldsymbol{\gamma}$  respectively. In Eq. (4.27), matrix  $\boldsymbol{\Gamma}$  was not adjusted according to Eq. (4.25).

#### 4.4.2 Estimating the Genomic Covariance Matrix

All the MLGSI parameters are associated with matrix  $\boldsymbol{\Gamma}$ ; thus, the estimation of this matrix in the testing population is very important. We estimate matrix  $\boldsymbol{\Gamma}$  according to the estimation method described in Chapter. that is, as

$$\hat{\boldsymbol{\Gamma}}_l = \left\{ \hat{\sigma}_{\gamma_{qq'}} \right\}, \quad (4.28)$$

where  $\hat{\sigma}_{\gamma_{qq'}} = \frac{1}{g} (\hat{\boldsymbol{\gamma}}_{ql} - \mathbf{1}\hat{\mu}_{\gamma_{ql}})' \mathbf{G}_l^{-1} (\hat{\boldsymbol{\gamma}}_{q'l} - \mathbf{1}\hat{\mu}_{\gamma_{q'l}})$  is the estimated covariance between  $\hat{\boldsymbol{\gamma}}_{ql} = \mathbf{X}_l \hat{\mathbf{u}}_q$  and  $\hat{\boldsymbol{\gamma}}_{q'l} = \mathbf{X}_l \hat{\mathbf{u}}_{q'}$  at stage  $l$  or selection cycle of the testing population;  $g$  is the number of genotypes;  $\hat{\mu}_{\gamma_{ql}}$  and  $\hat{\mu}_{\gamma_{q'l}}$  are the estimated arithmetic means of the values of  $\hat{\boldsymbol{\gamma}}_{ql}$  and  $\hat{\boldsymbol{\gamma}}_{q'l}$ ;  $\mathbf{1}$  is an  $g \times 1$  vector of 1s and  $\mathbf{G}_l = c^{-1} \mathbf{X}_l \mathbf{X}'_l$  is the additive genomic relationship matrix at stage  $l$  or selection cycle in the testing population.

#### 4.4.3 Numerical Examples

We illustrate the MLGSI theoretical results using the data described. simulated for eight phenotypic and seven genomic selection cycles,



each with four traits ( $T_1$ ,  $T_2$ ,  $T_3$  and  $T_4$ ), 500 genotypes, four replicates for each genotype, 2500 molecular markers, and 315 quantitative trait loci in one environment. The economic weights of  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$  were 1, -1, 1, and 1 respectively. In this subsection, and only for illustrative purposes, we use the data set from cycle 1.

The genotypic and genomic estimated covariance matrices in cycle 1 were  $\hat{\mathbf{C}} = \begin{bmatrix} 36.21 & -12.93 & 8.35 & 2.74 \\ -12.93 & 13.04 & -3.4 & -2.24 \\ 8.35 & -3.4 & 9.96 & 0.16 \\ 2.74 & -2.24 & 0.16 & 6.64 \end{bmatrix}$  and  $\hat{\mathbf{\Gamma}} = \begin{bmatrix} 16.26 & -6.51 & 5.60 & 2.29 \\ -6.51 & 5.79 & -2.23 & -1.62 \\ 5.60 & -2.23 & 3.75 & 0.94 \\ 2.29 & -1.62 & 0.94 & 2.62 \end{bmatrix}$

respectively, whereas  $\mathbf{w}' = [1 \ -1 \ 1 \ 1]$  was the vector of economic weights. Matrices  $\hat{\mathbf{P}}$  and  $\hat{\mathbf{C}}$  were obtained according to Eqs. (4.22) to (4.24), whereas matrix  $\hat{\mathbf{\Gamma}}$  was obtained according to Eq. (4.28).

Suppose that we select two traits at stages 1 and 2. Then, at stage 1,  $\hat{\mathbf{\Gamma}}_1 = \begin{bmatrix} 16.26 & -6.51 \\ -6.51 & 5.79 \end{bmatrix}$  and  $\hat{\mathbf{A}}_1 = \begin{bmatrix} 16.26 & -6.51 & 5.60 & 2.29 \\ -6.51 & 5.79 & -2.33 & -1.62 \end{bmatrix}$  are the estimated covariance matrices of  $\mathbf{\Gamma}_1$  and  $\mathbf{A}_1$  respectively, and the estimated MLGSI vector of coefficients was  $\hat{\boldsymbol{\beta}}'_1 = \mathbf{w}' \hat{\mathbf{A}}_1' \hat{\mathbf{\Gamma}}_1^{-1} = [1.39 \ -1.25]$ . Because at stage 2  $\hat{\boldsymbol{\beta}}'_2 = \mathbf{w}' \hat{\mathbf{A}} \hat{\mathbf{\Gamma}}^{-1} = \mathbf{w}' = [w_1 \ w_2 \ w_3 \ w_4]$ , the estimated MLGSI vector of coefficients is the vector of economic weights. Thus,  $\hat{\rho}_{I_1 I_2} = \frac{\hat{\boldsymbol{\beta}}'_1 \hat{\mathbf{A}}_1 \mathbf{w}}{\sqrt{\hat{\boldsymbol{\beta}}'_1 \hat{\mathbf{\Gamma}}_1 \hat{\boldsymbol{\beta}}_1} \sqrt{\mathbf{w}' \hat{\mathbf{\Gamma}} \mathbf{w}}} =$

0.97 was the estimated correlation between  $\hat{I}_1 = \hat{\boldsymbol{\beta}}'_1 \hat{\boldsymbol{\gamma}}_1$  and  $\hat{I}_2 = \mathbf{w}' \hat{\boldsymbol{\gamma}}$ , and assuming that the fixed proportion was 0.2 (20%),  $k_1 = 0.744$  and  $k_2 = 0.721$  were the approximated selection intensities for stages 1 and 2 respectively. The adjusted matrices  $\mathbf{\Gamma}^*$  and  $\mathbf{C}^*$  for previous selection on  $\hat{I}_1 = \hat{\boldsymbol{\beta}}'_1 \hat{\boldsymbol{\gamma}}_1$  were  $\hat{\mathbf{\Gamma}}^* = \begin{bmatrix} 7.96 & -2.11 & 2.71 & 0.88 \\ -2.11 & 3.46 & -0.80 & -0.87 \\ 2.71 & -0.80 & 2.75 & 0.45 \\ 0.88 & -0.87 & 0.45 & 2.38 \end{bmatrix}$  and  $\hat{\mathbf{C}}^* = \begin{bmatrix} 24.40 & -5.65 & 5.47 & 1.39 \\ -5.65 & 8.55 & -1.63 & -1.41 \\ 5.47 & -1.63 & 9.26 & -0.17 \\ 1.39 & -1.41 & -0.17 & 6.49 \end{bmatrix}$ .

The estimated MLGSI accuracy, selection response, and expected genetic gain for stage 1 in the testing population were  $\hat{\rho}_{HI_1} = \sqrt{\frac{\hat{\boldsymbol{\beta}}'_1 \hat{\mathbf{\Gamma}}_1 \hat{\boldsymbol{\beta}}_1}{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}}} = 0.71$ ,

$\hat{R}_1 = k_1 \sqrt{\hat{\boldsymbol{\beta}}'_1 \hat{\mathbf{\Gamma}}_1 \hat{\boldsymbol{\beta}}_1} = 5.90$ , and  $\hat{\mathbf{E}}'_1 = k_1 \frac{\hat{\mathbf{A}}_1' \hat{\boldsymbol{\beta}}_1}{\sqrt{\hat{\boldsymbol{\beta}}'_1 \hat{\mathbf{\Gamma}}_1 \hat{\boldsymbol{\beta}}_1}} = [2.88 \ -1.53 \ 1.00 \ 0.49]$

respectively, whereas at stage 2, the estimated MLGSI accuracy, selection response, and expected genetic gain were  $\hat{\rho}_{HI_2} = \sqrt{\frac{\mathbf{w}' \hat{\mathbf{\Gamma}}^* \mathbf{w}}{\mathbf{w}' \hat{\mathbf{C}}^* \mathbf{w}}} = 0.64$ ,  $\hat{R}_2 = k_2 \sqrt{\mathbf{w}' \hat{\mathbf{\Gamma}}^* \mathbf{w}} = 4.10$ ,

and  $\hat{\mathbf{E}}'_2 = k_2 \frac{\hat{\mathbf{\Gamma}}^* \mathbf{w}}{\sqrt{\mathbf{w}' \hat{\mathbf{\Gamma}}^* \mathbf{w}}} = [1.74 \ -0.92 \ 0.85 \ 0.58]$  respectively. The estimated

MLGSI accuracy, selection response, and expected genetic gain at stage 2 were



lower than at stage 1. This means that the adjusted matrices  $\widehat{\Gamma}^*$  and  $\widehat{\mathbf{C}}^*$  negatively affected the estimated MLPSI parameters at stage 2. The total estimated MLGSI selection response and expected genetic gain for stages 1 and 2 were  $\widehat{R}_1 + \widehat{R}_2 = 9.99$  and  $\widehat{\mathbf{E}}'_1 + \widehat{\mathbf{E}}'_2 = [4.62 \ -2.45 \ 1.85 \ 1.07]$ .

## 4.5 The Multistage Restricted Linear Genomic Selection Index (MRLGSI)

The restricted linear genomic selection index (RLGSI) described in Chap. 3 is extended to the multistage restricted linear genomic selection index (MRLGSI) context in a two-stage breeding selection scheme.

### 4.5.1 The MRLGSI Parameters

In Sect. 4.4.1, we indicated that the MLGSI vector of coefficients at stage 1 can be written as  $\boldsymbol{\beta}'_1 = \mathbf{w}'\mathbf{A}'_1\boldsymbol{\Gamma}_1^{-1} = [\beta_{11} \ \beta_{12}]$  and at stage 2 as  $\boldsymbol{\beta}'_2 = \mathbf{w}'\mathbf{A}\boldsymbol{\Gamma}^{-1} = \mathbf{w}' = [w_1 \ w_2 \ w_3 \ w_4]$ . It can be shown that the MRLGSI vector of coefficients is a linear transformation of vectors  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  made by matrix  $\mathbf{K}_G$ , which is a projector (see Chaps. 1 for details) that projects  $\boldsymbol{\beta}$  and  $\boldsymbol{\beta}_1$  into a space smaller than the original space of  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$ . Thus, at stages 1 and 2, the MRLGSI vector of coefficients is

$$\boldsymbol{\beta}_{R_1} = \mathbf{K}_{G_1}\boldsymbol{\beta}_1 \quad (4.29)$$

and

$$\boldsymbol{\beta}_{R_2} = \mathbf{K}_{G_2}\boldsymbol{\beta}_2 = \mathbf{K}_{G_2}\mathbf{w}, \quad (4.30)$$

respectively, where  $\mathbf{K}_{G_1} = [\mathbf{I} - \mathbf{Q}_{G_1}]$ ,  $\mathbf{Q}_{G_1} = \mathbf{U}_1(\mathbf{U}'_1\boldsymbol{\Gamma}_1\mathbf{U}_1)^{-1}\mathbf{U}'_1\boldsymbol{\Gamma}_1$ ,  $\mathbf{K}_{G_2} = [\mathbf{I} - \mathbf{Q}_{G_2}]$ , and  $\mathbf{Q}_{G_2} = \mathbf{U}_2(\mathbf{U}'_2\boldsymbol{\Gamma}\mathbf{U}_2)^{-1}\mathbf{U}'_2\boldsymbol{\Gamma}$  are matrix projectors. By Eqs. (4.29) and (4.30), the MRLGSI at stages 1 and 2 can be written as  $I_{R_1} = \boldsymbol{\beta}'_{R_1}\boldsymbol{\gamma}_1$  and  $I_{R_2} = \boldsymbol{\beta}'_{R_2}\boldsymbol{\gamma}$  respectively, where  $\boldsymbol{\gamma}'_1 = [\gamma_1 \ \gamma_2]$  and  $\boldsymbol{\gamma}' = [\gamma_1 \ \gamma_2 \ \gamma_3 \ \gamma_4]$  are vectors of genomic breeding values, which can be estimated using GEBVs, as described in Chap. 5. In Chap. 1 we described methods for constructing matrix  $\mathbf{U}'$  and estimating matrix  $\mathbf{K}_G$ ; those methods are also valid in the MRLGSI context.

In a similar manner to the MLGSI context, MRLGSI accuracies, expected genetic gains per trait, and selection responses for stages 1 and 2 in the testing population can be written as



$$\rho_{HI_1} = \sqrt{\frac{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{HI_2} = \sqrt{\frac{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}, \quad (4.31)$$

$$\mathbf{E}_{R_1} = k_1 \frac{\mathbf{A}'_1 \boldsymbol{\beta}_{R_1}}{\sqrt{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}}} \quad \text{and} \quad \mathbf{E}_{R_2} = k_2 \frac{\boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}{\sqrt{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}} \quad (4.32)$$

and

$$R_{R_1} = k_1 \sqrt{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}} \quad \text{and} \quad R_{R_2} = k_2 \sqrt{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}, \quad (4.33)$$

respectively. The total MRLGSI expected genetic gain per trait and selection response for both stages are equal to  $\mathbf{E}_{R_1} + \mathbf{E}_{R_2}$  and  $R_{R_1} + R_{R_2}$ . To simplify the notation, in Eqs. (4.32) and (4.33), we have omitted the intervals between stages or selection cycles ( $L_G$ ). Matrices  $\boldsymbol{\Gamma}^*$  and  $\mathbf{C}^*$  in Eqs. (4.31) to (4.33) are matrices  $\boldsymbol{\Gamma}$  and  $\mathbf{C}$  adjusted for previous selection.

In the MRLGSI context, matrices  $\boldsymbol{\Gamma}^*$  and  $\mathbf{C}^*$  can be obtained as

$$\boldsymbol{\Gamma}^* = \boldsymbol{\Gamma} - u \frac{\mathbf{A}'_1 \boldsymbol{\beta}_{R_1} \boldsymbol{\beta}'_{R_1} \mathbf{A}_1}{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}} \quad (4.34)$$

and

$$\mathbf{C}^* = \mathbf{C} - u \frac{\mathbf{G}'_1 \mathbf{b}_{R_1} \mathbf{b}'_{R_1} \mathbf{G}_1}{\mathbf{b}'_{R_1} \mathbf{P}_1 \mathbf{b}_{R_1}}, \quad (4.35)$$

where  $\boldsymbol{\beta}_{R_1}$  was defined in Eq. (4.29) and vector  $\mathbf{b}_{R_1}$  can be obtained according to the RLPSI as described in Chapter. The term  $u = k(k - \tau)$  was defined earlier.

The correlation between  $I_{R_1} = \boldsymbol{\beta}'_{R_1} \boldsymbol{\gamma}_1$  and  $I_{R_2} = \boldsymbol{\beta}'_{R_2} \boldsymbol{\gamma}$  can be written as

$$\rho_{I_{R_1} I_{R_2}} = \frac{\boldsymbol{\beta}'_{R_1} \mathbf{A}_1 \boldsymbol{\beta}_{R_2}}{\sqrt{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}} \sqrt{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma} \boldsymbol{\beta}_{R_2}}}, \quad (4.36)$$

where  $\sqrt{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}}$  and  $\sqrt{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma} \boldsymbol{\beta}_{R_2}}$  are the standard deviations of the variances of  $I_{R_1} = \boldsymbol{\beta}'_{R_1} \boldsymbol{\gamma}_1$  and  $I_{R_2} = \boldsymbol{\beta}'_{R_2} \boldsymbol{\gamma}$  respectively. In Eq. (4.36), matrix  $\boldsymbol{\Gamma}$  was not adjusted for previous selection on  $I_{R_1} = \boldsymbol{\beta}'_{R_1} \boldsymbol{\gamma}_1$ .

#### 4.5.2 Numerical Examples

To illustrate the MRLGSI theory in a two-stage breeding selection scheme, we use the simulated data described in Sect. 4.4.3. In that subsection we indicated that the



estimated covariance matrices of  $\boldsymbol{\Gamma}_1$  and  $\mathbf{A}_1$  were  $\widehat{\boldsymbol{\Gamma}}_1 = \begin{bmatrix} 16.26 & -6.51 \\ -6.51 & 5.79 \end{bmatrix}$  and  $\widehat{\mathbf{A}}_1 = \begin{bmatrix} 16.26 & -6.51 & 5.60 & 2.29 \\ -6.51 & 5.79 & -2.33 & -1.62 \end{bmatrix}$ , and that  $\widehat{\boldsymbol{\beta}}'_1 = \mathbf{w}' \widehat{\mathbf{A}}'_1 \widehat{\boldsymbol{\Gamma}}_1^{-1} = [1.39 \quad -1.25]$  was the estimated MLGSI vector of coefficients at stage 1. At stage 2, the estimated MLGSI vector of coefficients was  $\mathbf{w}' = [1 \quad -1 \quad 1 \quad 1]$ , the vector of economic weights.

Suppose that we restrict only trait 2; then at stages 1 and 2, matrix  $\mathbf{U}'_1 = [0 \quad 1]$  and matrix  $\mathbf{U}'_2 = [0 \quad 1 \quad 0 \quad 0]$  respectively. In addition,  $\widehat{\mathbf{Q}}_{G_1} = \mathbf{U}_1 (\mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1$ ,  $\widehat{\mathbf{Q}}_{G_2} = \mathbf{U}_2 (\mathbf{U}'_2 \widehat{\boldsymbol{\Gamma}} \mathbf{U}_2)^{-1} \mathbf{U}'_2 \widehat{\boldsymbol{\Gamma}}$ ,  $\widehat{\mathbf{K}}_{G_1} = [\mathbf{I} - \widehat{\mathbf{Q}}_{G_1}]$ , and  $\widehat{\mathbf{K}}_{G_2} = [\mathbf{I} - \widehat{\mathbf{Q}}_{G_2}]$  are the estimated matrices described in Eqs. (4.29) and (4.30) for stages 1 and 2. It can be shown that, at stages 1 and 2,  $\widehat{\boldsymbol{\beta}}'_{R_1} = \widehat{\boldsymbol{\beta}}'_1 \widehat{\mathbf{K}}'_{G_1} = [1.39 \quad 1.558]$  and  $\widehat{\boldsymbol{\beta}}'_{R_2} = \mathbf{w}' \widehat{\mathbf{K}}'_{G_2} = [1.0 \quad 1.81 \quad 1.01.0]$  are the MRLGSI vectors of coefficients respectively.

Suppose that the total proportion retained for the two stages was 20%, then at stage 1,  $k_1 = 0.744$  is an associated approximated selection intensity and the estimated MRLGSI selection response, expected genetic gain per trait, and accuracy were  $\widehat{R}_{R_1} = k_1 \sqrt{\widehat{\boldsymbol{\beta}}'_{R_1} \widehat{\boldsymbol{\Gamma}}_1 \widehat{\boldsymbol{\beta}}_{R_1}} = 3.083$ ,  $\widehat{\mathbf{E}}_{R_1} = [2.225 \quad 0 \quad 0.742 \quad 0.117]$ , and  $\widehat{\rho}_{HI_1} = \sqrt{\frac{\widehat{\boldsymbol{\beta}}'_{R_1} \widehat{\boldsymbol{\Gamma}}_1 \widehat{\boldsymbol{\beta}}_{R_1}}{\mathbf{w}' \widehat{\mathbf{C}} \mathbf{w}}} = 0.370$  respectively. The estimated MRLGSI expected genetic gain, accuracy, and selection response at stage 2 were  $\widehat{\mathbf{E}}_{R_2} = k_2 \frac{\widehat{\boldsymbol{\beta}}'_{R_2} \widehat{\boldsymbol{\Gamma}}^*}{\sqrt{\widehat{\boldsymbol{\beta}}'_{R_2} \widehat{\boldsymbol{\Gamma}}^* \widehat{\boldsymbol{\beta}}_{R_2}}} = [1.156 \quad 0 \quad 0.793 \quad 0.536]$ ,  $\widehat{\rho}_{HI_2} = \sqrt{\frac{\widehat{\boldsymbol{\beta}}'_{R_2} \widehat{\boldsymbol{\Gamma}}^* \widehat{\boldsymbol{\beta}}_{R_2}}{\mathbf{w}' \widehat{\mathbf{C}}^* \mathbf{w}}} = 0.32$ , and  $\widehat{R}_{R_2} = k_2 \sqrt{\widehat{\boldsymbol{\beta}}'_{R_2} \widehat{\boldsymbol{\Gamma}}^* \widehat{\boldsymbol{\beta}}_{R_2}} = 2.485$  respectively, where  $k_2 = 0.721$  was the approximated selection intensity value for stage 2.

The estimated total MRLGSI selection response and expected genetic gain at stages 1 and 2 were  $\widehat{R}_{R_1} + \widehat{R}_{R_2} = 5.568$  and  $\mathbf{E}'_{R_1} + \mathbf{E}'_{R_2} = [3.380 \quad 0 \quad 1.535 \quad 0.653]$  respectively. Note that, in effect, the expected genetic gain for trait 2 was 0, as expected.

## 4.6 The Multistage Predetermined Proportional Gain Linear Genomic Selection Index

The MPPG-LGSI is an adaptation of the predetermined proportional gain linear genomic selection index (PPG-LGSI) described in Chap. 6; thus, the theoretical results, properties, and objectives of both indices are similar. The MPPG-LGSI objective is to change  $\mu_q$  to  $\mu_q + d_q$ , where  $d_q$  is a predetermined change in  $\mu_q$ . We

solve this problem by minimizing the mean squared difference between  $I = \beta' \gamma$  and  $H = \mathbf{w}' \mathbf{g} (E[(H - I)^2])$  under the restriction  $\mathbf{U}' \Gamma \beta = \theta_G \mathbf{d}$ , where  $\theta_G$  is a proportionality constant,  $\mathbf{d}' = [d_1 \ d_2 \dots d_t]$  is the vector of predetermined restrictions,  $\mathbf{U}'$  is a matrix  $(t - 1) \times t$  of 1s and 0s, and  $\Gamma$  is a covariance matrix of additive genomic breeding values,  $\gamma' = [\gamma_1 \ \gamma_2 \dots \gamma_t]$ , where  $r$  is the number of predetermined restrictions and  $t$  the number of traits.

#### 4.6.1 The MPPG-LGSI Parameters

According to the results in Chap. 6, at stages 1 and 2, the MPPG-LGSI vector of coefficients can be written as

$$\beta_{P_1} = \beta_{R_1} + \theta_1 \mathbf{U}_1 (\mathbf{U}'_1 \Gamma_1 \mathbf{U}_1)^{-1} \mathbf{d} \quad (4.37)$$

and

$$\beta_{P_2} = \beta_{R_2} + \theta_2 \mathbf{U}_2 (\mathbf{U}'_2 \Gamma \mathbf{U}_2)^{-1} \mathbf{d}, \quad (4.38)$$

respectively, where  $\beta_{R_1} = \mathbf{K}_{G_1} \beta_1$ ,  $\beta_{R_2} = \mathbf{K}_{G_2} \beta_2 = \mathbf{K}_{G_2} \mathbf{w}$ ,  $\mathbf{K}_{G_1} = [\mathbf{I} - \mathbf{Q}_{G_1}]$ ,  $\mathbf{Q}_{G_1} = \mathbf{U}_1 (\mathbf{U}'_1 \Gamma_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 \Gamma_1$ ,  $\mathbf{K}_{G_2} = [\mathbf{I} - \mathbf{Q}_{G_2}]$ , and  $\mathbf{Q}_{G_2} = \mathbf{U}_2 (\mathbf{U}'_2 \Gamma \mathbf{U}_2)^{-1} \mathbf{U}'_2 \Gamma$  were described in Eqs. (4.29) and (4.30). Also, it can be shown that the proportionality constants for stages 1 ( $\theta_1$ ) and 2 ( $\theta_2$ ) are

$$\theta_1 = \frac{\mathbf{d}' (\mathbf{U}'_1 \Gamma_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 \mathbf{A}_1 \mathbf{w}}{\mathbf{d}' (\mathbf{U}'_1 \Gamma_1 \mathbf{U}_1)^{-1} \mathbf{d}} \quad \text{and} \quad \theta_2 = \frac{\mathbf{d}' (\mathbf{U}'_2 \Gamma \mathbf{U}_2)^{-1} \mathbf{U}'_2 \Gamma \mathbf{w}}{\mathbf{d}' (\mathbf{U}'_2 \Gamma \mathbf{U}_2)^{-1} \mathbf{d}}, \quad (4.39)$$

respectively. By Eqs. (4.37) to (4.39), the MPPG-LGSI for stages 1 and 2 can be written as  $I_{P_1} = \beta'_{P_1} \gamma_1$  and  $I_{P_2} = \beta'_{P_2} \gamma$  respectively, where  $\gamma_1$  and  $\gamma$  are vectors of genomic breeding values, which can be estimated using GEBVs (see Chap. 5 for details).

For stages 1 and 2, the MPPG-LGSI accuracies ( $\rho_{HI_1}$  and  $\rho_{HI_2}$ ), expected genetic gains per trait ( $E_{P_1}$  and  $E_{P_2}$ ), and selection responses ( $R_{P_1}$  and  $R_{P_2}$ ) can be written as

$$\rho_{HI_1} = \sqrt{\frac{\beta'_{P_1} \Gamma_1 \beta_{P_1}}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{HI_2} = \sqrt{\frac{\beta'_{P_2} \Gamma^* \beta_{P_2}}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}, \quad (4.40)$$

$$E_{P_1} = k_1 \frac{\mathbf{A}'_1 \beta_{P_1}}{\sqrt{\beta'_{P_1} \Gamma_1 \beta_{P_1}}} \quad \text{and} \quad E_{P_2} = k_2 \frac{\Gamma^* \beta_{P_2}}{\sqrt{\beta'_{P_2} \Gamma^* \beta_{P_2}}} \quad (4.41)$$

and



$$R_{P_1} = k_1 \sqrt{\boldsymbol{\beta}'_{P_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{P_1}} \quad \text{and} \quad R_{P_2} = k_2 \sqrt{\boldsymbol{\beta}'_{P_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{P_2}}, \quad (4.42)$$

respectively. The total MPPG-LGSI expected genetic gain per trait and selection response at both stages are equal to  $\mathbf{E}_{P_1} + \mathbf{E}_{P_2}$  and  $R_{P_1} + R_{P_2}$ . To simplify the notation, in Eqs. (4.41) and (4.42), we omitted the intervals between stages or selection cycles ( $L_G$ ). Matrices  $\boldsymbol{\Gamma}^*$  and  $\mathbf{C}^*$  are matrices  $\boldsymbol{\Gamma}$  and  $\mathbf{C}$  adjusted for previous selection on  $I_{P_1}$  according to Eqs. (4.34) and (4.35) respectively in the MPPG-LGSI context.

The correlation between  $I_{P_1} = \boldsymbol{\beta}'_{P_1} \boldsymbol{\gamma}_1$  and  $I_{P_2} = \boldsymbol{\beta}'_{P_2} \boldsymbol{\gamma}$  can be written as

$$\rho_{12} = \frac{\boldsymbol{\beta}'_{P_1} \mathbf{A}_1 \boldsymbol{\beta}_{P_2}}{\sqrt{\boldsymbol{\beta}'_{P_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{P_1}} \sqrt{\boldsymbol{\beta}'_{P_2} \boldsymbol{\Gamma} \boldsymbol{\beta}_{P_2}}}. \quad (4.43)$$

In Eq. (4.43), matrix  $\boldsymbol{\Gamma}$  was not adjusted for previous selection on  $I_{P_1} = \boldsymbol{\beta}'_{P_1} \boldsymbol{\gamma}_1$ .

## 4.6.2 Numerical Examples

To illustrate the MPPG-LGSI theory, we use the simulated data described in Sect. 4.4.3. Suppose that we select two traits at stages 1 and 2; then, at stage 1,  $\widehat{\boldsymbol{\Gamma}}_1 = \begin{bmatrix} 16.26 & -6.51 \\ -6.51 & 5.79 \end{bmatrix}$  and  $\widehat{\mathbf{A}}_1 = \begin{bmatrix} 16.26 & -6.51 & 5.60 & 2.29 \\ -6.51 & 5.79 & -2.33 & -1.62 \end{bmatrix}$  are the estimated covariance matrices of  $\boldsymbol{\Gamma}_1$  and  $\mathbf{A}_1$  respectively. We restricted trait 2 with  $\mathbf{d} = -2$ ; then, at the stage 1 matrix  $\mathbf{U}'_1 = [0 \ 1]$  and at the stage 2 matrix  $\mathbf{U}'_2 = [0 \ 1 \ 0 \ 0]$ . In addition,  $\widehat{\mathbf{Q}}_{G_1} = \mathbf{U}_1 (\mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1$ ,  $\widehat{\mathbf{Q}}_{G_2} = \mathbf{U}_2 (\mathbf{U}'_2 \widehat{\boldsymbol{\Gamma}}_2)^{-1} \mathbf{U}'_2 \widehat{\boldsymbol{\Gamma}}$ ,  $\widehat{\mathbf{K}}_{G_1} = [\mathbf{I} - \widehat{\mathbf{Q}}_{G_1}]$ , and  $\widehat{\mathbf{K}}_{G_2} = [\mathbf{I} - \widehat{\mathbf{Q}}_{G_2}]$  are the estimates of matrix projectors associated with stages 1 and 2 (Eqs. 4.37 and 4.38 for details).

In Sect. 4.4.3, we showed that the estimated MRLGSI vector of coefficients for stage 1 was  $\widehat{\boldsymbol{\beta}}'_{R_1} = \widehat{\boldsymbol{\beta}}'_1 \widehat{\mathbf{K}}'_{G_1} = [1.386 \ 1.550]$ . Thus, by Eq. (4.37), to obtain  $\widehat{\boldsymbol{\beta}}_{P_1} = \widehat{\boldsymbol{\beta}}_{R_1} + \widehat{\theta}_1 \mathbf{U}_1 (\mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d}$ , we only need to obtain  $\widehat{\theta}_1$  and  $\mathbf{U}_1 (\mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d}$ , where  $\mathbf{d} = -2$  and  $\widehat{\theta}_1 = \frac{\mathbf{d}' (\mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 \widehat{\mathbf{A}}_1 \mathbf{w}}{\mathbf{d}' (\mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d}}$ . It can be shown that  $\mathbf{U}_1 (\mathbf{U}'_1 \widehat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d} = \begin{bmatrix} 0 \\ -0.345 \end{bmatrix}$  and  $\widehat{\theta}_1 = 8.125$ ; therefore,  $\widehat{\boldsymbol{\beta}}'_{P_1} = [1.39 \ -1.25]$  is the MPPG-LGSI vector of coefficients at stage 1.

Suppose that the total proportion retained for the two stages was 20%; then,  $k_1 = 0.744$  is an approximate selection intensity associated with MPPG-LGSI and

the estimated MPPG-LGSI accuracy, selection response, and expected genetic gain at stage 1 were  $\hat{\rho}_{HI_1} = \sqrt{\frac{\hat{\beta}'_{P_1}\hat{\Gamma}_1\hat{\beta}_{P_1}}{\mathbf{w}'\hat{\mathbf{C}}\mathbf{w}}} = 0.71$ ,  $\hat{R}_{P_1} = k_1\sqrt{\hat{\beta}'_{P_1}\hat{\Gamma}_1\hat{\beta}_{P_1}} = 5.90$  and  $\hat{\mathbf{E}}'_{P_1} = k_1\frac{\hat{\mathbf{A}}'_1\hat{\beta}_{P_1}}{\sqrt{\hat{\beta}'_{P_1}\hat{\Gamma}_1\hat{\beta}_{P_1}}} = [2.88 \quad -1.53 \quad 1.00 \quad 0.49]$  respectively.

It can be shown that at stage 2,  $\mathbf{d}'(\mathbf{U}'_1\hat{\Gamma}_1\mathbf{U}_1)^{-1}\mathbf{U}'_1 = [0 \quad -0.345 \quad 0 \quad 0]$ ,  $\hat{\theta}_2 = 8.125$  and  $\hat{\beta}'_{P_2} = \mathbf{w}' = [1 \quad -1 \quad 1 \quad 1]$ . Thus, the estimated MPPG-LGSI accuracy, selection response, and expected genetic gain at this stage were  $\hat{\rho}_{HI_2} = \sqrt{\frac{\mathbf{w}'\hat{\Gamma}^*\mathbf{w}}{\mathbf{w}'\hat{\mathbf{C}}^*\mathbf{w}}} = 0.64$ ,  $\hat{R}_{P_2} = k_2\sqrt{\mathbf{w}'\hat{\Gamma}^*\mathbf{w}} = 4.10$ , and  $\hat{\mathbf{E}}'_{P_2} = k_2\frac{\hat{\Gamma}^*\mathbf{w}}{\sqrt{\mathbf{w}'\hat{\Gamma}^*\mathbf{w}}} = [1.74 \quad -0.92 \quad 0.85 \quad 0.58]$  respectively, where  $k_2 = 0.721$ . The estimated total MPPG-LGSI selection response and expected genetic gain for both stages were  $\hat{R}_{P_1} + \hat{R}_{P_2} = 9.99$  and  $\hat{\mathbf{E}}'_{P_1} + \hat{\mathbf{E}}'_{P_2} = [4.62 \quad -2.45 \quad 1.85 \quad 1.07]$  respectively. Note that the total expected genetic gain for trait 2 was  $-2.45$ , which is similar to  $\mathbf{d} = -2$ , the PPG imposed by the breeder. Finally, to simplify the notation, we omitted the intervals between stages or selection cycles ( $L_G$ ) in the estimated MPPG-LPSI selection response and expected genetic gain for both stages.

## References

- Arismendi JC (2013) Multivariate truncated moments. *J Multivar Anal* 117:41–75
- Cochran WG (1951) Improvement by means of selection. In: Neyman J (ed) Proc. 2nd Berkeley Symp. on Math., Stat. and Probability, pp 449–470
- Cunningham EP (1975) Multi-stage index selection. *Theor Appl Genet* 46:55–61
- Dekkers JCM (2014) Multiple stage selection. Armidale Animal Breeding Summer Course 2014, University of New England, Armidale, NSW, Australia
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longman, New York
- Hattaway JT (2010) Parameter estimation and hypothesis testing for the truncated normal distribution with applications to introductory statistics grades. All Theses and Dissertations. Paper 2053
- Hicks C, Muir WM, Stick DA (1998) Selection index updating for maximizing rate of annual genetic gain in laying hens. *Poult Sci* 77:1–7
- Mi X, Utz HF, Technow F, Melchinger AE (2014) Optimizing resource allocation for multistage selection in plant breeding with R package *Selectiongain*. *Crop Sci* 54:1413–1418
- Rausand M, Høyland A (2004) System reliability theory: models, statistical methods, and applications, 2nd edn. Wiley, Hoboken, NJ
- Saxton AM (1983) A comparison of exact and sequential methods in multi-stage index selection. *Theor Appl Genet* 66:23–28
- Thomas GB (2014) Thomas' calculus: early transcendentals, 3r edn. Pearson Education, Inc., Boston, MA
- Xu S, Muir WM (1992) Selection index updating. *Theor Appl Genet* 83:451–458
- Young SSY (1964) Multi-stage selection for genetic gain. *Heredity* 19:131–143

# 5

## Breeding Design and its Stochastic Simulation



Fernando H. Toledo, José Crossa, and Juan Burgueño

**Abstract** Stochastic simulation can contribute to a better understanding of the problem, and has already been successfully applied to evaluate other breeding scenarios. Despite all the theories developed in this book concerning different types of indices, including phenotypic data and/or data on molecular markers, no examples have been presented showing the long-term behavior of different indices. The objective of this chapter is to present some results and insights into the *in silico* (computer simulation) performance comparison of over 50 selection cycles of a recurrent and generic population breeding program with different selection indices, restricted and unrestricted. The selection indices included in this stochastic simulation were the linear phenotypic selection index (LPSI), the eigen selection index method (ESIM), the restrictive LPSI, and the restrictive ESIM.

### 5.1 Stochastic Simulation

Simulations were used to evaluate the accuracy, effectiveness, response to selection, and the decrease in the overall genetic variance in a recurrent selection scheme under the use of the Smith (1936) and Hazel (1943) index (or linear phenotypic selection index, LPSI, see Chap. 2 for details); the eigen selection index method (ESIM, see Chap. 7 for details); the Kempthorne and Nordskog (1959) restricted index (K&N or restricted phenotypic selection index, RLPSI, see Chap. for details); and the restricted eigen selection index method (RESIM, see Chap. for details). The different scenarios are described below and encompass variations in the nature of the genetic correlation between traits in addition to their expected heritabilities.

---

F. H. Toledo (✉) · J. Crossa · J. Burgueño

Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Mexico, Mexico  
e-mail: f.toledo@cgiar.org



### 5.1.1 Breeding Design

A total of 50 forward recurrent selection cycles of modern breeding were simulated, in which the breeder has the ability to select based on breeding value estimates of genetically correlated traits, and to apply the various above-mentioned selection indices. All simulated scenarios (described below) followed a common general breeding design. In each cycle, 350 full sib progenies ( $S_1$ ) were generated taking 700 parents at random from the base population. From each progeny, 100 double-haploid lines were randomly derived (which shortened the cycle interval by five inbreeding generations). The simulated phenotypic values of the 35,000 resulting lines were then evaluated in simulated trials. The selection was made by means of the progeny average performance. The selected progenies (top quarter) according to each index were then recombined by random mating a sample of the lines within the progeny to recover the population for the next cycle.

### 5.1.2 Simulating Quantitative Traits

Genetically correlated quantitative traits were simulated assuming a full pleiotropic model. This was carried out by randomly sampling genetic effects for all segregating sites from a multivariate normal distribution with zero mean and a previously stated variance–covariance. The genetic effects were in turn used to compute true breeding values (TBVs). An individual's phenotype was obtained by taking its TBV and adding a zero mean normally random term with variance consistent with the expected heritability ( $h^2$ ) for the trait at which phenotyping occurred. The genetic variance in each cycle was calculated as the variance of the TBV of the individuals in that generation. However, it was expressed as relative values of the genetic variance in the initial cycle. The realized response to selection was also standardized in units of the genetic standard deviation in cycle 0. Cycle 0 was used as the base generation because it represents the available genetic variability, and also to observe, from the start, the genetic changes in future breeding generations.

An empirical genome was considered comprising a set of 10 linkage groups (chromosomes), each 200 cM in length, and 1000 uniformly distributed segregating sites. To represent the historical evolution and recent breeding efforts up to the present day in addition to incorporating a steady state of known linkage disequilibrium (LD) structure existing in crops, the starting populations (cycle 0) were taken after 200 generations of random mating within an effective population size of 1000 segregating for all loci in which the allele frequency was 0.5.

The *in silico* meiosis reflected the Mendelian laws of segregation for diploid species, by a count-location process that mimics the Haldane map function (Haldane 1919). Thus, homologous chromosomes are paired into bivalents and recombined through randomly positioned chiasmata. The number of chiasmata follows a Poisson distribution, where the  $\lambda$  parameter represents the chromosome length in Morgans and their positions are uniformly distributed, i.e., without interference between crossovers or any mutagenesis process.



### 5.1.3 Simulated Scenarios

Three traits were considered, one with low heritability (the first,  $h^2 = 0.2$ ) and two with high heritability (the second and the third,  $h^2 = 0.5$ ). The correlations between the first and second trait vary from positive ( $\rho_G = 0.5$ ) to negative ( $\rho_G = -0.5$ ). The third trait was always considered with segregation independent from the two others.

The selection process involved two unrestricted indices: the LPSI (see Chap. ), which ranks the progenies based on the average merit of their lines considering equal economic weights for all traits, and the ESIM (see Chap. ), where the progenies were ranked in terms of ESIM values. Regarding the restricted selection indices, the RLPSI (or K&N) was employed (see Chap. ) with equal economic weights for the traits in addition to the RESIM (see Chap. ). Because of the restrictions, two different situations were evaluated in the latter cases, i.e., where the restrictions were applied for each of the first and second traits separately.

Thus, all simulated scenarios encompass a three-way factorial: four selection procedures (the LPSI, the ESIM, the RLPSI or K&N, and the RESIM); two correlation scenarios, positive ( $\rho_G = 0.5$ ) and negative correlations ( $\rho_G = -0.5$ ) between the first and second trait; and two constraint situations, where the restrictions were applied separately for the first and second traits.

To simulate genetically correlated traits a full pleiotropic model was assumed. Gene effects were sampled from a multivariate normal distribution with zero mean and a previously stated variance–covariance matrix. In that sense it is possible to represent a quantitative and infinitesimal model. Each genes has its own effect varying according to a probabilistic density i.e., genes with positive and negative effects varying its effects sizes; alleles with large effects at lower frequency (major genes) and alleles with modest effects at higher frequency (minor genes).

### 5.1.4 Inferences

Results are presented as summaries of 100 Monte Carlo replicates for each scenario and include the response to selection, decreases in the genetic variance, selection accuracy, and observed heritabilities. The meiosis routine was implemented in C++, and compiled, linked, and through the facilities provided by the Rccp R package (Eddelbuettel 2013). All simulations were performed, analyzed, and summarized in R version 3.3.3 (R Development Core Team 2017).

## 5.2 Results

Overriding the results of the simulations regarding the four selection indices under the different trait genetic correlations and restrictions, scenarios are presented in terms of the consistency of the observed heritabilities of the traits; the response to

selection and changes in genetic variance for each trait; and the accuracy of the indices' selection.

First of all, the results show the stability of the Monte Carlo replicates in terms of possible deviations in the observed heritability from that expected, which in turn may affect further inferences (Table 5.1). The *type I* error ( $\alpha$ ) of the *t test* comparing expected and observed heritabilities for all simulated scenarios did not show important and significant departures. Slight departures that may be due to Monte Carlo error ( $P < 0.05$ ) were found, namely: for both high and low heritability traits of the LPSI at cycle 5 when they were negatively correlated; for the independent trait also with the LPSI at cycle 50, but, when the other traits are positively correlated; for the high heritable trait at the first and last cycles, both under positive correlation in the ESIM and RESIM indices respectively; and for the low heritability trait in both restricted indices (RLPSI and RESIM) in cycles 0 and 5 for respective and negative and positive correlations.

A complementary estimate of the power (*type II* error or  $\beta$ ) of the tests was performed considering departures from the expected heritabilities of 1%. It was verified that the average power if the observed estimates was around 70%, which reinforces the appropriateness of the simulation findings.

### 5.2.1 Realized Genetic Gains

Figure 5.1 shows the average genetic gains (expressed as standard deviations from the mean of cycle 0) for cycles 0–50 for the traits (low and high heritabilities and the independent trait); the four selection indices (unrestricted: LPSI and ESIM and restricted: RLPSI and RESIM) when the correlations are positive and negative.

It is important to note that even after 50 recurrent cycles none of the scenarios has shown any indication that the selection plateau has been reached (Fig. 5.1). It is considered that even with the variation of the gains in the scenarios, there were increases in the merit of the target traits. Thus, the employment of selection indices is an effective way of achieving progress in long-term multi-trait selection.

As expected, the unrestricted selection indices have shown genetic gains higher than their restricted counterparts (Fig. 5.1). It must be highlighted that the restrictions proved their properties because when any trait was restricted, no gains were obtained for that trait (data not shown). The higher gains obtained with unrestricted indices is well known and justified in comparison with their restricted homologous because the net genetic merit is beneficited by the gains in all traits, while, with gains constrained to zero in some traits, there are no indirect gains that may be highlighted especially because of positive correlations.

The independent trait has presented the higher gains in comparison with the other traits for all correlation and selection process scenarios. The higher gains, however, were for the RESIM followed by the RLPSI in both positive and negative correlations (Fig. 5.1e and f). These findings may be understood both under the nature of the trait (independent inheritance) and over the properties of the restricted indices.

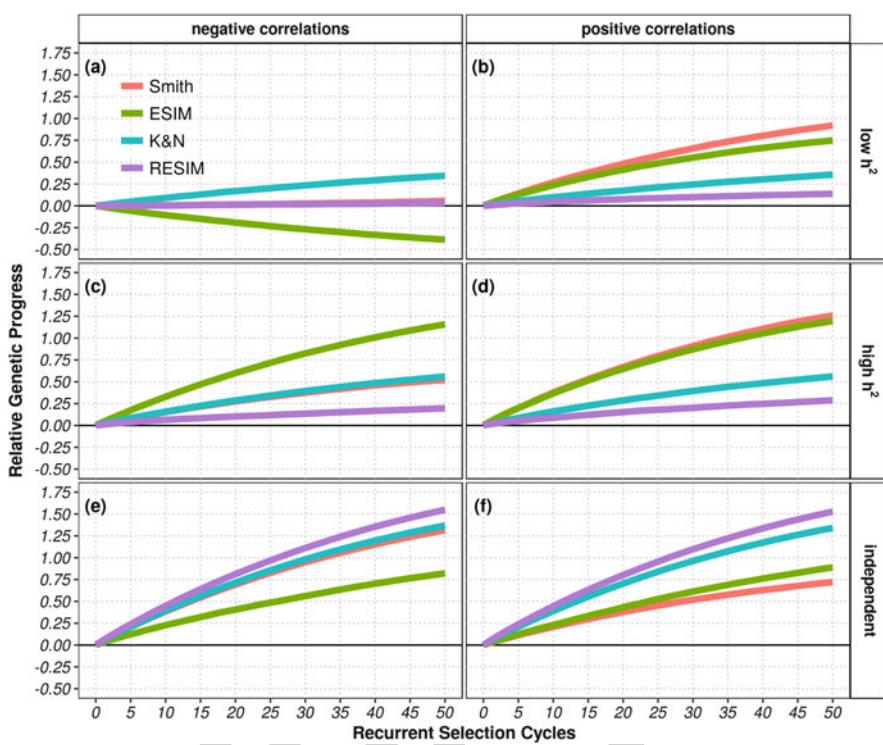
**Table 5.1** Mean ( $\mu$ ), standard deviation ( $\sigma$ ), type I error ( $\alpha$ ), and type III power error ( $\beta$ ) of the observed heritability in 100 Monte Carlo replicates for traits with low and high heritability ( $h^2$ ) and independent at cycles 0, 5, and 50 of a simulated selection given four indices, the linear phenotypic selection index (LPSI), the ESIM, the restricted linear phenotypic selection index (RLPSI), and the RESIM, with positive and negative correlations between the traits low  $h^2$  and high  $h^2$

Scenario/cycle	$\mu$	$\sigma$	Low $h^2$			High $h^2$			Independent					
			Errors		$\mu$	Errors		$\mu$	$\sigma$	$\alpha$	$\beta$	$\alpha$	$\beta$	
			$\alpha$	$\beta$		$\alpha$	$\beta$							
LPSI														
Negative	0	0.201	0.016	0.565	0.993	0.502	0.035	0.536	0.513	0.502	0.035	0.524	0.531	
	5	0.204	0.016	0.016	0.993	0.493	0.031	0.028	0.612	0.504	0.034	0.191	0.540	
	50	0.200	0.015	0.852	0.996	0.501	0.035	0.672	0.524	0.505	0.034	0.124	0.547	
Positive	0	0.200	0.014	0.782	0.999	0.500	0.037	0.902	0.485	0.503	0.033	0.347	0.556	
	5	0.200	0.016	0.886	0.995	0.499	0.032	0.758	0.601	0.502	0.033	0.561	0.562	
	50	0.200	0.015	0.950	0.997	0.504	0.035	0.214	0.513	0.507	0.035	0.046	0.511	
ESIM														
Negative	0	0.200	0.013	0.711	1.000	0.502	0.035	0.552	0.529	0.503	0.038	0.371	0.464	
	5	0.201	0.015	0.677	0.997	0.497	0.033	0.349	0.559	0.502	0.031	0.445	0.611	
	50	0.200	0.015	0.749	0.997	0.503	0.034	0.309	0.548	0.499	0.035	0.815	0.514	
Positive	0	0.199	0.015	0.559	0.996	0.493	0.032	0.029	0.593	0.502	0.030	0.530	0.642	
	5	0.200	0.014	0.892	0.999	0.498	0.032	0.490	0.589	0.505	0.037	0.222	0.472	
	50	0.202	0.018	0.405	0.975	0.504	0.029	0.132	0.679	0.500	0.033	0.888	0.560	
RLPSI														
Negative	0	0.198	0.016	0.178	0.994	0.497	0.033	0.405	0.557	0.503	0.032	0.438	0.587	
	5	0.203	0.014	0.042	0.999	0.505	0.033	0.110	0.567	0.501	0.034	0.471	0.541	
	50	0.202	0.015	0.173	0.996	0.501	0.032	0.742	0.602	0.500	0.033	0.397	0.572	
Positive	0	0.201	0.014	0.295	0.999	0.504	0.033	0.239	0.570	0.502	0.034	0.130	0.539	
	5	0.201	0.016	0.709	0.990	0.507	0.035	0.059	0.510	0.502	0.034	0.689	0.557	
	50	0.201	0.013	0.331	1.000	0.504	0.032	0.195	0.591	0.507	0.035	0.061	0.533	

(continued)

Table 5.1 (continued)

Scenario/cycle	Low $h^2$			High $h^2$			Independent		
	$\mu$	$\sigma$	Errors $\alpha$	Errors			$\mu$	$\sigma$	Errors $\alpha$
				$\mu$	$\sigma$	$\beta$			
RESIM									
Negative	0	0.201	0.017	0.504	0.987	0.506	0.034	0.075	0.545
	5	0.202	0.016	0.230	0.993	0.506	0.032	0.090	0.585
Positive	50	0.199	0.015	0.490	0.996	0.507	0.034	0.051	0.549
	0	0.203	0.014	0.018	0.999	0.506	0.035	0.075	0.510
	5	0.202	0.014	0.171	0.998	0.502	0.034	0.034	0.489
	50	0.200	0.017	0.911	0.988	0.508	0.036	0.034	0.495



**Fig.5.1** Average genetic gains in 100 Monte Carlo replicates for traits with low and high heritability ( $h^2$  0.2 and 0.5) and independent along cycles 0–50 of a simulated selection given four indices, the linear phenotypic selection index (LPSI), the ESIM, the restricted linear phenotypic selection index (RLPSI), and the RESIM with positive (0.5) and negative (-0.5) correlations between the traits low  $h^2$  and high  $h^2$ . (a) Gains for the trait with low heritability when it is negatively correlated with the high heritability trait. (b) Gains for the trait with low heritability when it is positively correlated with the high heritability trait. (c) Gains for the trait with high heritability when it is negatively correlated with the low heritability trait. (d) Gains for the trait with high heritability when it is positively correlated with the low heritability trait. (e) Gains for the independent trait when the other traits are negatively correlated. (f) Gains for the independent trait when the other traits are positively correlated

As the third trait becomes independent from the others, there are no indirect effects owing to the constraints in the gains of the other traits. With regard to the technical features of the RESIM, it must be emphasized that because of the eigen decomposition, the largest eigenvector obtains higher weight from the most variable trait and consequently ends in distinct gains, which in this case is the independent trait.

The Smith (or LPSI) and ESIM produce similar genetic gains for highly heritable traits when the genetic correlations are positive (Fig. 5.1d). The ESIM is simply another way of obtaining the LPSI based on the eigen decomposition theory, which avoids the assignment of economic weights. Thus, the results prove that the same results may be found with both indices. However, the ESIM is the preferred index



because of its advantages over the LPSI: no subjective decision for selecting economic weights, and better statistical sampling properties.

When the traits are negatively correlated, the trait with greater heritability has shown important realized genetic gains based on the ESIM and similar gains for the LPSI and its restricted analogous, i.e., the RLPSI (Fig. 5.1a and c). In addition, when traits are negatively correlated, restricting the traits with low heritability is an alternative, to ensure similar progress to the use of unrestricted indices for highly heritable traits. On the contrary, it is also interesting to note that the ESIM has the worst performance when the traits are negatively correlated for trait with lower heritability (Fig. 5.1a).

On the other hand, as already pointed out, the ESIM performance surpasses all the others with regard to the highly heritable trait (Fig. 5.1c and d). The reason for this is similar to the above-mentioned regarding the properties of the eigen decomposition. When the first trait is negatively correlated with the second one, heavier weight is given to the trait with higher heritability than to the trait with low heritability. However, when the traits are positively correlated, synergic and indirect effects increase both traits, one positively affecting the other.

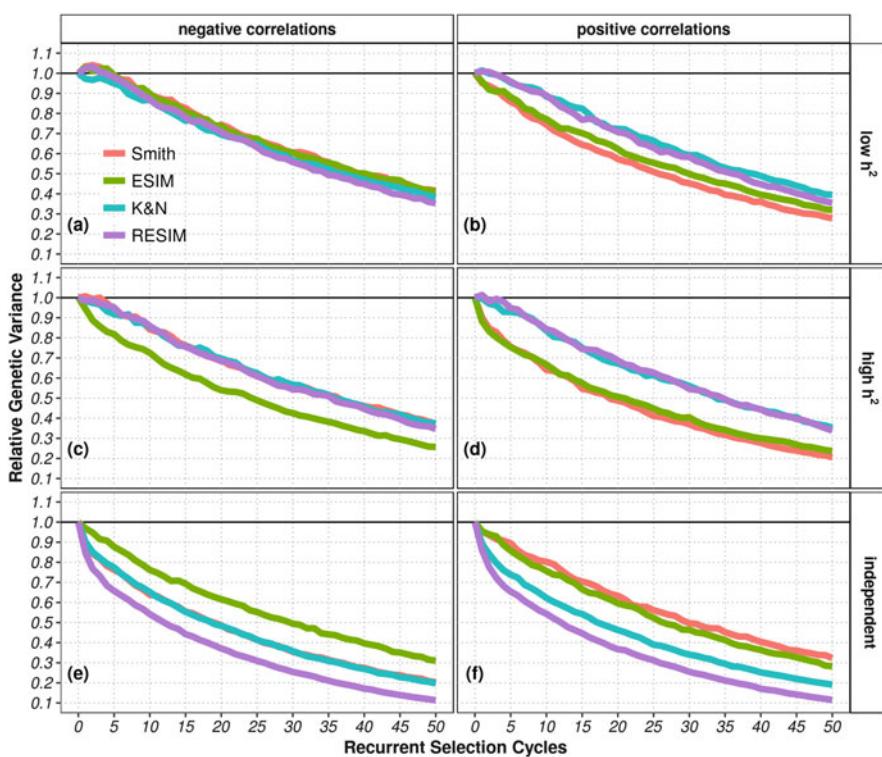
When the traits are positively correlated but with low heritability, the LPSI and the ESIM have similar realized genetic gains until cycle 25; after this selection cycle, the LPSI is superior to the ESIM (Fig. 5.1b). In this case, the two restrictive indices, the RLPSI and the RESIM, are given lower realized genetic gains than the LPSI and the ESIM (Fig. 5.1b). Finally, considering the third trait (the independent one), the RESIM provides the greater realized genetic gains (Fig. 5.1e and f).

## 5.2.2 *Genetic Variances*

In Fig. 5.2, the average relative decreases in the genetic variances along the 50 cycles of selection for the three traits (with low and high heritability traits in addition to the independent trait) under the selection system given by the four selection indices, restricted (the RLPSI and the RESIM) and unrestricted (the LPSI and the ESIM), both with negative and positive correlations between the first and second traits.

As a general result, it is clear that after selection there were decreases in the genetic variance along the recurrent cycles (Fig. 5.2). From the most conservative decrease (around 40% in Fig. 5.2a and b) to the sharp decrease (close to 10% in Fig. 5.2e and f) and in contrast to the trends in genetic gains, it is possible to conceive that the genetic variability was not yet exhausted by selection. This observation endorses what was said regarding the effectiveness of the selection indices as a criterion for long-term multi-trait selection.

As expected, the restricted indices are more conservative, maintaining greater genetic variance (Fig. 5.2). Their feature is to prevent the restricted trait from changing its genetic merit. Thus, they tend to keep its genetic variance unchanged,



**Fig. 5.2** Average genetic variances in 100 Monte Carlo replicates for traits with low and high heritability ( $h^2$  0.2 and 0.5) and independent along cycles 0–50 of a simulated selection given four selection indices, the LPSI, the ESIM, the RLPSI, and the RESIM, with positive (0.5) and negative (−0.5) correlations between the traits low  $h^2$  and high  $h^2$ . (a) Genetic variance of the low heritability trait when it is negatively correlated with the high heritability trait. (b) Genetic variance of the low heritability trait when it is positively correlated with the high heritability trait. (c) Genetic variance of the high heritability trait when it is negatively correlated with the low heritability trait. (d) Genetic variance of the high heritability trait when it is positively correlated with the low heritability trait. (e) Genetic variance of the independent trait when the other traits are negatively correlated. (f) Genetic variance of the independent trait when the other traits are positively correlated

which is reflected in the lower decreases in the genetic variance, even under the indirect effects of the other traits.

It should be noted that there was a slight increase in variance in the short term (up to cycle 3) for the trait with lower heritability when negatively correlated with the highly heritable one (Fig. 5.2a and b). This is an outcome of the changes in allele frequencies of the first trait due to the indirect effects of the second trait and/or the release of genetic disequilibrium owing to the assortative mating of the individuals given higher weights regarding the second trait (highly heritable).

Reflecting the findings regarding the genetic gains (Fig. 5.1), the trait with strong decreases in genetic variance on average was the one in which the response

to selection was more pronounced, i.e., the independent trait (Fig. 5.2e and f). This trait has shown stronger decreases over the selection through the ESIM index in both positive and negative correlation scenarios. As mentioned before, as the third trait is independent of the others, a greater response to selection was achieved in that trait and consequently strong changes in allele frequencies, which drove the decreases in genetic variance.

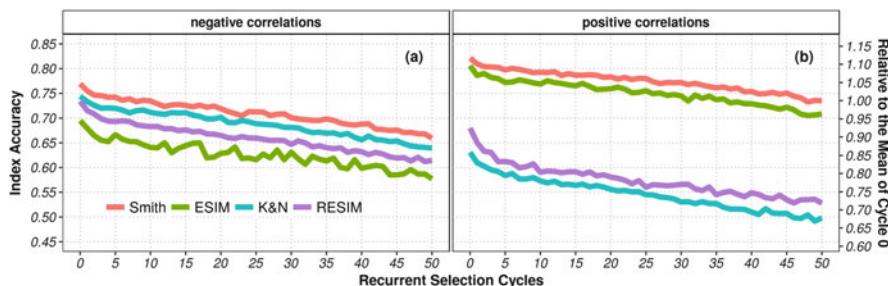
When the heritability is high, it is easy to differentiate the trends in the decrease in the genetic variance between restricted and unrestricted indices (Fig. 5.2c). It is more evident, especially when the traits are positively correlated (Fig. 5.2d). Thus, the ESIM has the highest decreases followed by the LPSI. Nevertheless, for the traits with low heritability, the decreases in genetic variance are indistinguishable between the indices, showing that the effectiveness of the response to selection is a function of the heritability (Fig. 5.2a and b).

### 5.2.3 Selection Accuracy

The accuracy of the selection was measured as the square root of the correlation between the net genetic merit and the estimated linear function of each index. Figure 5.3 shows the absolute accuracies (left axis) and relative values in relation to the mean accuracy of the first cycle (right axis) for all indices in both negative (Fig. 5.3a) and positive (Fig. 5.3b) correlation scenarios.

In all cases, a reduction in the selection precision of all the indices was observed. The effect of selection is the improvement in the genetic merit of the traits by means of changes in allele frequencies that also affect/decrease the genetic variance. However, as a side effect, the selection becomes harder and has lower precision.

The LPSI has shown greater accuracy in comparison with the other indices in any situation (Fig. 5.3a and b). Its main feature is precisely maximizing the correlation between the net genetic merit and the linear combination of the trait. It may be



**Fig. 5.3** Average absolute and relative accuracy of selection in 100 Monte Carlo replicates for traits with low and high heritability ( $h^2$ ) and independent along cycles 0–50 of a simulated selection given four selection indices, the LPSI, the ESIM, the RLPSI, and the RESIM with positive and negative correlations between the traits low  $h^2$  and high  $h^2$



argued that the ESIM also does that; however, only when the phenotypic and genotypic variances and covariances are known are they the best linear predictors. Thus, according to what was found, it is possible to note that the ESIM was more affected by the sampling properties when estimating matrices of variance and covariance (Fig. 5.3a).

For the scenario with positive correlations, the differences between the two types of indices, the restricted ones and the unrestricted ones, were clear, as the unrestricted indices have shown greater selection accuracy (Fig. 5.3b). This reflects the fact that the restricted index constrains the gains by means of restrictions in the correlation between the net genetic merit and the linear combination of the traits.

## References

- Eddelbuettel D (2013) Seamless R and C++ Integration with Rcpp. Springer, New York  
Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. *J Genet* 8:299–309  
Hazel IN (1943) The genetics basis for constructing selection indexes. *Genetics* 28:476–490  
Kempthorne O, Nordskog AW (1959) Restricted selection indices. *Biometrics* 15:10–19  
R Core Team (2017) R: A language and environment for statistical computing  
Smith HF (1936) A discriminant function for plant selection. *Ann Eugenics* 7:240–250

# 6

## RIndSel Selection Indices for Breeding



**Gregorio Alvarado, Angela Pacheco, Sergio Pérez-Elizalde, Juan Burgueño, and Francisco M. Rodríguez**

**Abstract** RIndSel is a graphical unit interface that uses selection index theory to select individual candidates as parents for the next selection cycle. The index can be a linear combination of phenotypic values, genomic estimated breeding values, or a linear combination of phenotypic values and marker scores. Based on the restriction imposed on the expected genetic gain per trait, the index can be unrestricted, null restricted, or predetermined proportional gain indices. RIndSel is compatible with any of the following versions of Windows: XP, 7, 8, and 10. Furthermore, it can be installed on 32-bit and 64-bit computers. In the context of fixed and mixed models, RIndSel estimates the phenotypic and genetic covariance using two main experimental designs: randomized complete block design and lattice or alpha lattice design. In the following, we explain how RIndSel can be used to determine individual candidates as parents for the next cycle of improvement.

### 6.1 Background

The linear selection index theory (see Chaps. to for details) can be difficult to apply without the use of specific codes developed in statistical analysis system (SAS) software. At the International Maize and Wheat Improvement Center (CIMMYT, for its Spanish acronym), codes were developed in SAS software version 9.4 (SAS institute 2017) that can help to determine individuals as parents for the next selection cycle. The SAS codes can be found at the following link: <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10242>.

---

G. Alvarado · A. Pacheco · J. Burgueño (✉) · F. M. Rodríguez  
Biometrics and Statistics Unit, International Maize and Wheat Improvement Center  
(CIMMYT), Mexico, Mexico  
e-mail: g.alvarado@cgiar.org; J.Burgueno@cgiar.org

S. Pérez-Elizalde  
Departamento de Socioeconomía Estadística e Informática, Colegio de Postgraduados, Mexico,  
Mexico



Afterward, the SAS codes were translated to R language as scripts (Pacheco et al. 2017) and denoted by RIIndSel (R software to analyze Selection Indices), with the objective of creating a user-friendly graphical unit interface (GUI) in JAVA. The link to download the software is: <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10854>.

## 6.2 Requirements, Installation, and Opening

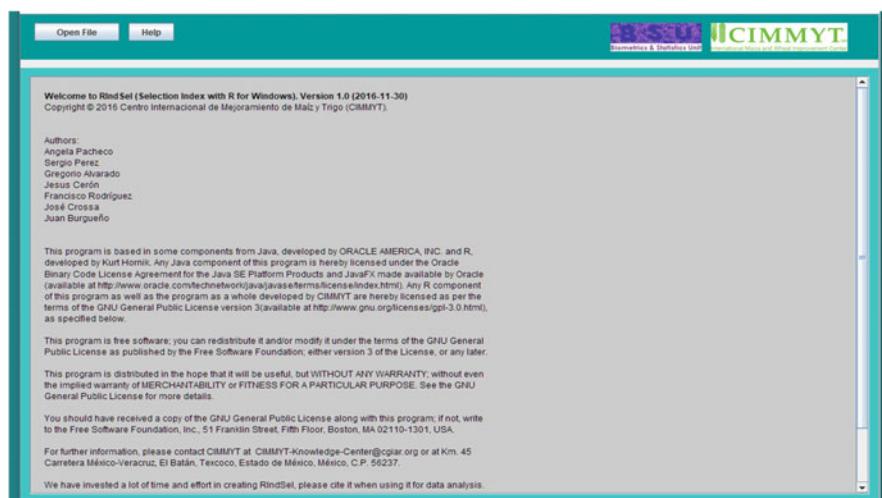
RIIndSel is compatible with a Windows platform, in any of the following versions: XP, 7, 8, and 10; furthermore, it can be installed on 32-bit and 64-bit computers. To install RIIndSel on a computer, the user must double-click on the executable file downloaded over the link given above and then follow the instructions that appear in the installation box. Once RIIndSel has been installed, it can be opened by:

1. Double-clicking on the shortcut located in the desktop.
2. Locating it in the Windows menu and clicking.
3. Locating the software via the pathway C:/RIIndSel, and double-clicking on RIIndSel.exe.

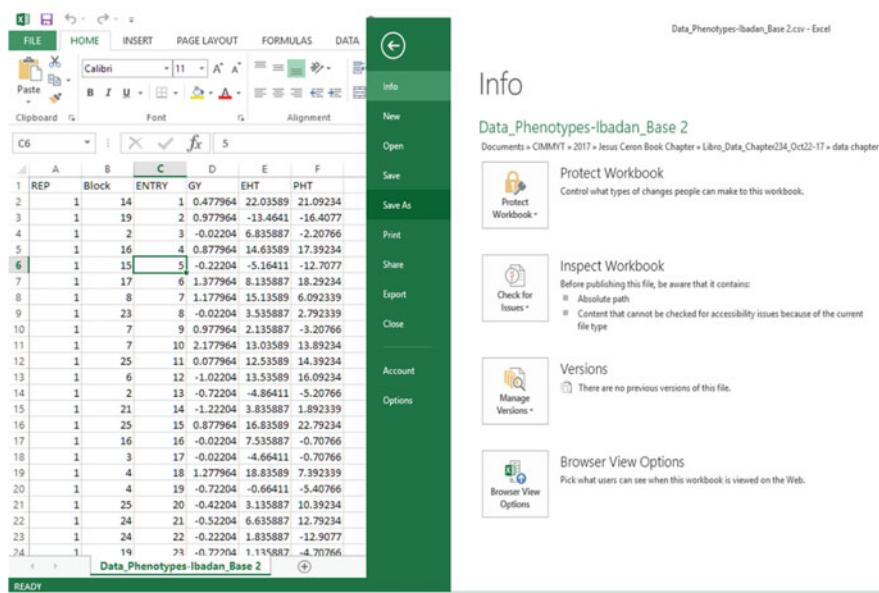
As we shall see, the software has been partitioned into two modules.

## 6.3 First Module: Data Reading and Helping

This module (Fig. 6.1) deploys two small boxes upper left denoted by “Open File” and “Help.” With *Open File*, the user may access a set of files where he/she can open, for example, the file of phenotypic data, which should contain information



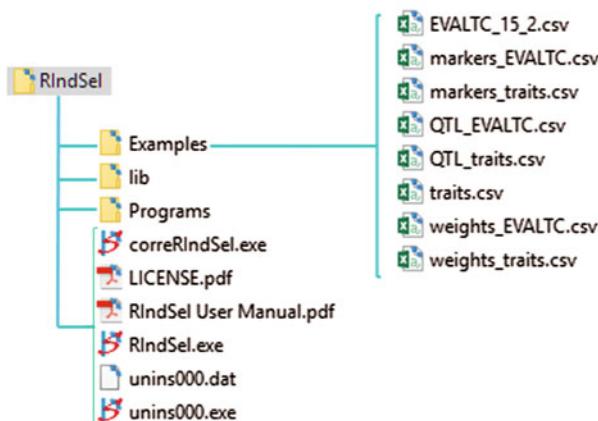
**Fig. 6.1** Module for reading data



**Fig. 6.2** Steps for saving a comma delimited file

associated with the experimental design. This file contains information about the field book where the experimental design variables can be identified in the first columns, whereas the remaining columns contain information about traits measured in the field; design variables and traits are connected by the plot number. Previously, the data set should have been captured in a spreadsheet using Excel or any other similar software and saved as a comma delimited file. To save the data as a comma delimited file in Excel, the following steps should be taken. In the Excel file that contains the data set (Fig. 6.2), select from the main menu: FILE → Save As → Browser View Options (look for the path where the data will be saved) → Save as type (look for CSV, comma separated values). The end of the file name should be “.csv,” indicating that the file is ready to be used.

The small box “Help” (Fig. 6.1) shows basic features such as the installation manual and software licenses. The installation manual provides a brief description of the selection indices that can be calculated and the pathway to where the software is located (Fig. 6.3). Furthermore, it shows folders related to the software features such as how the software could be used. There is also a folder called “Examples,” where the user can find data for test phenotypic selection indices, selection indices of coded score markers, and wide genome selection indices. The folders “Lib” and “Programs” contain information related to the software functioning; therefore, the authors highly recommend not modifying these folders.



**Fig. 6.3** Tree diagram of the RIndSel structure

## 6.4 Second Module: Capturing Parameters to Run

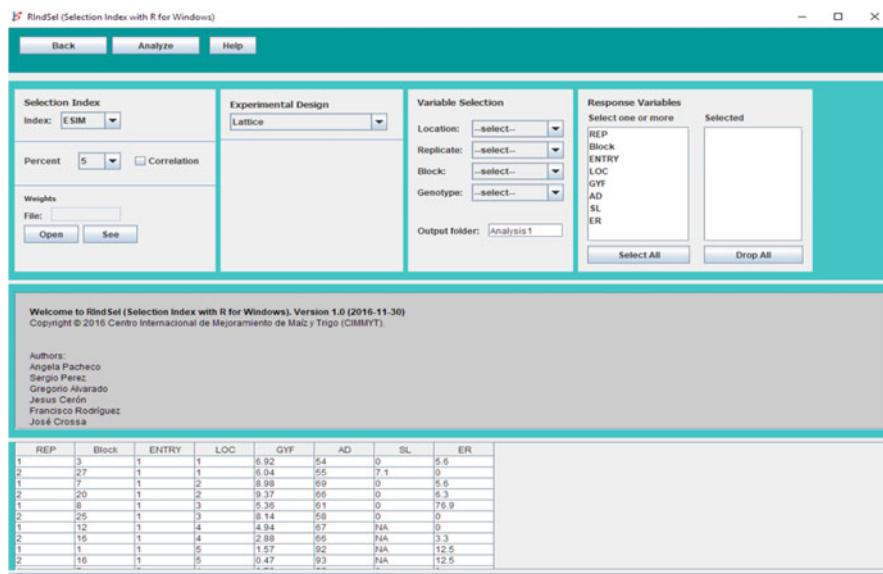
Once the data have been read (first module), RIndSel moves to the second module (Fig. 6.4), where some feedback is required:

1. To choose the selection index to calculate.
2. To select the experimental design.
3. To identify the variables of experimental design.
4. To choose the traits that will be used to calculate the selection index in the data file.

This module is structured in such a way that calculating any selection index is relatively easy. There are three other small buttons located upper left of the module: “Back,” “Analyze,” and “Help.” Back returns to the previous module (Fig. 6.1), Analyze executes and calculates the selection index, and Help provides the same functions as described in the previous section. In addition, there are four windows, each of which must be filled with the correct parameters. The first one is related to the indices that RIndSel is able to calculate (Fig. 6.5).

## 6.5 Selection Index

In this menu, it is necessary to define the percentage of genotypes that will be selected. By default, it is 5%, but any other percentage can be chosen. RIndSel uses the correlation matrix or the variance–covariance matrix to obtain the index; however, by default, the variance–covariance matrix is used. To work with the correlation matrix box, “Correlation” should be checked. The sign for “economic weights”



**Fig. 6.4** RIndSel module of analysis

can be used to determine the behavior of the expected genetic gain of the traits. For example, with  $-1$ , the mean of the traits tends to decrease, whereas with  $1$ , it increases. It is also possible to use the trait heritability. The economic weights can be assigned by creating a comma-delimited file with the name of the trait and economic weight sign (Fig. 6.6a). Once the file has been created, it can be browsed by pressing the open button and where the \*.csv file is located (Fig. 6.6b).

To calculate the restricted linear phenotypic selection index (RLPSI or K&N, see it is necessary to create the same file and incorporate an additional column called “*Restrictions*.”) This last column must be filled with the number one for those traits that remain fixed (restricted) and zeros for those traits that change (Fig. 6.7). An additional option is to ignore the “*Weights*” box, which means that RIndSel automatically presents an Excel file covering the options for capturing economic weights; the only requirement is that the file must be saved as a comma delimited file.

## 6.6 Experimental Design

The menu allows the user to select the field array design to be used. There are two choices:

1. Lattice or alpha-lattice
2. Random complete block designs

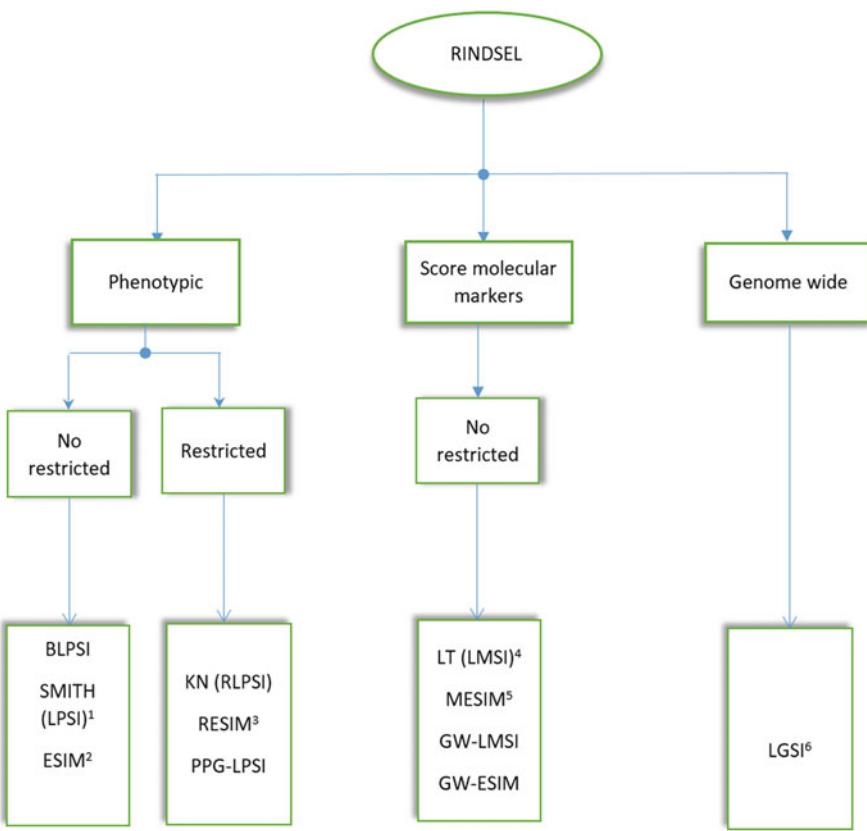


Fig. 6.5 Flow diagram of the selection indices that RIndSel is able to calculate; <sup>1</sup>Smith (1936), <sup>2,3</sup>Cerón-Rojas (2008a), <sup>4</sup>Lande R, Thompson R (1990), <sup>5</sup>Cerón-Rojas (2008b), <sup>6</sup>Cerón-Rojas (2015)

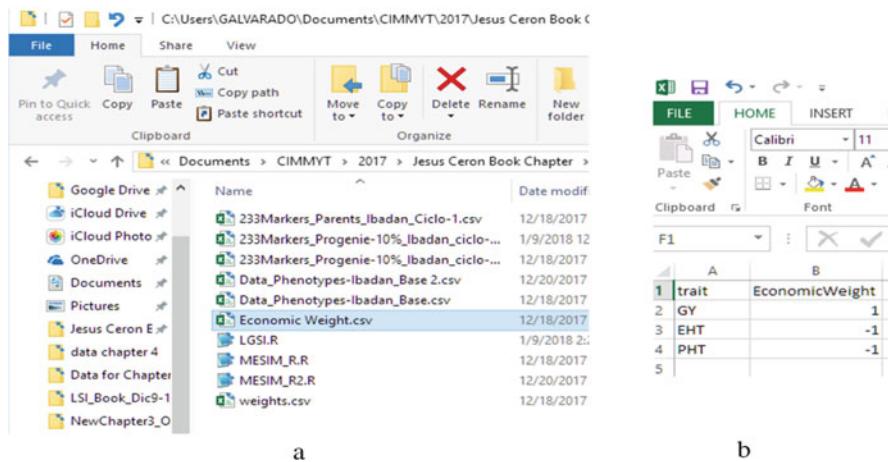


Fig. 6.6 Example of content for (a) economic weights of (b) file location

The screenshot shows a Microsoft Excel interface with the 'HOME' tab selected. The ribbon also includes 'FILE', 'INSERT', and 'PAGE LAYOUT'. Below the ribbon, the 'Clipboard' section shows 'Paste' options, and the 'Font' section shows 'Calibri' font, size '11', bold ('B'), italic ('I'), underline ('U'), and alignment ('A'). The 'Font' dropdown also includes a color palette. Below the font section is a toolbar with a dropdown for 'H12', a clear button, a checkmark button, and a formula editor ('fx'). The main area displays a table:

	A	B	C	D
1	trait	theta	rest	
2	T1		1	1
3	T2		-1	0
4	T3		1	0
5	T4		1	0

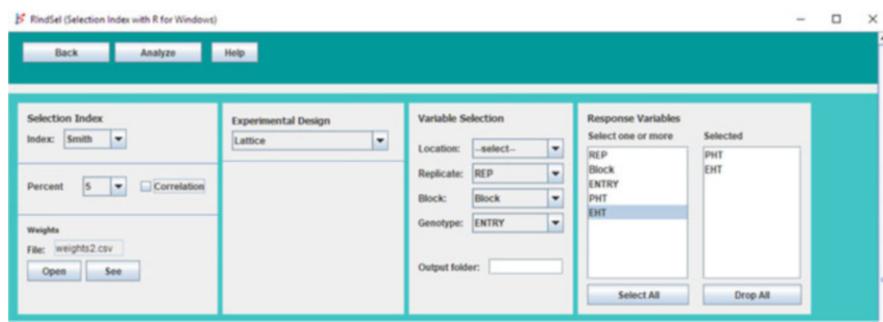
**Fig. 6.7** Economic weights for restricted selection indices

## 6.7 Variable Selection

Experimental design is strongly related to the “Variable Selection” menu, where it is possible to identify the variables that constitute the experimental design. Thus, we can choose variables that match with the “Location,” replicate for random complete block design and block, provided that we have a lattice or alpha-lattice experiment.

## 6.8 Response Variables

In this menu, the user can select traits to be used to calculate the selection index. It can be activated by clicking on the trait to be selected. Figure 6.8 shows an example of how this window must be filled when a Smith phenotypic selection index is calculated.



**Fig. 6.8** Example of parameters that could be used to calculate a phenotypic selection index

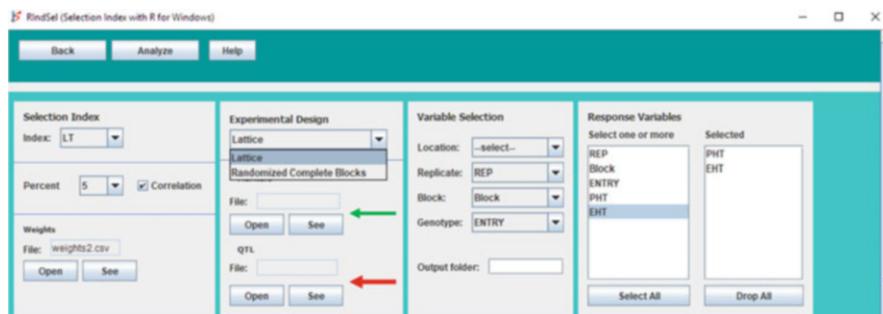
## 6.9 Molecular Selection Indices

If the selection index to be calculated is molecular, such as the Lande and Thompson (1990) or the linear molecular selection index (Fig. 6.9, and see Table 1.1, Chap. 1, for details), two additional files are required:

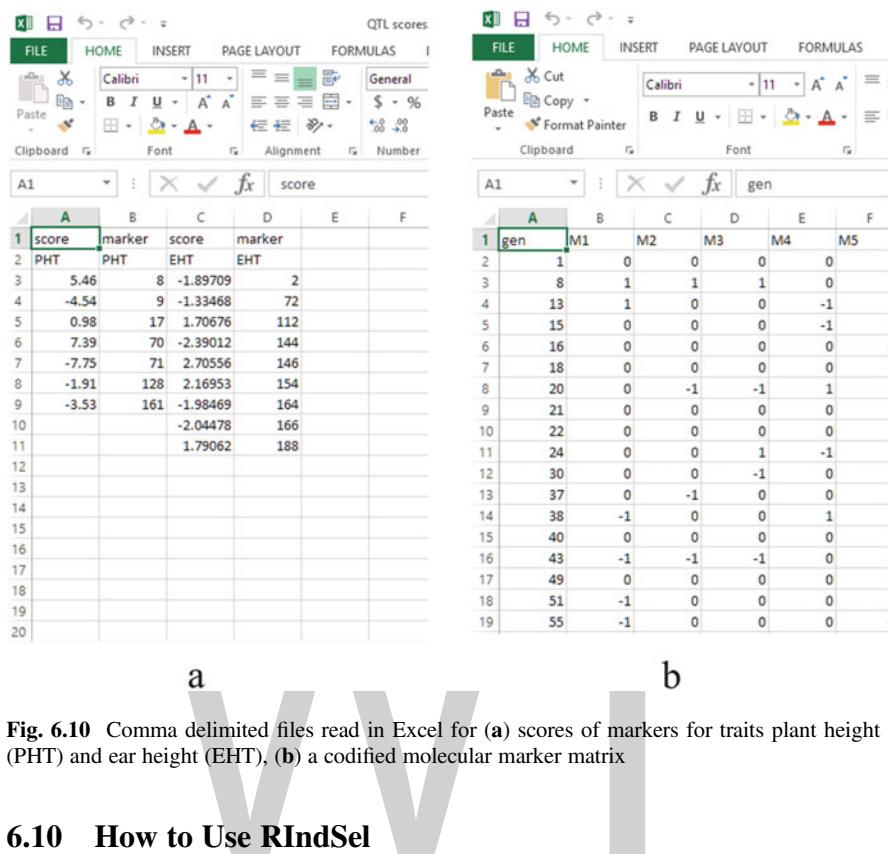
1. Whole molecular markers matrix (green arrow).
2. Marker scores or estimated quantitative trait loci values (red arrow).

Marker scores can be obtained by making a regression of the phenotypic values on a codified molecular markers matrix. The file can be created in Excel and must have the score with its respective marker for each trait; this file is saved with a .csv extension. An example of how these kinds of files must be generated is shown in Fig. 6.10a.

To calculate the scores in an F2 population, it is important for the molecular marker to have previously been codified as -1, 0, and 1 for genotypes aa, Aa, and AA respectively. When data come from an recombinant inbred line population, the molecular marker should be codified as -1 and 1 for homozygous genotype aa and AA respectively. In the genomic selection indices (LGSi) context (see Chap. for details), it is only necessary to codify the molecular marker matrix (Fig. 6.10b), as these indices do not require a marker score.



**Fig. 6.9** Example of parameters that could be used to calculate a molecular selection index



**Fig. 6.10** Comma delimited files read in Excel for (a) scores of markers for traits plant height (PHT) and ear height (EHT), (b) a codified molecular marker matrix

## 6.10 How to Use RIndSel

The use of RIndSel can be illustrated with an example from the Smith linear phenotypic selection index (LPSI) (Smith 1936, see Chap. 2 for details). Figure 6.11 shows the phenotypic data (Fig. 6.11a), together with the file of economic weights (Fig. 6.11b). Three simulated traits (T1, T2, and T3) described in Chap. were used. T1 and T3 are positive (economic value = 1), whereas trait T2 is negative (economic value = -1). It is important to remember that all data files must be saved in comma delimited format (\*.csv).

After the data and economic weights files have been generated, the data need to be loaded into RIndSel; thus, it is important to be able to find the pathway to where the files are located (e.g., “C://Book/datafile/C1\_PSI\_05\_Phen.csv”). Once the data file has been located, it must be uploaded, which can be done by clicking on the file, causing it to automatically begin this process. It is then possible go to the second module (Fig. 6.12) and select subsequent parameters from the menus. In this case, **Selection Index: Smith; Percent: 5; Weights: here we must look for where the economic weights are**, for example “C://Book/datafile/C1\_PSI\_05\_Phen\_Weights.csv.” Once this file has been located, it must be selected by clicking.

**Table (a) Data:**

	A	B	C	D	E	F
1	rep	block	entry	T1	T2	T3
2	1	1	1	165.8352	41.5268	34.4409
3	1	1	2	150.7461	53.7737	35.0019
4	1	1	3	142.1326	50.8256	37.4618
5	1	1	4	166.9536	47.2296	31.9817
6	1	1	5	152.4967	49.6705	31.3214
7	1	1	6	179.5407	39.5706	29.5517
8	1	1	7	152.5788	54.2718	30.0375
9	1	1	8	154.8235	37.7201	29.5861
10	1	1	9	155.0756	42.5305	35.3189
11	1	1	10	161.504	43.849	37.3255
12	1	1	11	140.6271	42.5848	36.4823
13	1	1	12	168.0173	46.9266	39.2436
14	1	1	13	147.714	57.2239	31.8897
15	1	1	14	142.9153	48.7155	28.3857

**Table (b) Data:**

	A	B	C
1	trait	EconomicWeight	
2	T1	1	
3	T2	-1	
4	T3	1	
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			

a

b

**Fig. 6.11** Simulated data from Chap. 2 with (a) array in an alpha-lattice and (b) economic weights required to test the Smith linear phenotypic selection index (LPSI)

**RindSel (Selection Index with R for Windows)**

**Selection Index**: Index: Smith, Percent: 5, Correlation:

**Experimental Design**: Lattice

**Weights**: File: C1\_PSI\_05\_Phen Weights.csv

trait	EconomicWeight
T1	1
T2	-1
T3	1

**Variable Selection**: Location: --select--, Replicate: rep, Block: block, Genotype: entry, Output folder: Smith-Simulated

**Response Variables**: Selected: T1, T2, T3

**Welcome to Rind Sel (Selection Index with R for Windows), Version 1.0 (2016-11-30)**  
Copyright © 2016 Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT).

**Authors:**  
Ángela Sanchez  
Sergio Perez  
Gregorio Alvarado  
Jesus Cárdenas  
Francisco Rodriguez  
Inaki Ormea

rep	block	entry	T1	T2	T3
1	1	1	165.8352	41.5268	34.4409
1	1	2	150.7461	53.7737	35.0019
1	1	3	142.1326	50.8256	37.4618
1	1	4	166.9536	47.2296	31.9817
1	1	5	152.4967	49.6705	31.3214

**Fig. 6.12** Example of filling in a phenotypic selection index without restrictions



After the selection index windows are filled, the following menu is called: **Experimental design**, which allows the user to select the appropriate design – (for example, a lattice). To select the design variables, the user must navigate to the **Variable Selection**. In this example, the experiment has only one location, and the following should be selected: *rep* as **Replicate**, *block* as **Block** and *entry* as **Genotype**. An output name of the index must be assigned by writing its name in the **Box Output** folder, which is below the **Variable Selection** menu. For the Smith LPSI, the name chosen was *SmithSimulated*. Finally, the **Response Variables** menu should be filled by selecting the traits T1, T2, and T3.

## 6.11 RIndSel Output

This section explains the structure of the RIndSel output. First, RIndSel presents the genotypic variance–covariance matrix and the phenotypic variance–covariance matrix (Table 6.1). In addition, when the selection index involves molecular data, RIndSel presents an additional molecular variance–covariance matrix, which contains the additive variability associated with the markers (Table 6.2).

RIndSel also presents a table with the estimated values of the index parameters (Table 6.3). These estimates are the covariance of the selection index, the variance of the selection index, the net genetic merit (breeding value), the correlation between the selection index and the net genetic merit, the selection response, and the heritability of the index.

Additional results are presented in Table 6.4, which show the ranked selected individuals; this ranking was done as a function of the estimated selection index values. Table 6.4 also presents the means of the traits of the selected individuals; the means of the traits of the total population; the selection differential.

**Table 6.1** Matrices of variance–covariance deployed by RIndSel

rownames	T1	T2	T3
Genetic covariance matrix			
T1	36.21	-12.93	8.35
T2	-12.93	13.04	-3.40
T3	8.35	-3.40	9.96
Phenotypic covariance matrix			
T1	62.50	-12.74	8.53
T2	-12.74	17.52	-3.38
T3	8.53	-3.38	12.31

**Table 6.2** Molecular covariance matrix

rownames	T1	T2	T3
T1	62.50	-12.74	8.53
T2	-12.74	17.52	-3.38
T3	8.53	-3.38	12.31

**Table 6.3** Estimated selection index parameters given by the RIndSel output

Parameter	Output
Covariance between the selection index and the breeding value	86.7185
Variance of the selection index	86.7185
Variance of the breeding value	108.5746
Correlation between the selection index and the breeding value	0.8937
Response to selection	16.3431
Heritability	0.8168

**Table 6.4** Values of the three traits for selected individuals and the values of the Smith linear phenotypic selection index, means and gains with  $k = 5\%$ 

rownames	T1	T2	T3	Index
Entry 353	189.68	38.16	36.13	103.97
Entry 370	178.27	34.38	37.79	103.45
Entry 480	174.84	42.72	45.12	100.66
Entry 300	177.38	39.15	40.34	100.65
Entry 273	181.18	35.94	35.14	100.52
Entry 275	167.94	36.82	42.20	99.92
Entry 148	173.37	37.07	39.62	99.86
Entry 137	185.48	46.48	42.55	99.77
Entry 351	173.79	38.38	40.52	99.68
Entry 236	182.85	37.88	34.96	99.20
Entry 217	175.13	38.48	39.16	98.84
Entry 356	171.09	39.60	41.98	98.47
Entry 167	175.39	38.73	37.73	97.17
Entry 230	169.73	37.10	38.69	96.80
Entry 243	171.90	41.53	41.45	96.29
Entry 55	170.02	36.92	37.76	96.15
Entry 68	172.56	37.18	36.70	96.13
Entry 36	175.80	38.86	36.34	95.75
Entry 164	173.61	38.37	36.42	95.14
Entry 140	170.53	42.52	41.97	95.05
Entry 146	177.40	39.64	35.50	94.89
Entry 432	174.01	40.73	38.26	94.84
Entry 378	176.62	42.69	38.47	94.44
Entry 288	172.14	39.37	37.26	94.23
Entry 386	175.77	42.89	38.81	94.13
Mean of selected individuals	175.46	39.26	38.83	
Mean of all individuals	161.88	45.19	34.39	
Selection differential	13.58	-5.92	4.44	
Expected genetic gain 5%	9.51	-5.48	4.22	



**Table 6.5** First 20 values of the entries and their corresponding selection index for all individuals when three traits are analyzed

<i>rownames</i>	T1	T2	T3	Index
Entry 1	164.46	39.63	34.66	86.81
Entry 2	144.39	50.77	34.65	63.82
Entry 3	157.48	48.04	37.90	77.52
Entry 4	167.30	47.98	30.49	74.97
Entry 5	164.11	49.89	32.03	72.85
Entry 6	166.26	40.44	29.93	81.81
Entry 7	154.59	52.22	30.31	63.22
Entry 8	160.00	42.91	31.23	77.12
Entry 9	158.51	46.32	34.52	76.25
Entry 10	163.63	45.43	35.73	81.35
Entry 11	156.16	46.75	35.58	75.62
Entry 12	171.38	41.17	35.13	89.52
Entry 13	153.17	54.18	36.23	66.79
Entry 14	149.89	52.33	31.13	61.39
Entry 15	159.63	49.01	31.72	70.96
Entry 16	160.70	42.51	32.99	79.85
Entry 17	157.07	45.49	28.40	69.68
Entry 18	167.50	41.69	36.73	88.55
Entry 19	159.17	50.60	36.25	73.93
Entry 20	161.80	46.58	37.33	80.85

and the expected genetic gain per trait. Selected individuals can be identified by the first column called “*rownames*,” as columns 2 to 4 contain the best linear and unbiased estimator for each mean trait. Finally, column 5 presents the estimated selection index values.

Comparison between means of selected individuals and all individuals is done by selection differential, where in general traits whose economic weight was 1 are positive, whereas those traits whose economic weight was -1 are negative. The expected genetic gain is an inferential tool based on normal distribution that depends on the percentage of selected individuals and gives the estimated index expected genetic gain per trait.

Finally, Table 6.5 shows the best linear and unbiased estimators for all individuals accompanied by its respective selection index. In this case, only the first 20 individuals were included. This table output is important, because on some occasions, it is necessary to determine the specific behavior of a group of genotypes that may not have a good performance, even though they have shown a good general performance from previous analyses. Another possibility is that a group of individuals belongs to a specific population group; thus, it is possible to select the best individual for this population group.



## References

- Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Crossa J (2008a) A restricted selection index method based on eigenanalysis. *J Agric Biol Environ Stat* 13(4):421–438
- Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Benítez-Riquelme I, Crossa J (2008b) A molecular selection index method based on eigenanalysis. *Genetics* 180:547–557
- Cerón-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. *Genes/Genomes/Genetics* 5:2155–2164
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756
- Pacheco A, Pérez S, Alvarado G, Ceron J, Rodríguez F, Crossa J, Burgueño J (2017) RIIndSel: selection indices for plant breeding. [hdl:11529/10854](https://hdl.handle.net/11529/10854), CIMMYT Research Data & Software Repository Network, V1
- SAS Institute (2017) SAS user's guide: statistics module. Version 9.4. Ed. Cary, NC
- Smith HF (1936) A discriminant function for plant selection. In: Papers on quantitative genetics and related topics. Department of Genetics, North Carolina State College, Raleigh, NC, pp 466–476

