

# Chapter 9

## Multistage Linear Selection Indices



**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.

### 9.1 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 in Chap. 2.

### 9.1.1 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

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

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

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  $\mathbf{b}'_2 = \mathbf{w}'\mathbf{G}'\mathbf{P}^{-1} = [b_{21} \ b_{22} \ b_{23} \ b_{24}]$  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

$$\rho_{HI_1} = \sqrt{\frac{\mathbf{b}'_1\mathbf{P}_1\mathbf{b}_1}{\mathbf{w}'\mathbf{C}\mathbf{w}}} \quad \text{and} \quad \rho_{HI_2} = \sqrt{\frac{\mathbf{b}'_2\mathbf{P}^*\mathbf{b}_2}{\mathbf{w}'\mathbf{C}^*\mathbf{w}}}, \quad (9.1)$$

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

$$\mathbf{E}_1 = k_1 \frac{\mathbf{G}'_1\mathbf{b}_1}{\sqrt{\mathbf{b}'_1\mathbf{P}_1\mathbf{b}_1}} \quad \text{and} \quad \mathbf{E}_2 = k_2 \frac{\mathbf{b}'_2\mathbf{C}^*}{\sqrt{\mathbf{b}'_2\mathbf{P}^*\mathbf{b}_2}}, \quad (9.2)$$

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

$$R_1 = k_1 \sqrt{\mathbf{b}'_1\mathbf{P}_1\mathbf{b}_1} \quad \text{and} \quad R_2 = k_2 \sqrt{\mathbf{b}'_2\mathbf{P}^*\mathbf{b}_2}, \quad (9.3)$$

and the total selection response for the two stages is  $R_1 + R_2$ . In Eqs. (9.1) to (9.3), 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

$$\text{Cov}(X, Y)^* = \text{Cov}(X, Y) - u \frac{\text{Cov}(X, W)\text{Cov}(Y, W)}{\text{Var}(W)}, \quad (9.4)$$

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. (9.4), it is possible to obtain matrices  $\mathbf{P}^*$  and  $\mathbf{C}^*$  using the following two equations:

$$\begin{aligned} \mathbf{P}^* &= \text{Var}(\mathbf{y})^* = \mathbf{P} - u \frac{\text{Cov}(\mathbf{y}, \mathbf{x}_1) \mathbf{b}_1 \mathbf{b}'_1 \text{Cov}(\mathbf{x}_1, \mathbf{y})}{\mathbf{b}'_1 \text{Var}(\mathbf{x}_1) \mathbf{b}_1} \\ &= \mathbf{P} - u \frac{\begin{bmatrix} \mathbf{P}_1 & \\ & \mathbf{P}_{21} \end{bmatrix} \mathbf{b}_1 \mathbf{b}'_1 [\mathbf{P}_1 & \mathbf{P}_{21}]}{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1} \end{aligned} \quad (9.5)$$

and

$$\mathbf{C}^* = \text{Var}(\mathbf{g})^* = \mathbf{C} - u \frac{\text{Cov}(\mathbf{g}, \mathbf{x}_1) \mathbf{b}_1 \mathbf{b}'_1 \text{Cov}(\mathbf{x}_1, \mathbf{g})}{\mathbf{b}'_1 \text{Var}(\mathbf{x}_1) \mathbf{b}_1} = \mathbf{C} - u \frac{\mathbf{G}'_1 \mathbf{b}_1 \mathbf{b}'_1 \mathbf{G}_1}{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}. \quad (9.6)$$

With the Eq. (9.5) result, the correlation between  $I_1 = \mathbf{b}'_1 \mathbf{x}_1$  and  $I_2 = \mathbf{b}'_2 \mathbf{y}$  is

$$\text{Corr}(I_1, I_2) = \frac{\mathbf{b}'_1 [\mathbf{P}_1 \quad \mathbf{P}_{21}] \mathbf{b}_2}{\sqrt{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1} \sqrt{\mathbf{b}'_2 \mathbf{P} \mathbf{b}_2}} = \rho_{12}, \quad (9.7)$$

where  $\sqrt{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}$  and  $\sqrt{\mathbf{b}'_2 \mathbf{P} \mathbf{b}_2}$  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.

### 9.1.2 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

$$p = \prod_{i=1}^N q_i, \quad (9.8)$$

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. 9.7), 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  $I_1 = \mathbf{b}'_1 \mathbf{x}_1$  and  $I_2 = \mathbf{b}'_2 \mathbf{y}$  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)

$$\int_{c_1}^{\infty} \int_{c_2-v_1}^{\infty} f(v_1, v_2) dv_1 dv_2 \tag{9.9}$$

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

$$k_1 = \frac{z(c_1)Q(a)}{p} + \frac{z(c_3)Q(b)\sqrt{(1+\rho_{12})/2}}{p} \tag{9.10}$$

and

$$k_2 = \frac{\rho_{12}z(c_1)Q(a)}{p} + \frac{z(c_3)Q(b)\sqrt{(1+\rho_{12})/2}}{p} \tag{9.11}$$

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  $b = \frac{2c_1 - c_2}{\sqrt{2(1-\rho_{12})}}$ , 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. (9.10) and (9.11). That is, as  $c_1 = 0.80$ ,  $c_2 = 1.6$ ,  $\rho_{12} = 0.8$ , and  $p = 0.2$ , then  $z(c_1) = \frac{\exp\{-0.5(0.8)^2\}}{\sqrt{2\pi}} = 0.290$ ,  $z(c_3) = \frac{\exp\{-0.5[(1.6)^2/2(1.8)]\}}{\sqrt{2\pi}} = 0.28$ ,  $a = \frac{1.6-0.8(1.8)}{\sqrt{1-(0.8)^2}} = 0.27$ ,  $b = \frac{2(0.8)-1.6}{\sqrt{2(0.2)}} = 0$ ,  $\Phi(a) = 0.6064$ ,  $\Phi(b) = 0.5$ ,  $Q(a) = 1 - \Phi(a) = 0.3936$ , 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. (9.10) and (9.11) 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.

### 9.1.3 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

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

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

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

$$\hat{\mathbf{P}}_1 = \begin{bmatrix} 137.178 & -90.957 \\ -90.957 & 201.558 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{G}}_1 = \begin{bmatrix} 14.634 & -18.356 & -0.109 & 1.233 \\ -18.356 & 32.029 & 0.103 & -2.574 \end{bmatrix}$$

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  $\hat{\mathbf{b}}'_2 = \mathbf{w}'\hat{\mathbf{C}}\hat{\mathbf{P}}^{-1} = [-0.59 \ 2.78 \ -49.45 \ 3.75]$  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  $\hat{\rho}_{12} = \frac{\hat{\mathbf{b}}'_1[\hat{\mathbf{P}}_1 \ \hat{\mathbf{P}}_{21}]\hat{\mathbf{b}}_2}{\sqrt{\hat{\mathbf{b}}'_1\hat{\mathbf{P}}_1\hat{\mathbf{b}}_1}\sqrt{\hat{\mathbf{b}}'_2\hat{\mathbf{P}}\hat{\mathbf{b}}_2}} = 0.88$ , where  $\sqrt{\hat{\mathbf{b}}'_1\hat{\mathbf{P}}_1\hat{\mathbf{b}}_1}$  and  $\sqrt{\hat{\mathbf{b}}'_2\hat{\mathbf{P}}\hat{\mathbf{b}}_2}$

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  $\hat{R}_1 = k_1\sqrt{\hat{\mathbf{b}}'_1\hat{\mathbf{P}}_1\hat{\mathbf{b}}_1} = 29.85$ ,  $\hat{\mathbf{E}}'_1 =$

$$k_1 \frac{\hat{\mathbf{G}}'_1\hat{\mathbf{b}}_1}{\sqrt{\hat{\mathbf{b}}'_1\hat{\mathbf{P}}_1\hat{\mathbf{b}}_1}} = [-1.046 \ 1.702 \ 0.006 \ -0.133], \quad \text{and} \quad \hat{\rho}_{HI_1} = \sqrt{\frac{\hat{\mathbf{b}}'_1\hat{\mathbf{P}}_1\hat{\mathbf{b}}_1}{\mathbf{w}'\hat{\mathbf{C}}\mathbf{w}}} = 0.353$$

respectively.

According to the  $k_1 = 0.744$  value, the approached value of  $u$  was  $u = 0.554$ , and by Eqs. (9.5) and (9.6), the estimated and adjusted phenotypic ( $\hat{\mathbf{P}}^*$ ) and genetic ( $\hat{\mathbf{C}}^*$ ) covariance matrices for the second stage were

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

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

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  $\hat{\mathbf{E}}_1' + \hat{\mathbf{E}}_2' = [-1.488 \quad 2.506 \quad -0.081 \quad -0.219]$ .

## 9.2 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.

### 9.2.1 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 ( $\mathbf{b}_2 = \mathbf{P}^{-1} \mathbf{C} \mathbf{w}$ ) described in Sect. 9.1.1 of this chapter, and should be written as

$$\mathbf{b}_{R_1} = \mathbf{K}_1 \mathbf{b}_1 \quad (9.12)$$

and

$$\mathbf{b}_{R_2} = \mathbf{K}_2 \mathbf{b}_2, \quad (9.13)$$



respectively, where, at stage 1,  $\mathbf{K}_1 = [\mathbf{I}_1 - \mathbf{Q}_1]$ ,  $\mathbf{Q}_1 = \mathbf{P}_1^{-1}\Psi_1(\Psi_1'\mathbf{P}_1^{-1}\Psi_1)^{-1}\Psi_1'$ ,  $\Psi_1' = \mathbf{U}'\mathbf{G}'_1$ ,  $\mathbf{I}_1$  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,  $\mathbf{K}_2 = [\mathbf{I}_2 - \mathbf{Q}_2]$ ,  $\mathbf{Q}_2 = \mathbf{P}^{-1}\Psi_2(\Psi_2'\mathbf{P}^{-1}\Psi_2)^{-1}\Psi_2'$ ,  $\Psi_2' = \mathbf{U}'\mathbf{C}$ ,  $\mathbf{I}_2$  is an identity matrix of the same size as  $\mathbf{P}$ , and  $\mathbf{P}^{-1}$  is the inverse of matrix  $\mathbf{P}$ . By Eqs. (9.12) and (9.13), 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. 9.10 and 9.11) respectively, and let  $\mathbf{P}^*$  and  $\mathbf{C}^*$  be the covariance matrices adjusted in the MRLPSI context according to Eqs. (9.5) and (9.5) respectively. The maximized MRLPSI selection response, expected genetic gain per trait, and accuracy at stages 1 and 2 can be written as

$$R_{R_1} = k_1 \sqrt{\mathbf{b}'_{R_1}\mathbf{P}_1\mathbf{b}_{R_1}} \quad \text{and} \quad R_{R_2} = k_2 \sqrt{\mathbf{b}'_{R_2}\mathbf{P}^*\mathbf{b}_{R_2}}, \quad (9.14)$$

$$\mathbf{E}_{R_1} = k_1 \frac{\mathbf{G}'_1\mathbf{b}_{R_1}}{\sqrt{\mathbf{b}'_{R_1}\mathbf{P}_1\mathbf{b}_{R_1}}} \quad \text{and} \quad \mathbf{E}_{R_2} = k_2 \frac{\mathbf{b}'_{R_2}\mathbf{C}^*}{\sqrt{\mathbf{b}'_{R_2}\mathbf{P}^*\mathbf{b}_{R_2}}} \quad (9.15)$$

and

$$\rho_{R_1} = \sqrt{\frac{\mathbf{b}'_{R_1}\mathbf{P}_1\mathbf{b}_{R_1}}{\mathbf{w}'\mathbf{C}\mathbf{w}}} \quad \text{and} \quad \rho_{R_2} = \sqrt{\frac{\mathbf{b}'_{R_2}\mathbf{P}^*\mathbf{b}_{R_2}}{\mathbf{w}'\mathbf{C}^*\mathbf{w}}}, \quad (9.16)$$

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  $\mathbf{E}_{R_1} + \mathbf{E}_{R_2}$ .

## 9.2.2 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

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

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  $\mathbf{K}_1 = [\mathbf{I}_1 - \mathbf{Q}_1]$ ,  $\mathbf{Q}_1 = \mathbf{P}_1^{-1} \Psi_1 (\Psi_1' \mathbf{P}_1^{-1} \Psi_1)^{-1} \Psi_1'$ ,  $\Psi_1 = \mathbf{U}' \mathbf{G}'_1$ ,  $\mathbf{K}_2 = [\mathbf{I}_2 - \mathbf{Q}_2]$ ,  $\mathbf{Q}_2 = \mathbf{P}^{-1} \Psi_2 (\Psi_2' \mathbf{P}^{-1} \Psi_2)^{-1} \Psi_2'$ , and  $\Psi_2 = \mathbf{U}' \mathbf{C}$ , which are used in this subsection.

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

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

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  $\hat{\Psi}'_1 = \mathbf{U} \hat{\mathbf{G}}'_1 = \begin{bmatrix} 44 & 11 & -11 & -3 \\ 11 & 26 & 24 & 7 \end{bmatrix}$ ; 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

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

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

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

The estimated vector  $\mathbf{b}'_{R_1} = \mathbf{b}'_1 \mathbf{K}'_1$  was  $\hat{\mathbf{b}}'_{R_1} = \hat{\mathbf{b}}'_1 \hat{\mathbf{K}}'_1 = [0.044 \ -0.095 \ 0.0450.131]$ , where  $\hat{\mathbf{b}}'_1 = \mathbf{w}' \hat{\mathbf{G}}'_1 \hat{\mathbf{P}}_1^{-1} = [-0.067 \ 0.125 \ 0.045 \ 0.167]$ , and  $\hat{I}_{R_1} = \hat{\mathbf{b}}'_{R_1} \mathbf{x}_1$  was the estimated MRLPSI at stage 1. The estimated MRLPSI vector of coefficients at stage 2 was  $\hat{\mathbf{b}}'_{R_2} = \hat{\mathbf{b}}'_2 \hat{\mathbf{K}}'_2 = [0.045 \ -0.068 \ 0.028 \ -0.057 \ 0.099 \ 0.106]$  and  $\hat{I}_{R_2} = \hat{\mathbf{b}}'_{R_2} \mathbf{y}$  was the estimated MRLPSI at stage 2.

The estimated correlation value ( $\hat{\rho}_{R_{12}}$ ) between  $\hat{I}_{R_1} = \hat{\mathbf{b}}'_{R_1} \mathbf{x}_1$  and  $\hat{I}_{R_2} = \hat{\mathbf{b}}'_{R_2} \mathbf{y}$  was

$$\hat{\rho}_{R_{12}} = \frac{\hat{\mathbf{b}}'_{R_1} [\hat{\mathbf{P}}_1 \ \hat{\mathbf{P}}_{21}] \hat{\mathbf{b}}_{R_2}}{\sqrt{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_{R_1}} \sqrt{\hat{\mathbf{b}}'_{R_2} \hat{\mathbf{P}} \hat{\mathbf{b}}_{R_2}}} = 0.564, \quad \text{where} \quad \sqrt{\hat{\mathbf{b}}'_{R_1} \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_{R_1}} \quad \text{and} \quad \sqrt{\hat{\mathbf{b}}'_{R_2} \hat{\mathbf{P}} \hat{\mathbf{b}}_{R_2}}$$

the estimated standard deviations of the variance of  $\hat{I}_{R_1} = \hat{\mathbf{b}}'_{R_1} \mathbf{x}_1$  and  $\hat{I}_{R_2} = \hat{\mathbf{b}}'_{R_2} \mathbf{y}$  respectively. According to Young (1964, Fig. 8), and Eqs. (9.10) and (9.11), 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  $\widehat{R}_{R_1} = k_1 \sqrt{\widehat{\mathbf{b}}'_{R_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{R_1}} = 0.973$  and  $\widehat{R}_{R_2} = k_2 \sqrt{\widehat{\mathbf{b}}'_{R_2} \widehat{\mathbf{P}}^* \widehat{\mathbf{b}}_{R_2}} = 0.930$ ,  $\widehat{\mathbf{E}}'_{R_1} = k_1 \frac{\widehat{\mathbf{G}}'_1 \widehat{\mathbf{b}}_{R_1}}{\sqrt{\widehat{\mathbf{b}}'_{R_1} \widehat{\mathbf{P}}_1 \widehat{\mathbf{b}}_{R_1}}} = [0 \ 0 \ 1.271 \ 0.870 \ 1.482 \ 0.974]$  and  $\widehat{\mathbf{E}}'_{R_2} = k_2 \frac{\widehat{\mathbf{C}}^* \widehat{\mathbf{b}}_{R_2}}{\sqrt{\widehat{\mathbf{b}}'_{R_2} \widehat{\mathbf{P}}^* \widehat{\mathbf{b}}_{R_2}}} = [0 \ 0 \ 1.419 \ 1.014 \ 2.037 \ 1.349]$ , whereas  $\widehat{R}_{R_1} + \widehat{R}_{R_2} = 1.903$  and  $\widehat{\mathbf{E}}'_{R_1} + \widehat{\mathbf{E}}'_{R_2} = [0 \ 0 \ 2.691 \ 1.884 \ 3.519 \ 2.322]$  were the total estimated MRLPSI selection response and expected genetic gain per trait respectively.

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

### 9.3 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 Chap. 3, 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.

#### 9.3.1 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 ( $\mathbf{b}_2 = \mathbf{P}^{-1} \mathbf{C} \mathbf{w}$ ), and should be written as

$$\mathbf{b}_{M_1} = \mathbf{K}_{M_1} \mathbf{b}_1 \quad (9.17)$$

and

$$\mathbf{b}_{M_2} = \mathbf{K}_{M_2} \mathbf{b}_2, \quad (9.18)$$

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. (9.17) and (9.18), 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. 9.10 and 9.11) respectively and let  $\mathbf{P}^*$  and  $\mathbf{C}^*$  be the adjusted matrices according to Eqs. (9.5) and (9.6) in the MPPG-LPSI context. Then, the MPPG-LPSI selection response and expected genetic gain per trait for both stages can be written as

$$R_{M_1} = k_1 \sqrt{\mathbf{b}'_{M_1} \mathbf{P}_1 \mathbf{b}_{M_1}} \quad \text{and} \quad R_{M_2} = k_2 \sqrt{\mathbf{b}'_{M_2} \mathbf{P}^* \mathbf{b}_{M_2}} \quad (9.19)$$

and

$$\mathbf{E}_{M_1} = k_1 \frac{\mathbf{G}'_1 \mathbf{b}_{M_1}}{\sqrt{\mathbf{b}'_{M_1} \mathbf{P}_1 \mathbf{b}_{M_1}}} \quad \text{and} \quad \mathbf{E}_{M_2} = k_2 \frac{\mathbf{b}'_{M_2} \mathbf{C}^*}{\sqrt{\mathbf{b}'_{M_2} \mathbf{P}^* \mathbf{b}_{M_2}}}, \quad (9.20)$$

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

$$\rho_{M_1} = \sqrt{\frac{\mathbf{b}'_{M_1} \mathbf{P}_1 \mathbf{b}_{M_1}}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{M_2} = \sqrt{\frac{\mathbf{b}'_{M_2} \mathbf{P}^* \mathbf{b}_{M_2}}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}. \quad (9.21)$$

### 9.3.2 Numerical Examples

We use the real data set described in Sect. 9.2.2 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  $\mathbf{b}'_2 = \mathbf{w}' \mathbf{G}'_2 \mathbf{P}_2^{-1}$ . In Chap. 3 we have given methods to estimates  $\mathbf{K}_M = [\mathbf{I} - \mathbf{Q}_M]$ ,  $\mathbf{Q}_M = \mathbf{P}^{-1} \mathbf{M} (\mathbf{M}' \mathbf{P}^{-1} \mathbf{M})^{-1} \mathbf{M}'$ ,  $\mathbf{M}' = \mathbf{D}' \mathbf{\Psi}'$ , and  $\mathbf{\Psi}' = \mathbf{U}' \mathbf{C}$ , which will be used in this subsection.

The estimated phenotypic and genotypic covariance matrices at stage 1 were

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

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  $\hat{\mathbf{M}}'_1 = \mathbf{D}' \mathbf{\Psi}'_1 = \begin{bmatrix} 242 & 7 & -178 & -61 \\ 88 & 58 & -66 & -34 \end{bmatrix}$  and

$$\hat{\mathbf{K}}_{M_1} = \begin{bmatrix} 0.176 & 0.205 & 0.606 & 0.159 \\ 0.031 & 0.032 & -0.007 & 0.199 \\ 0.195 & 0.235 & 0.852 & -0.098 \\ 0.130 & 0.130 & -0.098 & 0.940 \end{bmatrix} \quad \text{respectively, where } \hat{\mathbf{\Psi}}'_1 = \mathbf{U}' \hat{\mathbf{G}}'_1.$$

At stages 1 and 2, the estimated MPPG-LPSI vector of coefficients were  $\hat{\mathbf{b}}'_{M_1} = \hat{\mathbf{b}}'_1 \hat{\mathbf{K}}'_{M_1} = [0.068 \ 0.035 \ 0.039 \ 0.160]$  and  $\hat{\mathbf{b}}'_1 = \mathbf{w}' \mathbf{G}'_1 \mathbf{P}_1^{-1} = [-0.067 \ 0.125 \ 0.045 \ 0.167]$ , whence the estimated MPPG-LPSI were  $\hat{I}_{M_1} = \hat{\mathbf{b}}'_{M_1} \mathbf{x}_1$  and  $\hat{I}_{M_2} = \hat{\mathbf{b}}'_{M_2} \mathbf{y}$ . The estimated correlation value ( $\hat{\rho}_{M_{12}}$ ) between  $\hat{I}_{M_1}$

$= \hat{\mathbf{b}}'_{M_1} \mathbf{x}_1$  and  $\hat{I}_{M_2} = \hat{\mathbf{b}}'_{M_2} \mathbf{y}$  was  $\hat{\rho}_{M_{12}} = \frac{\hat{\mathbf{b}}'_{M_1} [\hat{\mathbf{P}}_1 \ \hat{\mathbf{P}}_{21}] \hat{\mathbf{b}}_{M_2}}{\sqrt{\hat{\mathbf{b}}'_{M_1} \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_{M_1}} \sqrt{\hat{\mathbf{b}}'_{M_2} \hat{\mathbf{P}}_{M_2} \hat{\mathbf{b}}_{M_2}}} = 0.870$ , where

$\sqrt{\hat{\mathbf{b}}'_{M_1} \hat{\mathbf{P}}_1 \hat{\mathbf{b}}_{M_1}}$  and  $\sqrt{\hat{\mathbf{b}}'_{M_2} \hat{\mathbf{P}}_{M_2} \hat{\mathbf{b}}_{M_2}}$  were the estimated standard deviations of variance of  $\hat{I}_{M_1} = \hat{\mathbf{b}}'_{M_1} \mathbf{x}_1$  and  $\hat{I}_{M_2} = \hat{\mathbf{b}}'_{M_2} \mathbf{y}$  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. 9.10 and 9.11) respectively.

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

$\widehat{\mathbf{E}}'_{M_2} = k_2 \frac{\widehat{\mathbf{C}}^{*'} \widehat{\mathbf{b}}_{M_2}}{\sqrt{\widehat{\mathbf{b}}'_{M_2} \widehat{\mathbf{P}}^* \widehat{\mathbf{b}}_{M_2}}} = [0.878 \quad 1.346 \quad 2.604 \quad 1.433 \quad 2.506 \quad 1.602]$ , whereas  $\widehat{R}_{M_1} + \widehat{R}_{M_2} = 2.954$  and  $\widehat{\mathbf{E}}'_{M_1} + \widehat{\mathbf{E}}'_{M_2} = [1.755 \quad 2.662 \quad 4.797 \quad 2.561 \quad 4.161 \quad 2.639]$  were the total estimated MPPGLPSI selection response and expected genetic gain per trait respectively. Note that the vector of predetermined restriction was  $\mathbf{d}' = [2 \quad 3 \quad 5]$ . 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.

## 9.4 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.

### 9.4.1 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  $\boldsymbol{\gamma}_i$  and  $\mathbf{g}_i$  is equal to the variance of  $\boldsymbol{\gamma}_i$ , i.e.,  $Cov(\mathbf{g}_i, \boldsymbol{\gamma}_i) = s_i^2$ , and that the GEBV associated with the  $i$ th trait is a predictor of the  $i$ th vector of genomic breeding values ( $\boldsymbol{\gamma}_i$ ). 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  $\boldsymbol{\gamma}' = [\gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4]$ ,  $\mathbf{g}' = [g_1 \quad g_2 \quad g_3 \quad g_4]$ , and  $\mathbf{w}' = [w_1 \quad w_2 \quad w_3 \quad w_4]$  are the vectors of genomic breeding values ( $\boldsymbol{\gamma}$ ), true breeding values ( $\mathbf{g}$ ), and

economic weights ( $\mathbf{w}$ ) respectively. Let  $\mathbf{\Gamma} = \text{Var}(\boldsymbol{\gamma}) = \begin{bmatrix} s_1^2 & s_{12} & s_{13} & s_{14} \\ s_{21} & s_2^2 & s_{23} & s_{24} \\ s_{31} & s_{32} & s_3^2 & s_{34} \\ s_{41} & s_{42} & s_{43} & s_4^2 \end{bmatrix}$  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,  $\mathbf{\Gamma}_1 = \text{Var}(\boldsymbol{\gamma}_1) = \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix}$  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  $\mathbf{\Gamma}$ .

At stage 2,  $\mathbf{\Gamma} = \text{Var}(\boldsymbol{\gamma})$  is the covariance matrix of  $\boldsymbol{\gamma}$  and  $\mathbf{A} = \mathbf{\Gamma}$  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  $\boldsymbol{\beta}'_1 = \mathbf{w}'\mathbf{A}'_1\mathbf{\Gamma}_1^{-1} = [\beta_{11} \ \beta_{12}]$  and  $\boldsymbol{\beta}'_2 = \mathbf{w}'\mathbf{A}\mathbf{\Gamma}^{-1} = \mathbf{w}' = [w_1 \ w_2 \ w_3 \ w_4]$  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 ( $\rho_{HI_1}$  and  $\rho_{HI_2}$ ), expected genetic gains per trait ( $\mathbf{E}_1$  and  $\mathbf{E}_2$ ) and selection responses ( $R_1$  and  $R_2$ ) can be written as

$$\rho_{HI_1} = \sqrt{\frac{\boldsymbol{\beta}'_1\mathbf{\Gamma}_1\boldsymbol{\beta}_1}{\mathbf{w}'\mathbf{C}\mathbf{w}}} \quad \text{and} \quad \rho_{HI_2} = \sqrt{\frac{\mathbf{w}'\mathbf{\Gamma}^*\mathbf{w}}{\mathbf{w}'\mathbf{C}^*\mathbf{w}}}, \quad (9.22)$$

$$\mathbf{E}_1 = k_1 \frac{\mathbf{A}'_1\boldsymbol{\beta}_1}{\sqrt{\boldsymbol{\beta}'_1\mathbf{\Gamma}_1\boldsymbol{\beta}_1}} \quad \text{and} \quad \mathbf{E}_2 = k_2 \frac{\mathbf{\Gamma}^*\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{\Gamma}^*\mathbf{w}}} \quad (9.23)$$

and

$$R_1 = k_1 \sqrt{\boldsymbol{\beta}'_1\mathbf{\Gamma}_1\boldsymbol{\beta}_1} \quad \text{and} \quad R_2 = k_2 \sqrt{\mathbf{w}'\mathbf{\Gamma}^*\mathbf{w}}. \quad (9.24)$$

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. (9.23) and (9.24), we have omitted the intervals between stages or selection cycles ( $L_G$ ). Matrices  $\mathbf{C}^*$  and  $\mathbf{\Gamma}^*$  in Eqs. (9.22) to (9.23) are matrices  $\mathbf{\Gamma}$  and  $\mathbf{C}$  adjusted for previous selection on  $I_1$ .

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

$$\mathbf{\Gamma}^* = \mathbf{\Gamma} - u \frac{\mathbf{A}'_1 \boldsymbol{\beta}_1 \boldsymbol{\beta}'_1 \mathbf{A}_1}{\boldsymbol{\beta}'_1 \mathbf{\Gamma}_1 \boldsymbol{\beta}_1} \quad (9.25)$$

and

$$\mathbf{C}^* = \mathbf{C} - u \frac{\mathbf{G}'_1 \mathbf{b}_1 \mathbf{b}'_1 \mathbf{G}_1}{\mathbf{b}'_1 \mathbf{P}_1 \mathbf{b}_1}, \quad (9.26)$$

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. (9.26) were defined in Eq. (9.6).

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

$$\text{Corr}(I_1, I_2) = \frac{\boldsymbol{\beta}'_1 \mathbf{A}_1 \mathbf{w}}{\sqrt{\boldsymbol{\beta}'_1 \mathbf{\Gamma}_1 \boldsymbol{\beta}_1} \sqrt{\mathbf{w}' \mathbf{\Gamma} \mathbf{w}}} = \rho_{I_1 I_2}, \quad (9.27)$$

where  $\sqrt{\boldsymbol{\beta}'_1 \mathbf{\Gamma}_1 \boldsymbol{\beta}_1}$  and  $\sqrt{\mathbf{w}' \mathbf{\Gamma} \mathbf{w}}$  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. (9.27), matrix  $\mathbf{\Gamma}$  was not adjusted according to Eq. (9.25).

### 9.4.2 Estimating the Genomic Covariance Matrix

All the MLGSI parameters are associated with matrix  $\mathbf{\Gamma}$ ; thus, the estimation of this matrix in the testing population is very important. We estimate matrix  $\mathbf{\Gamma}$  according to the estimation method described in Chap. 5 (Eq. 5.25), that is, as

$$\widehat{\mathbf{\Gamma}}_l = \left\{ \widehat{\sigma}_{\gamma_{q'l}} \right\}, \quad (9.28)$$

where  $\widehat{\sigma}_{\gamma_{q'l}} = \frac{1}{g} (\widehat{\boldsymbol{\gamma}}_{ql} - \mathbf{1} \widehat{\mu}_{\gamma_{ql}})' \mathbf{G}_l^{-1} (\widehat{\boldsymbol{\gamma}}_{q'l} - \mathbf{1} \widehat{\mu}_{\gamma_{q'l}})$  is the estimated covariance between  $\widehat{\boldsymbol{\gamma}}_{ql} = \mathbf{X}_l \widehat{\mathbf{u}}_q$  and  $\widehat{\boldsymbol{\gamma}}_{q'l} = \mathbf{X}_l \widehat{\mathbf{u}}_{q'}$  at stage  $l$  or selection cycle of the testing population;  $g$  is the number of genotypes;  $\widehat{\mu}_{\gamma_{ql}}$  and  $\widehat{\mu}_{\gamma_{q'l}}$  are the estimated arithmetic means of the values of  $\widehat{\boldsymbol{\gamma}}_{ql}$  and  $\widehat{\boldsymbol{\gamma}}_{q'l}$ ;  $\mathbf{1}$  is 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 (see Chap. 5 for details).

### 9.4.3 Numerical Examples

We illustrate the MLGSI theoretical results using the data described in Chap. 2, Sect. 2.8.1 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. (2.22) to (2.24), whereas matrix  $\hat{\mathbf{\Gamma}}$  was obtained according to Eq. (9.28).

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  $\boldsymbol{\beta}'_2 = \mathbf{w}'\mathbf{A}\mathbf{\Gamma}^{-1} = \mathbf{w}' = [w_1 \ w_2 \ w_3 \ w_4]$ , the estimated MLGSI vector of

coefficients is the vector of economic weights. Thus,  $\hat{\rho}_{I_1I_2} = \frac{\hat{\boldsymbol{\beta}}'_1\hat{\mathbf{A}}_1\mathbf{w}}{\sqrt{\hat{\boldsymbol{\beta}}'_1\hat{\mathbf{\Gamma}}_1\hat{\boldsymbol{\beta}}_1}\sqrt{\mathbf{w}'\hat{\mathbf{\Gamma}}\mathbf{w}}} =$

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

$$\hat{\mathbf{\Gamma}}^* = \begin{bmatrix} 7.96 & -2.11 & 2.71 & 0.88 \\ -2.11 & 3.46 & -0.80 & -0.87 \\ 2.71 & -0.80 & 2.75 & 0.45 \\ 0.88 & -0.87 & 0.45 & 2.38 \end{bmatrix} \text{ and } \hat{\mathbf{C}}^* = \begin{bmatrix} 24.40 & -5.65 & 5.47 & 1.39 \\ -5.65 & 8.55 & -1.63 & -1.41 \\ 5.47 & -1.63 & 9.26 & -0.17 \\ 1.39 & -1.41 & -0.17 & 6.49 \end{bmatrix}.$$

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

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

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  $\widehat{\mathbf{E}}'_1 + \widehat{\mathbf{E}}'_2 = [4.62 \quad -2.45 \quad 1.85 \quad 1.07]$ .

## 9.5 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.

### 9.5.1 The MRLGSI Parameters

In Sect. 9.4.1, we indicated that the MLGSI vector of coefficients at stage 1 can be written as  $\boldsymbol{\beta}'_1 = \mathbf{w}'\mathbf{A}'_1\boldsymbol{\Gamma}_1^{-1} = [\beta_{11} \quad \beta_{12}]$  and at stage 2 as  $\boldsymbol{\beta}'_2 = \mathbf{w}'\mathbf{A}\boldsymbol{\Gamma}^{-1} = \mathbf{w}' = [w_1 \quad w_2 \quad w_3 \quad w_4]$ . 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. 3 and 6 for details) that projects  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  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

$$\boldsymbol{\beta}_{R_1} = \mathbf{K}_{G_1}\boldsymbol{\beta}_1 \quad (9.29)$$

and

$$\boldsymbol{\beta}_{R_2} = \mathbf{K}_{G_2}\boldsymbol{\beta}_2 = \mathbf{K}_{G_2}\mathbf{w}, \quad (9.30)$$

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. (9.29) and (9.30), 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  $\boldsymbol{\gamma}'_1 = [\gamma_1 \quad \gamma_2]$  and  $\boldsymbol{\gamma}' = [\gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4]$  are vectors of genomic breeding values, which can be estimated using GEBVs, as described in Chap. 5. In Chap. 6 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

$$\rho_{HI_1} = \sqrt{\frac{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}}{\mathbf{w}' \mathbf{C} \mathbf{w}}} \quad \text{and} \quad \rho_{HI_2} = \sqrt{\frac{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}{\mathbf{w}' \mathbf{C}^* \mathbf{w}}}, \quad (9.31)$$

$$\mathbf{E}_{R_1} = k_1 \frac{\mathbf{A}'_1 \boldsymbol{\beta}_{R_1}}{\sqrt{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}}} \quad \text{and} \quad \mathbf{E}_{R_2} = k_2 \frac{\boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}{\sqrt{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}} \quad (9.32)$$

and

$$R_{R_1} = k_1 \sqrt{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}} \quad \text{and} \quad R_{R_2} = k_2 \sqrt{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{R_2}}, \quad (9.33)$$

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. (9.32) and (9.33), we have omitted the intervals between stages or selection cycles ( $L_G$ ). Matrices  $\boldsymbol{\Gamma}^*$  and  $\mathbf{C}^*$  in Eqs. (9.31) to (9.33) 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

$$\boldsymbol{\Gamma}^* = \boldsymbol{\Gamma} - u \frac{\mathbf{A}'_1 \boldsymbol{\beta}_{R_1} \boldsymbol{\beta}'_{R_1} \mathbf{A}_1}{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}} \quad (9.34)$$

and

$$\mathbf{C}^* = \mathbf{C} - u \frac{\mathbf{G}'_1 \mathbf{b}_{R_1} \mathbf{b}'_{R_1} \mathbf{G}_1}{\mathbf{b}'_{R_1} \mathbf{P}_1 \mathbf{b}_{R_1}}, \quad (9.35)$$

where  $\boldsymbol{\beta}_{R_1}$  was defined in Eq. (9.29) and vector  $\mathbf{b}_{R_1}$  can be obtained according to the RLPSI as described in Chap. 3. 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

$$\rho_{I_{R_1} I_{R_2}} = \frac{\boldsymbol{\beta}'_{R_1} \mathbf{A}_1 \boldsymbol{\beta}_{R_2}}{\sqrt{\boldsymbol{\beta}'_{R_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{R_1}} \sqrt{\boldsymbol{\beta}'_{R_2} \boldsymbol{\Gamma} \boldsymbol{\beta}_{R_2}}}, \quad (9.36)$$

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. (9.36), matrix  $\boldsymbol{\Gamma}$  was not adjusted for previous selection on  $I_{R_1} = \boldsymbol{\beta}'_{R_1} \boldsymbol{\gamma}_1$ .

## 9.5.2 Numerical Examples

To illustrate the MRLGSI theory in a two-stage breeding selection scheme, we use the simulated data described in Sect. 9.4.3. In that subsection we indicated that the

estimated covariance matrices of  $\Gamma_1$  and  $\mathbf{A}_1$  were  $\hat{\Gamma}_1 = \begin{bmatrix} 16.26 & -6.51 \\ -6.51 & 5.79 \end{bmatrix}$  and  $\hat{\mathbf{A}}_1 = \begin{bmatrix} 16.26 & -6.51 & 5.60 & 2.29 \\ -6.51 & 5.79 & -2.33 & -1.62 \end{bmatrix}$ , and that  $\hat{\boldsymbol{\beta}}'_1 = \mathbf{w}'\hat{\mathbf{A}}_1\hat{\Gamma}_1^{-1} = [1.39 \quad -1.25]$  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,  $\hat{\mathbf{Q}}_{G_1} = \mathbf{U}_1(\mathbf{U}'_1\hat{\Gamma}_1\mathbf{U}_1)^{-1}\mathbf{U}'_1\hat{\Gamma}_1$ ,  $\hat{\mathbf{Q}}_{G_2} = \mathbf{U}_2(\mathbf{U}'_2\hat{\Gamma}_1\mathbf{U}_2)^{-1}\mathbf{U}'_2\hat{\Gamma}_1$ ,  $\hat{\mathbf{K}}_{G_1} = [\mathbf{I} - \hat{\mathbf{Q}}_{G_1}]$ , and  $\hat{\mathbf{K}}_{G_2} = [\mathbf{I} - \hat{\mathbf{Q}}_{G_2}]$  are the estimated matrices described in Eqs. (9.29) and (9.30) for stages 1 and 2. It can be shown that, at stages 1 and 2,  $\hat{\boldsymbol{\beta}}'_{R_1} = \hat{\boldsymbol{\beta}}'_1\hat{\mathbf{K}}'_{G_1} = [1.39 \quad 1.558]$  and  $\hat{\boldsymbol{\beta}}'_{R_2} = \mathbf{w}'\hat{\mathbf{K}}'_{G_2} = [1.0 \quad 1.81 \quad 1.010]$  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  $\hat{R}_{R_1} = k_1\sqrt{\hat{\boldsymbol{\beta}}'_{R_1}\hat{\Gamma}_1\hat{\boldsymbol{\beta}}_{R_1}} = 3.083$ ,  $\hat{\mathbf{E}}_{R_1} = [2.225 \quad 0 \quad 0.742 \quad 0.117]$ , and  $\hat{\rho}_{HI_1} = \sqrt{\frac{\hat{\boldsymbol{\beta}}'_{R_1}\hat{\Gamma}_1\hat{\boldsymbol{\beta}}_{R_1}}{\mathbf{w}'\hat{\mathbf{C}}\mathbf{w}}} = 0.370$  respectively. The estimated MRLGSI expected genetic gain, accuracy, and selection response at stage 2 were  $\hat{R}_{R_2} = k_2\frac{\hat{\boldsymbol{\beta}}'_{R_2}\hat{\Gamma}^*\hat{\boldsymbol{\beta}}_{R_2}}{\sqrt{\hat{\boldsymbol{\beta}}'_{R_2}\hat{\Gamma}^*\hat{\boldsymbol{\beta}}_{R_2}}} = [1.156 \quad 0 \quad 0.793 \quad 0.536]$ ,  $\hat{\rho}_{HI_2} = \sqrt{\frac{\hat{\boldsymbol{\beta}}'_{R_2}\hat{\Gamma}^*\hat{\boldsymbol{\beta}}_{R_2}}{\mathbf{w}'\hat{\mathbf{C}}^*\mathbf{w}}} = 0.32$ , and  $\hat{R}_{R_2} = k_2\sqrt{\hat{\boldsymbol{\beta}}'_{R_2}\hat{\Gamma}^*\hat{\boldsymbol{\beta}}_{R_2}} = 2.485$  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  $\hat{R}_{R_1} + \hat{R}_{R_2} = 5.568$  and  $\hat{\mathbf{E}}'_{R_1} + \hat{\mathbf{E}}'_{R_2} = [3.380 \quad 0 \quad 1.535 \quad 0.653]$  respectively. Note that, in effect, the expected genetic gain for trait 2 was 0, as expected.

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

### 9.6.1 The OMPPG-LGSI Parameters

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

$$\boldsymbol{\beta}_{P_1} = \boldsymbol{\beta}_{R_1} + \theta_1\mathbf{U}_1(\mathbf{U}'_1\boldsymbol{\Gamma}_1\mathbf{U}_1)^{-1}\mathbf{d} \tag{9.37}$$

and

$$\boldsymbol{\beta}_{P_2} = \boldsymbol{\beta}_{R_2} + \theta_2\mathbf{U}_2(\mathbf{U}'_2\boldsymbol{\Gamma}_2\mathbf{U}_2)^{-1}\mathbf{d}, \tag{9.38}$$

respectively, where  $\boldsymbol{\beta}_{R_1} = \mathbf{K}_{G_1}\boldsymbol{\beta}_1$ ,  $\boldsymbol{\beta}_{R_2} = \mathbf{K}_{G_2}\boldsymbol{\beta}_2 = \mathbf{K}_{G_2}\mathbf{w}$ ,  $\mathbf{K}_{G_1} = [\mathbf{I} - \mathbf{Q}_{G_1}]$ ,  $\mathbf{Q}_{G_1} = \mathbf{U}_1(\mathbf{U}'_1\boldsymbol{\Gamma}_1\mathbf{U}_1)^{-1}\mathbf{U}'_1\boldsymbol{\Gamma}_1$ ,  $\mathbf{K}_{G_2} = [\mathbf{I} - \mathbf{Q}_{G_2}]$ , and  $\mathbf{Q}_{G_2} = \mathbf{U}_2(\mathbf{U}'_2\boldsymbol{\Gamma}_2\mathbf{U}_2)^{-1}\mathbf{U}'_2\boldsymbol{\Gamma}_2$  were described in Eqs. (9.29) and (9.30). Also, it can be shown that the proportionality constants for stages 1 ( $\theta_1$ ) and 2 ( $\theta_2$ ) are

$$\theta_1 = \frac{\mathbf{d}'(\mathbf{U}'_1\boldsymbol{\Gamma}_1\mathbf{U}_1)^{-1}\mathbf{U}'_1\mathbf{A}_1\mathbf{w}}{\mathbf{d}'(\mathbf{U}'_1\boldsymbol{\Gamma}_1\mathbf{U}_1)^{-1}\mathbf{d}} \quad \text{and} \quad \theta_2 = \frac{\mathbf{d}'(\mathbf{U}'_2\boldsymbol{\Gamma}_2\mathbf{U}_2)^{-1}\mathbf{U}'_2\boldsymbol{\Gamma}_2\mathbf{w}}{\mathbf{d}'(\mathbf{U}'_2\boldsymbol{\Gamma}_2\mathbf{U}_2)^{-1}\mathbf{d}}, \tag{9.39}$$

respectively. By Eqs. (9.37) to (9.39), the MPPG-LGSI for stages 1 and 2 can be written as  $I_{P_1} = \boldsymbol{\beta}'_{P_1}\boldsymbol{\gamma}_1$  and  $I_{P_2} = \boldsymbol{\beta}'_{P_2}\boldsymbol{\gamma}$  respectively, where  $\boldsymbol{\gamma}_1$  and  $\boldsymbol{\gamma}$  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 ( $\mathbf{E}_{P_1}$  and  $\mathbf{E}_{P_2}$ ), and selection responses ( $R_{P_1}$  and  $R_{P_2}$ ) can be written as

$$\rho_{HI_1} = \sqrt{\frac{\boldsymbol{\beta}'_{P_1}\boldsymbol{\Gamma}_1\boldsymbol{\beta}_{P_1}}{\mathbf{w}'\mathbf{C}\mathbf{w}}} \quad \text{and} \quad \rho_{HI_2} = \sqrt{\frac{\boldsymbol{\beta}'_{P_2}\boldsymbol{\Gamma}^*\boldsymbol{\beta}_{P_2}}{\mathbf{w}'\mathbf{C}^*\mathbf{w}}}, \tag{9.40}$$

$$\mathbf{E}_{P_1} = k_1 \frac{\mathbf{A}'_1\boldsymbol{\beta}_{P_1}}{\sqrt{\boldsymbol{\beta}'_{P_1}\boldsymbol{\Gamma}_1\boldsymbol{\beta}_{P_1}}} \quad \text{and} \quad \mathbf{E}_{P_2} = k_2 \frac{\boldsymbol{\Gamma}^*\boldsymbol{\beta}_{P_2}}{\sqrt{\boldsymbol{\beta}'_{P_2}\boldsymbol{\Gamma}^*\boldsymbol{\beta}_{P_2}}} \tag{9.41}$$

and

$$R_{P_1} = k_1 \sqrt{\boldsymbol{\beta}'_{P_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{P_1}} \quad \text{and} \quad R_{P_2} = k_2 \sqrt{\boldsymbol{\beta}'_{P_2} \boldsymbol{\Gamma}^* \boldsymbol{\beta}_{P_2}}, \quad (9.42)$$

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. (9.41) and (9.42), 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. (9.34) and (9.35) 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

$$\rho_{12} = \frac{\boldsymbol{\beta}'_{P_1} \mathbf{A}_1 \boldsymbol{\beta}_{P_2}}{\sqrt{\boldsymbol{\beta}'_{P_1} \boldsymbol{\Gamma}_1 \boldsymbol{\beta}_{P_1}} \sqrt{\boldsymbol{\beta}'_{P_2} \boldsymbol{\Gamma} \boldsymbol{\beta}_{P_2}}}. \quad (9.43)$$

In Eq. (9.43), matrix  $\boldsymbol{\Gamma}$  was not adjusted for previous selection on  $I_{P_1} = \boldsymbol{\beta}'_{P_1} \boldsymbol{\gamma}_1$ .

## 9.6.2 Numerical Examples

To illustrate the MPPG-LGSI theory, we use the simulated data described in Sect. 9.4.3. Suppose that we select two traits at stages 1 and 2; then, at stage 1,  $\hat{\boldsymbol{\Gamma}}_1 = \begin{bmatrix} 16.26 & -6.51 \\ -6.51 & 5.79 \end{bmatrix}$  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  $\boldsymbol{\Gamma}_1$  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,  $\hat{\mathbf{Q}}_{G_1} = \mathbf{U}_1 (\mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1$ ,  $\hat{\mathbf{Q}}_{G_2} = \mathbf{U}_2 (\mathbf{U}'_2 \hat{\boldsymbol{\Gamma}} \mathbf{U}_2)^{-1} \mathbf{U}'_2 \hat{\boldsymbol{\Gamma}}$ ,  $\hat{\mathbf{K}}_{G_1} = [\mathbf{I} - \hat{\mathbf{Q}}_{G_1}]$ , and  $\hat{\mathbf{K}}_{G_2} = [\mathbf{I} - \hat{\mathbf{Q}}_{G_2}]$  are the estimates of matrix projectors associated with stages 1 and 2 (Eqs. 9.37 and 9.38 for details).

In Sect. 9.4.3, we showed that the estimated MRLGSI vector of coefficients for stage 1 was  $\hat{\boldsymbol{\beta}}'_{R_1} = \hat{\boldsymbol{\beta}}'_1 \hat{\mathbf{K}}'_{G_1} = [1.386 \ 1.550]$ . Thus, by Eq. (9.37), to obtain  $\hat{\boldsymbol{\beta}}_{P_1} = \hat{\boldsymbol{\beta}}_{R_1} + \hat{\theta}_1 \mathbf{U}_1 (\mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d}$ , we only need to obtain  $\hat{\theta}_1$  and  $\mathbf{U}_1 (\mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d}$ , where  $\mathbf{d} = -2$  and  $\hat{\theta}_1 = \frac{\mathbf{d}' (\mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 \hat{\mathbf{A}}_1 \mathbf{w}}{\mathbf{d}' (\mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d}}$ . It can be shown that  $\mathbf{U}_1 (\mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{d} = \begin{bmatrix} 0 \\ -0.345 \end{bmatrix}$  and  $\hat{\theta}_1 = 8.125$ ; therefore,  $\hat{\boldsymbol{\beta}}'_{P_1} = [1.39 \ -1.25]$  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  $\hat{\rho}_{HI_1} = \sqrt{\frac{\hat{\boldsymbol{\beta}}'_{P_1} \hat{\boldsymbol{\Gamma}}_1 \hat{\boldsymbol{\beta}}_{P_1}}{\mathbf{w}' \hat{\mathbf{C}} \mathbf{w}}} = 0.71$ ,  $\hat{R}_{P_1} = k_1 \sqrt{\hat{\boldsymbol{\beta}}'_{P_1} \hat{\boldsymbol{\Gamma}}_1 \hat{\boldsymbol{\beta}}_{P_1}} = 5.90$  and  $\hat{\mathbf{E}}'_{P_1} = k_1 \frac{\hat{\mathbf{A}}'_1 \hat{\boldsymbol{\beta}}_{P_1}}{\sqrt{\hat{\boldsymbol{\beta}}'_{P_1} \hat{\boldsymbol{\Gamma}}_1 \hat{\boldsymbol{\beta}}_{P_1}}} = [2.88 \quad -1.53 \quad 1.00 \quad 0.49]$  respectively.

It can be shown that at stage 2,  $\mathbf{d}'(\mathbf{U}'_1 \hat{\boldsymbol{\Gamma}}_1 \mathbf{U}_1)^{-1} \mathbf{U}'_1 = [0 \quad -0.345 \quad 0 \quad 0]$ ,  $\hat{\theta}_2 = 8.125$  and  $\hat{\boldsymbol{\beta}}'_{P_2} = \mathbf{w}' = [1 \quad -1 \quad 1 \quad 1]$ . Thus, the estimated MPPG-LGSI accuracy, selection response, and expected genetic gain at this stage were  $\hat{\rho}_{HI_2} = \sqrt{\frac{\mathbf{w}' \hat{\boldsymbol{\Gamma}}^* \mathbf{w}}{\mathbf{w}' \hat{\mathbf{C}}^* \mathbf{w}}} = 0.64$ ,  $\hat{R}_{P_2} = k_2 \sqrt{\mathbf{w}' \hat{\boldsymbol{\Gamma}}^* \mathbf{w}} = 4.10$ , and  $\hat{\mathbf{E}}'_{P_2} = k_2 \frac{\hat{\boldsymbol{\Gamma}}^* \mathbf{w}}{\sqrt{\mathbf{w}' \hat{\boldsymbol{\Gamma}}^* \mathbf{w}}} = [1.74 \quad -0.92 \quad 0.85 \quad 0.58]$  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, 3rd 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

**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.

