

# Chapter 7

## Linear Phenotypic Eigen Selection Index Methods



**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.

### 7.1 The Linear Phenotypic Eigen Selection Index Method

The conditions described in Chap. 2 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.

### 7.1.1 The ESIM Parameters

The theoretical ESIM selection response can be written as

$$R_I = k_I \sigma_H \rho_{HI}, \quad (7.1)$$

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  $H$ ,  $\rho_{HI} = \frac{\mathbf{w}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}\mathbf{b}'\mathbf{P}\mathbf{b}}}$  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. (7.1), 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

$$f(\mathbf{b}, \mathbf{w}, \mu, \phi) = (\mathbf{w}'\mathbf{C}\mathbf{b})^2 - \mu(\mathbf{b}'\mathbf{P}\mathbf{b} - \sigma_I^2) - \phi(\mathbf{w}'\mathbf{C}\mathbf{w} - \sigma_H^2) \quad (7.2)$$

with respect to  $\mathbf{b}$ ,  $\mathbf{w}$ ,  $\mu$ , and  $\phi$ , where  $\mu$  and  $\phi$  are Lagrange multipliers. The derivative results of Eq. (7.2) with respect to  $\mathbf{b}$ ,  $\mathbf{w}$ ,  $\mu$ , and  $\phi$  are:

$$(\mathbf{w}'\mathbf{C}\mathbf{b})\mathbf{C}\mathbf{w} - \mu\mathbf{P}\mathbf{b} = \mathbf{0}, \quad (7.3)$$

$$(\mathbf{w}'\mathbf{C}\mathbf{b})\mathbf{C}\mathbf{b} - \phi\mathbf{C}\mathbf{w} = \mathbf{0}, \quad (7.4)$$

$$\mathbf{b}'\mathbf{P}\mathbf{b} = \sigma_I^2 \text{ and } \mathbf{w}'\mathbf{C}\mathbf{w} = \sigma_H^2, \quad (7.5)$$

respectively, where Eq. (7.5) 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. (7.3) and (7.4) can be written as

$$\theta^{1/2} \mathbf{C}\mathbf{w} - \frac{\theta}{\sigma_I^2} \mathbf{P}\mathbf{b} = \mathbf{0} \quad (7.6)$$

and

$$\theta^{1/2} \mathbf{C}\mathbf{b} - \frac{\theta}{\sigma_H^2} \mathbf{C}\mathbf{w} = \mathbf{0}, \quad (7.7)$$

respectively. Equation (7.6) is equal to  $\mathbf{C}\mathbf{w} = \frac{\theta^{1/2}}{\sigma_I^2} \mathbf{P}\mathbf{b}$ ; then, vector  $\mathbf{w}$  can be written as

$$\mathbf{w}_E = \frac{\theta^{1/2}}{\sigma_I^2} \mathbf{C}^{-1} \mathbf{P}\mathbf{b}. \quad (7.8)$$

By the result of Eq. (7.8), 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. (7.7), we get

$$\mathbf{C}\mathbf{b} - \frac{(\mathbf{w}'_E \mathbf{C}\mathbf{b})^2}{\sigma_I^2 \sigma_{H_E}^2} \mathbf{P}\mathbf{b} = \mathbf{0}, \quad (7.9)$$

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. (7.9) can be written as

$$(\mathbf{P}^{-1} \mathbf{C} - \lambda_E^2 \mathbf{I}) \mathbf{b}_E = \mathbf{0}, \quad (7.10)$$

and the optimized ESIM index is  $I_E = \mathbf{b}'_E \mathbf{y}$ . Note that in Eq. (7.10)  $\mathbf{P}^{-1} \mathbf{C}$  is the multi-trait heritability. By Eqs. (7.8) and (7.10), 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

$$\rho_{H_E I_E} = \frac{\sigma_{I_E}}{\sigma_{H_E}}, \quad (7.11)$$

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. (7.11) 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. (7.10) 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 (7.10) can also be written as  $\mathbf{Cb}_E = \lambda_E^2 \mathbf{Pb}_E$ , 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  $h_I^2 = \frac{\mathbf{b}'\mathbf{Cb}}{\mathbf{b}'\mathbf{Pb}}$ , then another way of writing Eq. (7.1) is

$$R_I = k_I \sigma_I h_I^2 = k_I \frac{\mathbf{b}'\mathbf{Cb}}{\sqrt{\mathbf{b}'\mathbf{Pb}}}, \quad (7.12)$$

which is similar to the univariate breeder's equation (see Chap. 2, Eq. 2.4). All the parameters of Eq. (7.12) were defined earlier.

The derivative of the ratio  $\frac{\mathbf{b}'\mathbf{Cb}}{\sqrt{\mathbf{b}'\mathbf{Pb}}}$  (Eq. 7.12) with respect to  $\mathbf{b}$  can be written as  $2(\mathbf{b}'\mathbf{Pb})^{1/2}\mathbf{Cb} - (\mathbf{b}'\mathbf{Pb})^{-1/2}(\mathbf{b}'\mathbf{Cb})\mathbf{Pb} = \mathbf{0}$ , and, except by a proportionality constant, the result is

$$(\mathbf{P}^{-1}\mathbf{C} - h_I^2 \mathbf{I})\mathbf{b}_E = \mathbf{0}, \quad (7.13)$$

where  $h_{I_E}^2 = \frac{\mathbf{b}'_E \mathbf{Cb}_E}{\mathbf{b}'_E \mathbf{Pb}_E}$  is the maximized ESIM heritability. Let  $\lambda_E^2 = \rho_E^2 = h_{I_E}^2$ , then Eq. (7.13) is equal to Eq. (7.10) and can be written as  $\mathbf{b}'_E \mathbf{Cb}_E = \lambda_E^2 \mathbf{b}'_E \mathbf{Pb}_E$ , whence the maximized  $\rho_E^2$  in terms of  $h_{I_E}^2$  is

$$\rho_E^2 = \frac{\mathbf{b}'_E \mathbf{Cb}_E}{\mathbf{b}'_E \mathbf{Pb}_E}, \quad (7.14)$$

which should give a equivalent result to that of Eq. (7.11).

By Eq. (7.11) and  $\sigma_{H_E} = \sqrt{\mathbf{b}'_E \mathbf{P} \mathbf{C}^{-1} \mathbf{Pb}_E}$ , the maximized ESIM selection response and expected genetic gain per trait can be written as

$$R_E = k_I \sqrt{\mathbf{b}'_E \mathbf{Pb}_E} \quad (7.15)$$

and

$$\mathbf{E}_E = k_I \frac{\mathbf{Cb}_E}{\sqrt{\mathbf{b}'_E \mathbf{Pb}_E}}, \quad (7.16)$$

respectively. Equations (7.15) and (7.16) 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. (7.15) can be written as  $R_E = k_I \lambda_E$ . When  $\sigma_{H_E}^2 \neq 1$  Eq. (7.15) 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. (7.13) can be written as  $\mathbf{T} \mathbf{I} \mathbf{b}_E = \lambda_E^2 \mathbf{I} \mathbf{b}_E$ , 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. (7.10) and (7.13) is

$$(\mathbf{T}_2 - \lambda_E^2 \mathbf{I}) \boldsymbol{\beta} = \mathbf{0}, \quad (7.17)$$

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. (7.13) 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. (7.17). Vector  $\boldsymbol{\beta} = \mathbf{F} \mathbf{b}_E$  can substitute  $\mathbf{b}_E$  in Eqs. (7.15) and (7.16); and in this case, the optimized ESIM index should be written as  $I_E = \boldsymbol{\beta}' \mathbf{y}$ .

### 7.1.2 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 Chap. 2 for details). In Eq. (7.11), 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,

$$\sigma_{I_E}^2 = \mathbf{b}'_E \mathbf{P} \mathbf{b}_E \text{ and } \sigma_{H_E I_E} = \mathbf{w}'_E \mathbf{C} \mathbf{b}_E = \mathbf{b}'_E \mathbf{P} \mathbf{C}^{-1} \mathbf{C} \mathbf{b}_E = \mathbf{b}'_E \mathbf{P} \mathbf{b}_E, \quad (7.18)$$

respectively; that is,  $\sigma_{I_E}^2 = \sigma_{H_E I_E}$ . By Eq. (7.18), the VPE of the ESIM can be written as

$$E[(H_E - I_E)^2] = \sigma_{H_E}^2 + \sigma_{I_E}^2 - 2\sigma_{H_E I_E} = \sigma_{H_E}^2 - \sigma_{I_E}^2 = (1 - \rho_E^2) \sigma_{H_E}^2. \quad (7.19)$$

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  $\sigma_{I_E}^2$ , 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

$$\sigma_{I_E}^2 = \rho_E^2 \sigma_{H_E}^2, \quad (7.20)$$

and the relative mean squared effect can be measured by  $\rho_E^2$  (Anderson 2003). If in Eq. (7.20)  $\rho_E^2 = 1$ ,  $\sigma_{I_E}^2 = \sigma_{H_E}^2$ , 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.

### 7.1.3 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. 7.1) 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  $\mathbf{y}$  ( $I_E = \mathbf{b}'_{E_j} \mathbf{y}$ ) and the  $j$ th linear combination of  $\mathbf{g}$  ( $H_E = \mathbf{w}'_{E_j} \mathbf{g}$ ) is the  $j$ th canonical correlation ( $\lambda_j$ ) value obtained from equation  $(\mathbf{P}^{-1} \mathbf{C} - \lambda_j^2 \mathbf{I}) \mathbf{b}_{E_j} = \mathbf{0}$ , where  $\mathbf{b}_{E_j}$  is the  $j$ th *canonical vector* ( $j = 1, 2, \dots, t$ ) of matrix  $\mathbf{P}^{-1} \mathbf{C}$ , and  $\mathbf{w}_{E_j} = \mathbf{C}^{-1} \mathbf{P} \mathbf{b}_{E_j}$ . Thus, in the canonical correlation context,  $I_E = \mathbf{b}'_{E_j} \mathbf{y}$  and  $H_E = \mathbf{w}'_{E_j} \mathbf{g}$  are *canonical variables*.

In the ESIM, the first eigenvector ( $\mathbf{b}_{E_1}$ ) of matrix  $\mathbf{P}^{-1} \mathbf{C}$  should be used on  $I_E = \mathbf{b}'_{E_1} \mathbf{y}$ ; the first eigenvalue ( $\lambda_1^2$ ) and  $\mathbf{b}_{E_1}$  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.

### 7.1.4 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 Chap. 2. With matrices  $\widehat{\mathbf{C}}$  and  $\widehat{\mathbf{P}}$ , we constructed matrix  $\widehat{\mathbf{T}} = \widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}}$  and equation

$$(\widehat{\mathbf{T}} - \widehat{\lambda}_{Ej}^2 \mathbf{I})\widehat{\mathbf{b}}_{Ej} = \mathbf{0}, \quad (7.21)$$

$j = 1, 2, \dots, t$ , 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

$$\widehat{\mathbf{T}} = \mathbf{V}_1 \mathbf{L}^{1/2} \mathbf{V}_2', \quad (7.22)$$

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 ( $\widehat{\lambda}_{E1}^2 \geq \widehat{\lambda}_{E2}^2 \geq \dots \geq \widehat{\lambda}_{Eq}^2 > 0$ ) 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  $\widehat{\lambda}_{E1}^2 \geq \widehat{\lambda}_{E2}^2 \geq \dots \geq \widehat{\lambda}_{Eq}^2 > 0$  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  $\widehat{\mathbf{b}}_{E1}$  and  $\widehat{\lambda}_{E1}^2$  of the first eigenvector  $\mathbf{b}_{E1}$  and the first eigenvalue  $\lambda_{E1}^2$  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  $\widehat{\lambda}_{E1}^2$  and  $\widehat{\mathbf{b}}_{E1}$  in the ESIM context can be obtained from

$$(\widehat{\mathbf{T}}\widehat{\mathbf{T}}' - \widehat{\mu}_j \mathbf{I})\widehat{\mathbf{b}}_{Ej} = \mathbf{0}, \quad (7.23)$$

where  $\widehat{\mu}_j = \widehat{\lambda}_{Ej}^4$ ,  $j = 1, 2, \dots, t$ . In the asymptotic context,  $\widehat{\lambda}_{E1}^2$  and  $\widehat{\mathbf{b}}_{E1}$  are consistent and unbiased estimators (Anderson 2003).

The latter results allow the ESIM index ( $I_E = \mathbf{b}'_E \mathbf{y}$ ) as  $\widehat{I}_E = \widehat{\mathbf{b}}'_{E1} \mathbf{y}$  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  $\mathbf{b}_{E_j}$  ( $\widehat{\mathbf{b}}_{E_j}$ ) has multivariate normal distribution with expectation  $E(\widehat{\mathbf{b}}_{E_j}) = \mathbf{b}_{E_j}$  and variance

$$\text{Var}(\widehat{\mathbf{b}}_{E_j}) = \frac{1}{2n} \mathbf{b}_{E_j} \mathbf{b}'_{E_j} + \frac{1}{n} (1 - \lambda_{E_j}^2) \sum_{i \neq j}^t \frac{\lambda_{E_j}^2 + \lambda_{E_i}^2 - 2\lambda_{E_i}^2 \lambda_{E_j}^2}{(\lambda_{E_i}^2 - \lambda_{E_j}^2)^2} \mathbf{b}_{E_i} \mathbf{b}'_{E_i}, \quad (7.24)$$

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

$$\text{Cov}(\widehat{\mathbf{b}}_{E_i}, \widehat{\mathbf{b}}_{E_j}) = \frac{(1 - \lambda_{E_j}^2)(1 - \lambda_{E_i}^2)(\lambda_{E_i}^2 + \lambda_{E_j}^2)}{n(\lambda_{E_i}^2 - \lambda_{E_j}^2)^2} \mathbf{b}_{E_j} \mathbf{b}'_{E_i}, \quad (7.25)$$

where  $n$  is the number of individuals or genotypes (Anderson 1999). The variance of  $\widehat{\mathbf{b}}_{E_j}$  and the covariance between  $\widehat{\mathbf{b}}_{E_i}$  and  $\widehat{\mathbf{b}}_{E_j}$  depend not only on  $n$ , but also on eigenvalues  $\lambda_{E_i}^2$  and  $\lambda_{E_j}^2$ . Suppose that  $\lambda_{E_j}^2 > \lambda_{E_i}^2$ ; then, when  $\lambda_{E_j}^2$  is very close to 1,  $\text{Var}(\widehat{\mathbf{b}}_{E_j}) \approx \frac{1}{2n} \mathbf{b}_{E_j} \mathbf{b}'_{E_j}$  (“ $\approx$ ” denotes an approximation) and  $\text{Cov}(\widehat{\mathbf{b}}_{E_i}, \widehat{\mathbf{b}}_{E_j})$  is very close to 0. By the result of Eq. (7.24), the variance of the first eigenvector ( $\widehat{\mathbf{b}}_{E_1}$ ) of  $\widehat{\mathbf{P}}^{-1} \widehat{\mathbf{C}}$  can be written as  $\text{Var}(\widehat{\mathbf{b}}_{E_1}) = \frac{1}{2n} \mathbf{b}_{E_1} \mathbf{b}'_{E_1} + \frac{1}{n} (1 - \lambda_{E_1}^2) \sum_{j=2}^t \frac{\lambda_{E_1}^2 + \lambda_{E_j}^2 - 2\lambda_{E_1}^2 \lambda_{E_j}^2}{(\lambda_{E_1}^2 - \lambda_{E_j}^2)^2} \mathbf{b}_{E_j} \mathbf{b}'_{E_j}$ . If the first eigenvalue  $\lambda_{E_1}^2$  of  $\mathbf{P}^{-1} \mathbf{C}$  is very close to 1 ( $\lambda_{E_1}^2 \approx 1$ ),  $\text{Var}(\widehat{\mathbf{b}}_{E_1}) = \frac{1}{2n} \mathbf{b}_{E_1} \mathbf{b}'_{E_1}$  and  $\text{Cov}(\widehat{\mathbf{b}}_{E_1}, \widehat{\mathbf{b}}_{E_j}) \approx 0$ .

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

$$\text{Var}(\widehat{\lambda}_{E_j}) \approx \frac{(1 - \lambda_{E_j}^2)^2}{n}, \quad (7.26)$$

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

$$\text{Var}(\widehat{\lambda}_{E_j}^2) \approx \frac{4\lambda_{E_j}^2 (1 - \lambda_{E_j}^2)^2}{n}. \quad (7.27)$$



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

Equation (7.26) 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  $Var(\widehat{R}_E) \approx \frac{k_I^2(1-\lambda_{E1}^2)^2}{n}$  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.

### 7.1.5 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

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

tively. 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.,  $\widehat{\boldsymbol{\beta}} = \mathbf{F}\widehat{\mathbf{b}}_{E_1}$ , 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  $\widehat{\mathbf{T}} = \widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}} = \begin{bmatrix} 0.1102 & -0.0405 & -0.0280 \\ -0.0390 & 0.0864 & -0.0184 \\ -0.1833 & 0.0517 & 0.4115 \end{bmatrix}$  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  $\widehat{T}_E = -0.1701\text{RL} + 0.0259\text{SM} + 0.9851\text{EW}$ .

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

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

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  $\widehat{\boldsymbol{\beta}} = \mathbf{F}\widehat{\mathbf{b}}_{E_1}$  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

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

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

According to the theory of similar matrices (Harville 1997), the estimated maximized ESIM accuracy,  $\hat{\lambda}_{E_1} = 0.6782$ , 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  $\hat{\lambda}_{E_1} = 0.6782$  to LPSI accuracy  $\hat{\rho}_S = 0.362$ , i.e.,  $\frac{\hat{\lambda}_{E_1}}{\hat{\rho}_S} = \frac{0.6782}{0.362} = 1.873$ , or in percentage terms,  $\hat{p}_E = 100(1.873 - 1) = 87.3$  (see Chap. 5, Eq. 5.17). 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 Chap. 2. 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 7.1 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 7.1 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 Chap. 2 (Table 2.4). The averages of the estimated LPSI and ESIM selection responses were 4.70 and 6.31 respectively.

Table 7.2 presents the estimated ESIM expected genetic gain per trait, accuracy ( $\hat{\rho}_E$ ), and the values  $\hat{p}_E = 100(\hat{\lambda}_E - 1)$ , where  $\hat{\lambda}_E = \hat{\rho}_E / \hat{\rho}_S$  is the ratio of  $\hat{\rho}_E$  to the estimated LPSI accuracy ( $\hat{\rho}_S$ ), expressed as percentages. Table 7.2 also presents the accuracy of the PPG-ESIM and the estimated ratio ( $\hat{p}_{PE}$ ) 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  $\hat{p}_E = 100(\hat{\lambda}_E - 1)$  to compare ESIM efficiency versus LPSI efficiency.

The estimated LPSI expected genetic gains per trait were presented in Chap. 2, Table 2.4. According to the results shown in Table 2.4, the averages of the estimated

**Table 7.1** 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 7.2, 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 7.2). 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 7.1), not with the expected genetic gain per trait, because  $\hat{\lambda}_E = \frac{\hat{\rho}_E}{\hat{\rho}_S} \approx \frac{\hat{R}_E}{\hat{R}_S}$ , 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  $\frac{\hat{\rho}_E}{\hat{\rho}_S} = \frac{\hat{R}_E}{\hat{R}_S}$  is true only when the denominators of both estimated correlations are the same, as in the linear selection indices described in Chaps. 2–6.

**Table 7.2** 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

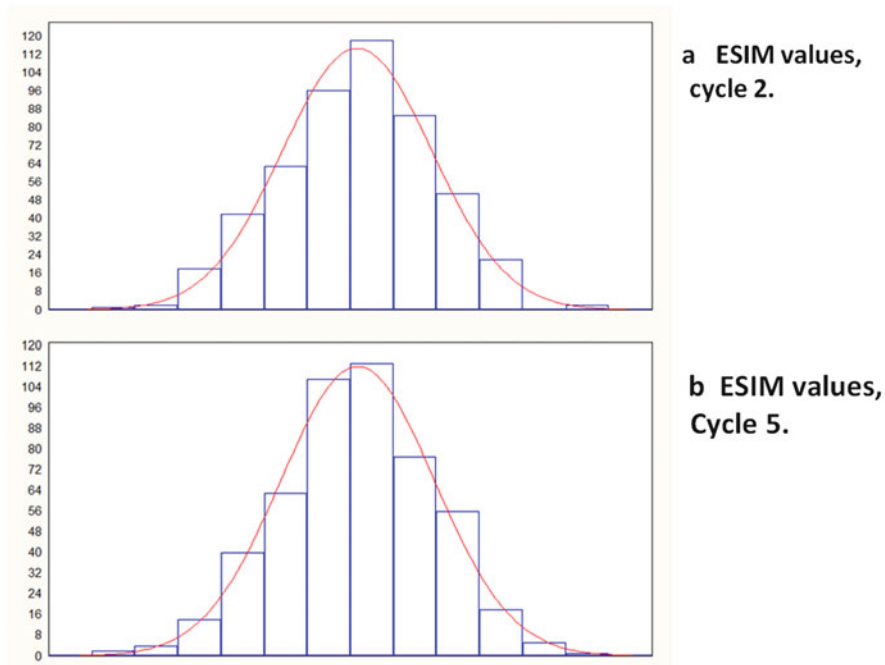
  

Cycle	PPG-ESIM accuracies for one, two, and three predetermined restrictions			$\hat{\rho}_P$ values (in %) for one, two, and three predetermined restrictions		
	1	2	3	1	2	3
1	0.98	0.96	0.99	9.34	8.90	20.99
2	0.96	0.96	0.98	10.94	12.46	25.20
3	0.97	0.97	1.00	10.04	9.71	41.43
4	0.95	0.94	0.99	10.35	13.98	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  $\hat{\rho}_S = \frac{\sqrt{\hat{\mathbf{b}}_S' \hat{\mathbf{P}} \hat{\mathbf{b}}_S}}{\sqrt{\hat{\mathbf{w}}' \hat{\mathbf{C}} \hat{\mathbf{w}}}}$  and  $\hat{\rho}_E = \frac{\sqrt{\hat{\mathbf{b}}_E' \hat{\mathbf{P}} \hat{\mathbf{b}}_E}}{\sqrt{\hat{\mathbf{w}}_E' \hat{\mathbf{C}} \hat{\mathbf{w}}_E}}$ , whereas  $\hat{R}_S = \sqrt{\hat{\mathbf{b}}_S' \hat{\mathbf{P}} \hat{\mathbf{b}}_S}$  and  $\hat{R}_E = \sqrt{\hat{\mathbf{b}}_E' \hat{\mathbf{P}} \hat{\mathbf{b}}_E}$ ; this means that if  $\sqrt{\hat{\mathbf{w}}_E' \hat{\mathbf{C}} \hat{\mathbf{w}}_E} \neq \sqrt{\hat{\mathbf{w}}' \hat{\mathbf{C}} \hat{\mathbf{w}}}$ ,  $\frac{\hat{\rho}_E}{\hat{\rho}_S} \neq \frac{\hat{R}_E}{\hat{R}_S}$ . For the Akbar et al. (1984) data,  $\hat{R}_E = 9.54$  and  $\hat{R}_S = 4.75$ , then  $\frac{\hat{R}_E}{\hat{R}_S} = 2.0$  but  $\frac{\hat{\lambda}_{E1}}{\hat{\rho}_S} = 1.873$ ; that is,  $\frac{\hat{\rho}_E}{\hat{\rho}_S} \approx \frac{\hat{R}_E}{\hat{R}_S}$ , where “ $\approx$ ” indicates an approximation.

Figure 7.1 presents the frequency distribution of 500 estimated ESIM values for cycle 2 (Fig. 7.1a) and cycle 5 (Fig. 7.1b), obtained from one selection cycle for 500 genotypes and four traits simulated in one environment. Figure 7.1a, b indicates that the frequency distribution of the estimated ESIM values approaches normal distribution.



**Fig. 7.1** 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

## 7.2 The Linear Phenotypic Restricted Eigen Selection Index Method

Similar to the RLPSI (see Chap. 2), 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  $Cov(I, \mathbf{g}) = \mathbf{C}\mathbf{b}$  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.,  $Cov(I, \mathbf{U}'\mathbf{g}) = \mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}$ , 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

$$\rho_{HI}^2 = \frac{(\mathbf{w}'\mathbf{C}\mathbf{b})^2}{(\mathbf{w}'\mathbf{C}\mathbf{w})(\mathbf{b}'\mathbf{P}\mathbf{b})}$$

with respect to vectors  $\mathbf{b}$  and  $\mathbf{w}$  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. 7.12) 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. (7.13). Both approaches give the same result, but it is easier to work with the second approach than with the first one.

### 7.2.1 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

$$f(\mathbf{b}, \mathbf{v}') = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}} - \mathbf{v}'\mathbf{U}'\mathbf{C}\mathbf{b} \quad (7.28a)$$

with respect to  $\mathbf{b}$  and  $\mathbf{v}'$ , where  $\mathbf{v}' = [v_1 \ v_2 \ \dots \ v_{r-1}]$  is a vector of Lagrange multipliers. The derivatives of Eq. (7.28a) with respect to  $\mathbf{b}$  and  $\mathbf{v}'$  can be written as

$$2(\mathbf{b}'\mathbf{P}\mathbf{b})^{1/2}\mathbf{C}\mathbf{b} - (\mathbf{b}'\mathbf{P}\mathbf{b})^{-1/2}(\mathbf{b}'\mathbf{C}\mathbf{b})\mathbf{P}\mathbf{b} - \mathbf{C}\mathbf{U}\mathbf{v} = \mathbf{0} \quad (7.28b)$$

and

$$\mathbf{U}'\mathbf{C}\mathbf{b} = \mathbf{0}, \quad (7.29)$$

respectively, where Eq. (7.29) denotes the restriction imposed for maximizing Eq. (7.28a). Using algebraic methods on Eq. (7.28b) similar to those used to obtain Eqs. (7.10) and (7.13), we get

$$\left(\mathbf{K}\mathbf{P}^{-1}\mathbf{C} - h_{I_R}^2\mathbf{I}_t\right)\mathbf{b}_R = \mathbf{0}, \quad (7.30)$$

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. (7.30) can be written as

$$\left(\mathbf{K}\mathbf{P}^{-1}\mathbf{C} - \lambda_R^2\mathbf{I}_t\right)\mathbf{b}_R = \mathbf{0}. \quad (7.31)$$

Thus, the optimized RESIM index is  $I = \mathbf{b}'_R\mathbf{y}$ . The only difference between Eqs. (7.31) and (7.13) is matrix  $\mathbf{K}$ . Equation (7.31) was obtained by Cerón-Rojas et al. (2008) by maximizing  $\rho_{HI}^2$  (Eq. 7.1) 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

$$\mathbf{w}_R = \mathbf{C}^{-1}[\lambda_R \mathbf{Pb}_R + \mathbf{\Psi v}], \quad (7.32)$$

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;  $\mathbf{\Psi} = \mathbf{C}\mathbf{U}$  and  $\mathbf{v} = \lambda_R^{-1}(\mathbf{\Psi}'\mathbf{P}^{-1}\mathbf{\Psi})^{-1}\mathbf{\Psi}'\mathbf{P}^{-1}\mathbf{C}\mathbf{b}_R$ . 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  $\mathbf{\Psi v}$  can be written as  $\mathbf{\Psi v} = \mathbf{\Psi}(\mathbf{\Psi}'\mathbf{P}^{-1}\mathbf{\Psi})^{-1}\mathbf{\Psi}'\mathbf{P}^{-1}\mathbf{C}\mathbf{b}_R = \mathbf{Q}'_R \mathbf{C}\mathbf{b}_R$ , where  $\mathbf{Q}'_R$  is the transpose of matrix  $\mathbf{Q}_R$  described in Eq. (7.30). Thus, another way of writing Eq. (7.32) is

$$\mathbf{w}_R = \mathbf{C}^{-1}[\mathbf{P} + \mathbf{Q}'_R \mathbf{C}] \mathbf{b}_R. \quad (7.33)$$

By Eq. (7.33) and the restriction  $\mathbf{b}'\mathbf{\Psi} = \mathbf{0}$ , 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

$$\sigma_{H_R I_R} = \mathbf{w}'_R \mathbf{C}\mathbf{b}_R = \mathbf{b}'_R \mathbf{P}\mathbf{b}_R + \mathbf{b}'_R \mathbf{Q}'_R \mathbf{C}\mathbf{b}_R = \mathbf{b}'_R \mathbf{P}\mathbf{b}_R, \quad (7.34)$$

where  $\mathbf{b}'_R \mathbf{Q}'_R \mathbf{C}\mathbf{b}_R = 0$  according to the restriction  $\mathbf{b}'\mathbf{\Psi} = \mathbf{0}$ . Equation (7.34) 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

$$\rho_{H_R I_R} = \frac{\sqrt{\mathbf{b}'_R \mathbf{P}\mathbf{b}_R}}{\sqrt{\mathbf{w}'_R \mathbf{C}\mathbf{w}_R}}, \quad (7.35)$$

where  $\mathbf{w}'_R \mathbf{C}\mathbf{w}_R = \sigma_{H_R}^2$  is the variance of  $H_R$ ,  $\mathbf{w}_R = \mathbf{C}^{-1}[\mathbf{P} + \mathbf{Q}'_R \mathbf{C}] \mathbf{b}_R$ ,  $\mathbf{Q}'_R = \mathbf{\Psi}(\mathbf{\Psi}'\mathbf{P}^{-1}\mathbf{\Psi})^{-1}\mathbf{\Psi}'\mathbf{P}^{-1}$ , and  $\mathbf{\Psi} = \mathbf{C}\mathbf{U}$ . 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  $\sigma_{I_R}^2 = \mathbf{b}'_R \mathbf{P}\mathbf{b}_R = \mathbf{b}'_E \mathbf{P}\mathbf{b}_E = \sigma_{I_E}^2$ , the variance of  $I_E$ . Hereafter, to simplify the notation, we write Eq. (7.35) 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

$$R_R = k_I \sqrt{\mathbf{b}'_R \mathbf{P}\mathbf{b}_R} \quad (7.36)$$

and

$$\mathbf{E}_R = k_I \frac{\mathbf{C}\mathbf{b}_R}{\sqrt{\mathbf{b}'_R \mathbf{P}\mathbf{b}_R}}, \quad (7.37)$$



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. (7.36) and (7.37), and in  $I_R = \mathbf{b}'_R \mathbf{y}$ .

Equation (7.36) 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. (7.36) 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. (7.34) 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

$$E[(H_R - I_R)^2] = \sigma_{H_R}^2 + \sigma_{I_R}^2 - 2\sigma_{H_R I_R} = \sigma_{H_R}^2 - \sigma_{I_R}^2 = (1 - \rho^2) \sigma_{H_R}^2. \quad (7.38)$$

Statistical properties associated with the ESIM and described in Sect. 7.1.2 are also valid for the RESIM.

## 7.2.2 Estimating the RESIM Parameters

We can estimate the RESIM parameters in a similar manner to the ESIM parameters in Sect. 7.1.4. 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

$$(\widehat{\mathbf{S}}_R \widehat{\mathbf{S}}'_R - \widehat{\mu}_{R_j} \mathbf{I}_t) \widehat{\mathbf{b}}_{R_j} = \mathbf{0}, \quad (7.39)$$

where  $\widehat{\mu}_{R_j} = \widehat{\lambda}_{R_j}^4$ ,  $j = 1, 2, \dots, t$ . 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  $\widehat{\mathbf{E}}_R = k_I \frac{\widehat{\mathbf{C}} \widehat{\mathbf{b}}_{R_1}}{\sqrt{\widehat{\mathbf{b}}'_{R_1} \widehat{\mathbf{P}} \widehat{\mathbf{b}}_{R_1}}}$  respectively, whereas the estimator of the maximized RESIM accuracy is  $\widehat{\lambda}_{R_1}$ .

## 7.2.3 Numerical Examples

We compare the RLPSI results with those of the RESIM using the Akbar et al. (1984) data described in Sect. 7.1.5. We restrict the trait RL (number of eggs) in both indices. In Chap. 3, Sect. 3.1.3, we indicated how to construct matrix  $\mathbf{U}'$  and, in Sect.

3.1.4 of the same chapter, we described how to obtain matrix  $\widehat{\mathbf{K}} = [\mathbf{I}_t - \widehat{\mathbf{Q}}]$  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% ( $k_I = 1.755$ ) 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. 7.1.5 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  $\widehat{\mathbf{S}}_R = \mathbf{V}_{R1}\mathbf{L}_R^{1/2}\mathbf{V}_{R1}'$ .

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  $\widehat{\mathbf{b}}'_{R1} = [0.0500 \quad -0.1446 \quad 0.9882]$ ,  $\widehat{I}_R = 0.0500\text{RL} - 0.1446\text{SM} + 0.9882\text{EW}$ ,

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

respectively.

The estimated RLPSI selection response was  $\widehat{R} = \frac{53.01}{34.25} = 1.55$ ; thus, the estimated RESIM selection response was higher than the estimated RLPSI response. In addition, the estimated RLPSI expected genetic gain per trait was  $\widehat{\mathbf{E}}' = [0 \quad -0.71 \quad 2.96]$ , 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  $\widehat{\lambda}_{E_1} = 0.5798$  to the RLPSI accuracy  $\widehat{\rho} = 0.26$ , i.e.,  $\frac{\widehat{\lambda}_{R_1}}{\widehat{\rho}_S} = \frac{0.5798}{0.26} = 2.23$ , or in percentage terms,  $\widehat{\rho}_E = 100(2.23 - 1) = 123$  (see Chap. 5, Eq. 5.17). 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 7.1 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 7.3 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. 3 (Table 3.3). According to the results shown in Table 3.3 (Chap. 3), 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 7.3, 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 7.3** 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

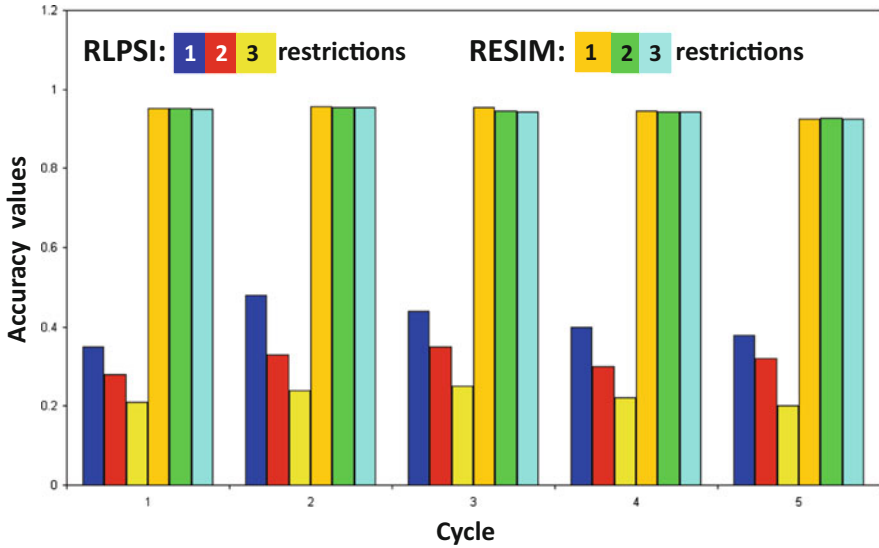
Cycle	Estimated PPG-ESIM expected genetic gain per trait											
	One predetermined restriction				Two predetermined restrictions				Three predetermined restrictions			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
1	7.81	-4.62	3.11	2.21	7.09	-3.04	3.12	2.76	6.62	-2.84	4.73	0.83
2	5.15	-2.98	2.31	3.48	5.41	-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% ( $k_l = 1.755$ ) 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

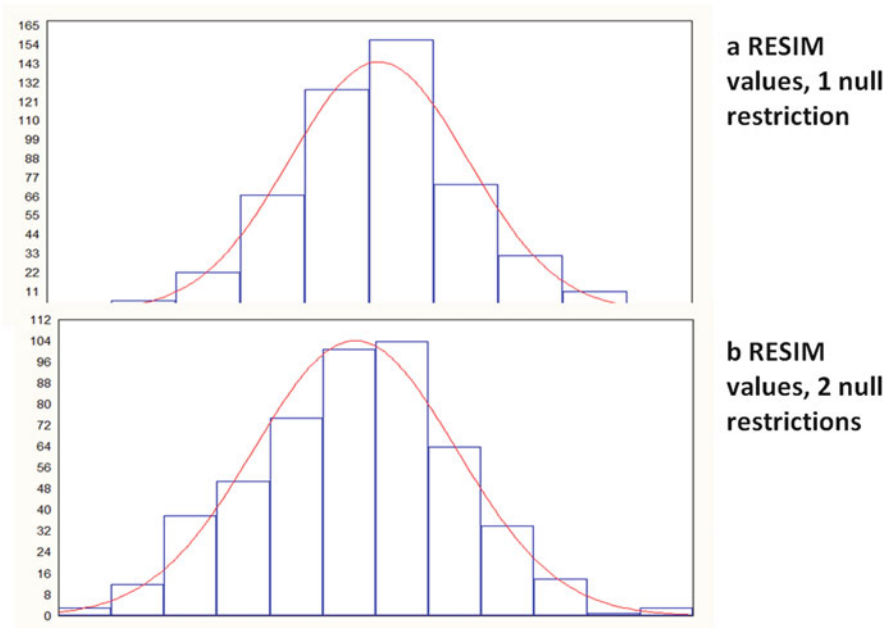
genetic gains per trait was higher than that of the estimated RLPSI expected genetic gains per trait only for trait 4.

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. 7.3. Figure 7.3a presents the frequency distribution of the estimated RESIM values with one null restriction for cycle 2, whereas Fig. 7.3b 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 Chap. 10 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. 7.2** Estimated correlation values between the restricted linear phenotypic selection index (RLPSI) and the net genetic merit ( $H = \mathbf{w}'\mathbf{g}$ ); 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. 7.3** 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

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

In a similar manner to the PPG-LPSI (see Chap. 3), in the PPG-ESIM the breeder pre-sets optimal levels (predetermined proportional gains) on certain traits before the selection is carried out. Let  $\mathbf{d}' = [d_1 \ d_2 \ \cdots \ d_r]$  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

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

$r$  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  $\mathbf{M}' = \mathbf{D}'\mathbf{\Psi}'$ , where  $\mathbf{\Psi}' = \mathbf{U}'\mathbf{C}$  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. 7.12) 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.

### 7.3.1 The PPG-ESIM Parameters

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

$$f(\mathbf{b}, \mathbf{v}') = \frac{\mathbf{b}'\mathbf{C}\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}}} - \mathbf{v}'\mathbf{M}'\mathbf{b} \quad (7.40)$$

with respect to vectors  $\mathbf{b}$  and  $\mathbf{v}'$ , where  $\mathbf{v}' = [v_1 \ v_2 \ \cdots \ v_{r-1}]$  is a vector of Lagrange multipliers. The derivatives of Eq. (7.40) with respect to  $\mathbf{b}$  and  $\mathbf{v}'$  were:

$$2(\mathbf{b}'\mathbf{P}\mathbf{b})^{1/2}\mathbf{C}\mathbf{b} - (\mathbf{b}'\mathbf{P}\mathbf{b})^{-1/2}(\mathbf{b}'\mathbf{C}\mathbf{b})\mathbf{P}\mathbf{b} - \mathbf{M}\mathbf{v} = \mathbf{0} \quad (7.41)$$

and

$$\mathbf{M}'\mathbf{b} = \mathbf{0}, \quad (7.42)$$

respectively, where Eq. (7.42) denotes the restriction imposed for maximizing Eq. (7.40). By using algebraic methods on Eq. (7.41) similar to those used to obtain Eq. (7.10) we get

$$(\mathbf{K}_P\mathbf{P}^{-1}\mathbf{C} - \lambda_P^2\mathbf{I}_t)\mathbf{b}_P = \mathbf{0}, \quad (7.43)$$

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 Chap. 3). Also, vector  $\mathbf{b}_P$  can be transformed as  $\boldsymbol{\beta}_P = \mathbf{F}\mathbf{b}_P$ ; 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 Chap. 3), 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

$$\mathbf{K}_P\mathbf{P}^{-1}\mathbf{C} = [\mathbf{I}_t - \mathbf{P}^{-1}\Psi\mathbf{S}^{-1}\Psi']\mathbf{P}^{-1}\mathbf{C} + \mathbf{A}_P, \quad (7.44)$$

where  $\Psi' = \mathbf{U}'\mathbf{C}$ ,  $\mathbf{A}_P = \delta\boldsymbol{\alpha}'$ ,  $\boldsymbol{\delta} = \mathbf{P}^{-1}\Psi(\Psi'\mathbf{P}^{-1}\Psi)^{-1}\mathbf{d}$ , 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. (7.44) is a mathematical equivalent form of matrix  $\mathbf{K}_P\mathbf{P}^{-1}\mathbf{C}$  and that Eq. (7.44) 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. (7.43) instead of matrix  $\mathbf{K}_P\mathbf{P}^{-1}\mathbf{C}$ .

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

$$\mathbf{w}_P = \mathbf{C}^{-1}[\lambda_P \mathbf{Pb}_P + \mathbf{Mv}_P], \quad (7.45)$$

whence  $H = \mathbf{w}'\mathbf{g}$  can be written as  $H_P = \mathbf{w}'_P \mathbf{g}$ . In Eq. (7.45),  $\lambda_P$  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  $\mathbf{K}_P \mathbf{P}^{-1} \mathbf{C}$ ,  $\mathbf{v}_P = \lambda_P^{-1} (\mathbf{M}' \mathbf{P}^{-1} \mathbf{M})^{-1} \mathbf{M}' \mathbf{P}^{-1} \mathbf{Cb}_P$ ,  $\mathbf{M}' = \mathbf{D}' \mathbf{\Psi}'$ , and  $\mathbf{\Psi}' = \mathbf{U}' \mathbf{C}$ . 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  $I_P = \mathbf{b}'_P \mathbf{y}$  ( $\sigma_{I_P}^2 = \mathbf{b}'_P \mathbf{Pb}_P$ ), that is,  $\sigma_{H_P I_P} = \mathbf{w}'_P \mathbf{Cb}_P = \mathbf{b}'_P \mathbf{Pb}_P$ .

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

$$\rho_{H_P I_P} = \frac{\sqrt{\mathbf{b}'_P \mathbf{Pb}_P}}{\sqrt{\mathbf{w}'_P \mathbf{Cw}_P}}, \quad (7.46)$$

where  $\sigma_{H_P}^2 = \mathbf{w}'_P \mathbf{Cw}_P = \mathbf{b}'_P \mathbf{PC}^{-1} \mathbf{Pb}_P + \mathbf{b}'_P \mathbf{PC}^{-1} \mathbf{Q}'_P \mathbf{Cb}_P$  is the variance of  $H_P$ . When  $\mathbf{D}' = \mathbf{U}'$ ,  $\mathbf{w}'_P \mathbf{Cw}_P = \mathbf{w}'_R \mathbf{Cw}_R$  (the variance of  $H_R$ ), and when  $\mathbf{U}'$  is a null matrix,  $\mathbf{w}'_P \mathbf{Cw}_P = \mathbf{w}'_E \mathbf{Cw}_E$  (the variance of  $H_E$ ). Hereafter, to simplify the notation, we write Eq. (7.46) as  $\rho_P$  or  $\lambda_P$ .

Let  $\boldsymbol{\beta}_P = \mathbf{Fb}_P$  be the PPG-ESIM transformed vector of coefficients by matrix  $\mathbf{F}$ . By Eqs. (7.1) and (7.46), the maximized selection response ( $R_P$ ) and expected genetic gain per trait ( $\mathbf{E}_P$ ) of the PPG-ESIM can be written as

$$R_P = k_I \sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P} \quad (7.47)$$

and

$$\mathbf{E}_P = k_I \frac{\mathbf{C} \boldsymbol{\beta}_P}{\sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P}}, \quad (7.48)$$

respectively, where  $\sqrt{\boldsymbol{\beta}'_P \mathbf{P} \boldsymbol{\beta}_P} = \sigma_{I_P}$  is the standard deviation of the variance of  $I_P = \mathbf{b}'_P \mathbf{y}$ . Equations (7.47) and (7.48) do not require economic weights. When  $\mathbf{F}$  is an identity matrix,  $\boldsymbol{\beta}_P = \mathbf{b}_P$ ,  $I_P = \mathbf{b}'_P \mathbf{y}$ ,  $R_P = k_I \sqrt{\mathbf{b}'_P \mathbf{Pb}_P}$ , and  $\mathbf{E}_P = k_I \frac{\mathbf{Cb}_P}{\sqrt{\mathbf{b}'_P \mathbf{Pb}_P}}$ .

Equation (7.47) can also be written as  $R_P = k_I \sigma_{H_P} \lambda_P$ , where  $\sigma_{H_P} = \sqrt{\mathbf{b}'_P \mathbf{PC}^{-1} \mathbf{Pb}_P + \mathbf{b}'_P \mathbf{PC}^{-1} \mathbf{Q}'_P \mathbf{Cb}_P}$  is the standard deviation of the variance of  $H_P$ , and  $\lambda_P$  is the canonical correlation between  $H_P$  and  $I_P = \mathbf{b}'_P \mathbf{y}$ . When  $\sigma_{H_P} = 1$ , Eq. (7.47) can be written as  $R_P = k_I \lambda_P$ , where  $\lambda_P$  is the covariance between  $I_P = \mathbf{b}'_P \mathbf{y}$  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. 7.46) can be used to construct the ratio of index accuracies. The PPG-ESIM mean square error or the VPE can be obtained as



$$E[(H_P - I_P)^2] = \sigma_{H_P}^2 + \sigma_{I_P}^2 - 2\sigma_{H_P I_P} = \sigma_{H_P}^2 - \sigma_{I_P}^2 = (1 - \rho_P^2)\sigma_{H_P}^2. \quad (7.49)$$

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

### 7.3.2 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

$$(\hat{\mathbf{S}} - \hat{\lambda}_{P_j}^2 \mathbf{I}_t) \hat{\mathbf{b}}_{P_j} = \mathbf{0}, \quad (7.50)$$

$j = 1, 2, \dots, t$ , 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  $\hat{\mathbf{Q}}_P = \hat{\mathbf{P}}^{-1} \hat{\Psi} \mathbf{D} (\mathbf{D}' \hat{\Psi}' \hat{\mathbf{P}}^{-1} \hat{\Psi} \mathbf{D})^{-1} \mathbf{D}' \hat{\Psi}'$ . As  $\hat{\mathbf{S}}$  is an asymmetric matrix, the values of  $\hat{\mathbf{b}}_{P_j}$  and  $\hat{\lambda}_{P_j}^2$  should be obtained using SVD (singular value decomposition).

According to SVD, we need to solve equation

$$(\hat{\mathbf{S}} \hat{\mathbf{S}}' - \hat{\mu}_{P_j} \mathbf{I}_t) \hat{\mathbf{b}}_{P_j} = \mathbf{0}, \quad (7.51)$$

where  $\hat{\mu}_{P_j} = \hat{\lambda}_{P_j}^4$  ( $j = 1, 2, \dots, t$ ). By Eq. (7.51), the estimated PPG-ESIM index ( $I_P = \mathbf{b}'_P \mathbf{y}$ ) is  $\hat{I}_P = \hat{\mathbf{b}}'_{P_1} \mathbf{y}$ . The estimator of the maximized PPG-ESIM selection response, and its expected genetic gain per trait, can be denoted as  $\hat{R}_P = k_I \sqrt{\hat{\mathbf{b}}'_{P_1} \hat{\mathbf{P}} \hat{\mathbf{b}}_{P_1}}$  and  $\hat{E}_P = k_I \frac{\hat{\mathbf{C}} \hat{\mathbf{b}}_{P_1}}{\sqrt{\hat{\mathbf{b}}'_{P_1} \hat{\mathbf{P}} \hat{\mathbf{b}}_{P_1}}}$  respectively, whereas the estimator of the maximized accuracy of the PPG-ESIM is  $\hat{\lambda}_{P_1}$ .

### 7.3.3 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  $\mathbf{d}' = [3 \quad -1]$ . In Chap. 3, Sect. 3.1.4, we indicated how to construct matrix  $\mathbf{U}'$  and, in Sect. 3.2.4 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  $\mathbf{w}' = [19.54 \quad -3.56 \quad 17.01]$ . The estimated PPG-LPSI vector of coefficients for two predetermined restrictions was  $\hat{\mathbf{b}}' = [1.70 \quad 1.04 \quad 2.93]$ , and its

estimated selection response, expected genetic gain per trait, accuracy, and heritability were  $\widehat{R} = 1.755\sqrt{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}} = 49.02$ ,  $\widehat{\mathbf{E}}' = 1.755\frac{\widehat{\mathbf{b}}'\widehat{\mathbf{C}}}{\sqrt{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}}} = [1.25 \quad -0.42 \quad 1.36]$ ,

$\widehat{\rho} = \frac{\sqrt{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}}}{\sqrt{\widehat{\mathbf{w}}'\widehat{\mathbf{C}}\widehat{\mathbf{w}}}} = 0.24$ , and  $\widehat{h}^2 = \frac{\widehat{\mathbf{b}}'\widehat{\mathbf{C}}\widehat{\mathbf{b}}}{\widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}}} = 0.12$  respectively. In this case,

$\widehat{\mathbf{b}}'\widehat{\mathbf{b}} = 12.57$ ; then, the estimated PPG-LPSI selection response using the normalized PPG-LPSI vector of coefficients was  $\widehat{R} = \frac{49.02}{12.57} = 3.90$ , whereas the rest of the estimated PPG-LPSI parameters were the same.

In the PPG-ESIM, we need matrix  $\widehat{\mathbf{S}} = \widehat{\mathbf{K}}_P\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}}$  to obtain the eigenvalues and eigenvectors of  $(\widehat{\mathbf{S}}\widehat{\mathbf{S}}' - \widehat{\mu}_{P_j}\mathbf{I}_t)\widehat{\mathbf{b}}_{P_j} = \mathbf{0}$  that make up matrices  $\mathbf{L}_P^{1/2}$ ,  $\mathbf{V}_{P_1}$ , and  $\widehat{\mathbf{S}} = \mathbf{V}_{P_1}\mathbf{L}_P^{1/2}\mathbf{V}_{P_1}'$ , where  $\widehat{\mu}_{P_j} = \widehat{\lambda}_{P_j}^4$ . It can be shown that  $\widehat{\mathbf{S}} =$

$$\widehat{\mathbf{K}}_P\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{C}} = \begin{bmatrix} 0.1047 & -0.0349 & -0.0279 \\ 0.0678 & -0.0226 & -0.0213 \\ -0.1970 & 0.0657 & 0.4119 \end{bmatrix}, \quad \widehat{\mathbf{S}}\widehat{\mathbf{S}}' = \begin{bmatrix} 0.0130 & 0.0085 & -0.0344 \\ 0.0085 & 0.0056 & -0.0236 \\ -0.0344 & -0.0236 & 0.2118 \end{bmatrix},$$

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  $\widehat{\mu}_{P_j} = \widehat{\lambda}_{P_j}^4$  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,  $\widehat{\mu}_{P_1} = \widehat{\lambda}_{P_1}^4 = 0.2214$ ,

$\widehat{\lambda}_{P_1}^2 = 0.4705$ , and the estimated maximized PPG-ESIM accuracy was  $\widehat{\lambda}_{P_1} = 0.6859$ .

We transformed the first eigenvector  $\widehat{\mathbf{b}}'_{P_1} = [-0.1663 \quad -0.1138 \quad 0.9795]$  using

matrix  $\mathbf{F} = \begin{bmatrix} -9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$  to obtain vector  $\widehat{\boldsymbol{\beta}}_P = \widehat{\mathbf{b}}'_{P_1}\mathbf{F} = [1.4968 \quad -0.1138 \quad 0.9795]$

and  $\widehat{\boldsymbol{\beta}}_P'\widehat{\boldsymbol{\beta}}_P = 3.21$ , whence the estimates of the index, the selection response, and expected genetic gain per trait of the PPG-ESIM were  $\widehat{I}_P = 1.4968\text{RL} - 0.1138\text{SM} +$

$0.9795\text{EW}$ ,  $\widehat{R}_P = \frac{1.755\sqrt{\widehat{\boldsymbol{\beta}}_P'\widehat{\mathbf{P}}\widehat{\boldsymbol{\beta}}_P}}{\widehat{\boldsymbol{\beta}}_P'\widehat{\boldsymbol{\beta}}_P} = \frac{43.01}{3.21} = 13.39$ , and  $\widehat{\mathbf{E}}'_P = 1.755\frac{\widehat{\boldsymbol{\beta}}_P'\widehat{\mathbf{C}}}{\sqrt{\widehat{\boldsymbol{\beta}}_P'\widehat{\mathbf{P}}\widehat{\boldsymbol{\beta}}_P}} =$

$[3.05 \quad -1.96 \quad 0.19]$  respectively. The estimated PPG-LPSI selection response was  $\widehat{R} = \frac{49.02}{12.57} = 3.90$ , 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 ( $\widehat{\lambda}_{P_1} = 0.6859$ ) to

PPG-LPSI accuracy ( $\widehat{\rho} = 0.24$ ), i.e.,  $\frac{\widehat{\lambda}_{P_1}}{\widehat{\rho}} = \frac{0.6859}{0.24} = 2.858$  or, in percentage terms,

$\widehat{\rho}_P = 100(2.858 - 1) = 185.80$ . 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 7.1 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 7.2 presents the estimated PPG-ESIM accuracy ( $\widehat{\rho}_p$ ) and the ratio of  $\widehat{\rho}_p$  to the estimated PPG-LPSI accuracy ( $\widehat{\rho}$ ), expressed in percentage terms,  $\widehat{\rho}_p = 100(\widehat{\lambda}_p - 1)$ , where  $\widehat{\lambda}_p = \widehat{\rho}_p/\widehat{\rho}$ , for one, two, and three predetermined restrictions for five simulated selection cycles. The estimated PPG-LPSI accuracies were presented in Chap. 3 (Table 3.6). 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 7.3 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 Chap. 3, Table 3.5, 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 7.3). 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. *Poult 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

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

