Chapter 6
Constrained Linear Genomic Selection Indices

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

6.1 The Restricted Linear Genomic Selection Index

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

J. J. Céron-Rojas, J. Crossa, *Linear Selection Indices in Modern Plant Breeding*,
https://doi.org/10.1007/978-3-319-91223-3_6

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

6.1.1 The Maximized RLGSI Parameters

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

$$f_R(\boldsymbol{\beta}, \mathbf{v}) = \mathbf{w}'\mathbf{C}\mathbf{w} + \boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta} - 2\mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta} + 2\mathbf{v}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} \tag{6.1}$$

with respect to vectors $\boldsymbol{\beta}$ and $\mathbf{v}' = [v_1 \;\; v_2 \;\; \cdots \;\; v_{r-1}]$, where $\mathbf{v}$ is a vector of Lagrange multipliers. In matrix notation, the derivative results of Eq. (6.1) are

$$\begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Gamma}\mathbf{U} \\ \mathbf{U}'\boldsymbol{\Gamma} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\Gamma}\mathbf{w} \\ \mathbf{0} \end{bmatrix}. \tag{6.2}$$

Following the procedure described in Chap. 3 (Eqs. 3.2 to 3.5), it can be shown that the RLGSI vector of coefficients that minimizes $E[(H - I_G)^2]$ under the restriction $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$ is

$$\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}, \tag{6.3}$$

where $\mathbf{K}_G = [\mathbf{I}_t - \mathbf{Q}_G]$, $\mathbf{Q}_G = \mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\Gamma}$, $\mathbf{w}$ is a vector of economic weights, and $\mathbf{I}_t$ is an identity matrix $t \times t$. When no restrictions are imposed on any of the traits, $\mathbf{U}'$ is a null matrix and $\boldsymbol{\beta}_{RG} = \mathbf{w}$, the optimized LGSI vector of coefficients (see Chap. 5 for details).

By Eq. (6.3), the RLGSI, and the maximized RLGSI selection response and expected genetic gain per trait can be written as

$$I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}, \tag{6.4}$$

$$R_{RG} = \frac{k_I}{L_G}\sqrt{\boldsymbol{\beta}'_{RG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}} \tag{6.5}$$

and

$$\mathbf{E}_{RG} = \frac{k_I}{L_G} \frac{\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}}{\sqrt{\boldsymbol{\beta}'_{RG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}}}, \tag{6.6}$$

respectively, where k_I is the standardized selection differential (or selection intensity) associated with the RLGSI, and L_G is the interval between selection cycles or the time required to complete a selection cycle using the RLGSI. Equations (6.4) to (6.6) depend only on GEBV information; thus, they are useful in testing populations.

6.1.2 Statistical Properties of RLGSI

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

1. Matrices $\mathbf{K}_G$ and $\mathbf{Q}_G$ are idempotent ($\mathbf{K}_G = \mathbf{K}_G^2$ and $\mathbf{Q}_G = \mathbf{Q}_G^2$) and orthogonal ($\mathbf{K}_G\mathbf{Q}_G = \mathbf{Q}_G\mathbf{K}_G = \mathbf{0}$), that is, they are projectors. Matrix $\mathbf{Q}_G$ projects vector $\boldsymbol{\beta} = \mathbf{w}$ into a space generated by the columns of matrix $\mathbf{U}'\boldsymbol{\Gamma}$ due to the restriction $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$ used when $f_R(\boldsymbol{\beta}, \mathbf{v})$ (Eq. 6.1) is minimized with respect to vectors $\boldsymbol{\beta}$ and $\mathbf{v}$, whereas matrix $\mathbf{K}_G$ projects $\mathbf{w}$ into a space perpendicular to that generated by the $\mathbf{U}'\boldsymbol{\Gamma}$ matrix columns.
2. Because of the restriction $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$, matrix $\mathbf{K}_G$ projects vector $\mathbf{w}$ into a space smaller than the original space of $\mathbf{w}$. The space reduction into which matrix $\mathbf{K}_G$ projects $\mathbf{w}$ is equal to the number of zeros that appears in Eq. (6.6).
3. Vector $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$ minimizes the mean square error under the restriction $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$.
4. The variance of $I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}$ ($\sigma^2_{I_{RG}} = \boldsymbol{\beta}'_{RG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}$) is equal to the covariance between $I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}$ and $H = \mathbf{w}'\mathbf{g}$ ($\sigma_{HI_{RG}} = \mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}$).
5. The maximized correlation between H and I_{RG} is equal to $\rho_{HI_{RG}} = \frac{\sigma_{I_{RG}}}{\sigma_H}$, where $\sigma_{I_{RG}} = \sqrt{\boldsymbol{\beta}'_{RG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{RG}}$ and $\sigma_H = \sqrt{\mathbf{w}'\mathbf{C}\mathbf{w}}$ are the standard deviations of $I_{RG} = \boldsymbol{\beta}'_{RG}\boldsymbol{\gamma}$ and $H = \mathbf{w}'\mathbf{g}$ respectively.
6. The variance of the predicted error, $Var(H - I_{RG}) = \left(1 - \rho^2_{HI_{RG}}\right)\sigma^2_H$, is minimal. Note that $Var(H - I_{RG}) = \sigma^2_{I_{RG}} + \sigma^2_H - 2\sigma_{HI_{RG}}$, and when $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$, $\sigma^2_{I_{RG}} = \sigma_{HI_{RG}}$, whence $Var(H - I_{RG}) = \sigma^2_H - \sigma^2_{I_{RG}} = \left(1 - \rho^2_{HI_{RG}}\right)\sigma^2_H$ is minimal.

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

6.1.3 Numerical Examples

To estimate the parameters associated with the RLGSI, we use the real data set described in Chap. 5, Sect. 5.1.8, where we found that, in the testing population, the

estimate of matrix $\boldsymbol{\Gamma}$ was $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.21 & 2.95 & 5.00 \\ 2.95 & 42.41 & 71.11 \\ 5.00 & 71.11 & 121.53 \end{bmatrix}$. We use this matrix and the GEBVs associated with the traits grain yield (GY, ton ha^{-1}), ear height (EHT, cm), and plant height (PHT, cm) to illustrate the RLGSI theoretical results.

Suppose that on the RLGSI expected genetic gain per trait we impose one and two null restrictions using matrices $\mathbf{U}_1' = [1 \quad 0 \quad 0]$ and $\mathbf{U}_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ (see Chap. 3, Sect. 3.1.3, for details about matrix $\mathbf{U}'$). We need to estimate the RLGSI vector of coefficients ($\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$) as $\widehat{\boldsymbol{\beta}}_{RG} = \widehat{\mathbf{K}}_G\mathbf{w}$, where $\widehat{\mathbf{K}}_G = \left[\mathbf{I}_3 - \widehat{\mathbf{Q}}_G\right]$ and $\widehat{\mathbf{Q}}_G = \mathbf{U}\left(\mathbf{U}'\widehat{\boldsymbol{\Gamma}}\mathbf{U}\right)^{-1}\mathbf{U}'\widehat{\boldsymbol{\Gamma}}$ are estimates of matrices $\mathbf{K}_G = [\mathbf{I}_3 - \mathbf{Q}_G]$ and $\mathbf{Q}_G = \mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\Gamma}$ respectively, and $\mathbf{I}_3$ is an identity matrix 3×3. The estimated $\mathbf{Q}_G$ matrices for restrictions $\mathbf{U}_1' = [1\ 0\ 0]$ and $\mathbf{U}_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ were $\widehat{\mathbf{Q}}_{G_1} = \mathbf{U}_1\left(\mathbf{U}_1'\widehat{\boldsymbol{\Gamma}}\mathbf{U}_1\right)^{-1}\mathbf{U}_1'\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 1.0 & 14.05 & 23.81 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\widehat{\mathbf{Q}}_{G_2} = \mathbf{U}_2\left(\mathbf{U}_2'\widehat{\boldsymbol{\Gamma}}\mathbf{U}_2\right)^{-1}\mathbf{U}_2'\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 1.0 & 0 & 11.18 \\ 0 & 1.0 & 0.90 \\ 0 & 0 & 0 \end{bmatrix}$ respectively, whereas the estimated $\mathbf{K}_G$ matrices for both restrictions were $\widehat{\mathbf{K}}_{G_1} = \left[\mathbf{I}_3 - \widehat{\mathbf{Q}}_{G_1}\right] = \begin{bmatrix} 0 & -14.05 & -23.81 \\ 0 & 1.0 & 0 \\ 0 & 0 & 1.0 \end{bmatrix}$ and $\widehat{\mathbf{K}}_{G_2} = \left[\mathbf{I}_3 - \widehat{\mathbf{Q}}_{G_2}\right] = \begin{bmatrix} 0 & 0 & -11.18 \\ 0 & 0 & -0.90 \\ 0 & 0 & 1.0 \end{bmatrix}$.

Let $\mathbf{w}' = [5 \quad -0.1 \quad -0.1]$ be the vector of economic weights; then the estimated RLGSI vector of coefficients for one and two null restrictions were $\widehat{\boldsymbol{\beta}}_{RG_1}' = \mathbf{w}'\widehat{\mathbf{K}}_{G_1}' = [3.78 \quad -0.1 \quad -0.1]$ and $\widehat{\boldsymbol{\beta}}_{RG_2}' = \mathbf{w}'\widehat{\mathbf{K}}_{G_2}' = [1.12 \quad 0.09 \quad -0.1]$ respectively, and the estimated RLGSI for both restrictions can be written as $\widehat{I}_{RG_1} = 3.78\text{GEBV}_1 - 0.1\text{GEBV}_2 - 0.1\text{GEBV}_3$ and $\widehat{I}_{RG_2} = 1.12\text{GEBV}_1 + 0.09\text{GEBV}_2 - 0.1\text{GEBV}_3$, where GEBV_1, GEBV_2, and GEBV_3 are the genomic estimated breeding values associated with traits GY, EHT, and PHT respectively in the testing population.

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

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

Table 6.1 Number of genotypes selected from 380 genotypes of a real testing population; genomic estimated breeding values (GEBVs) associated with three traits: grain yield (GY, ton ha^{-1}), ear height (EHT, cm), and plant height (PHT, cm) in the testing population, and estimated and ranked restricted linear genomic selection index (RLGSI) values obtained in the testing population for one null restriction

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

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

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

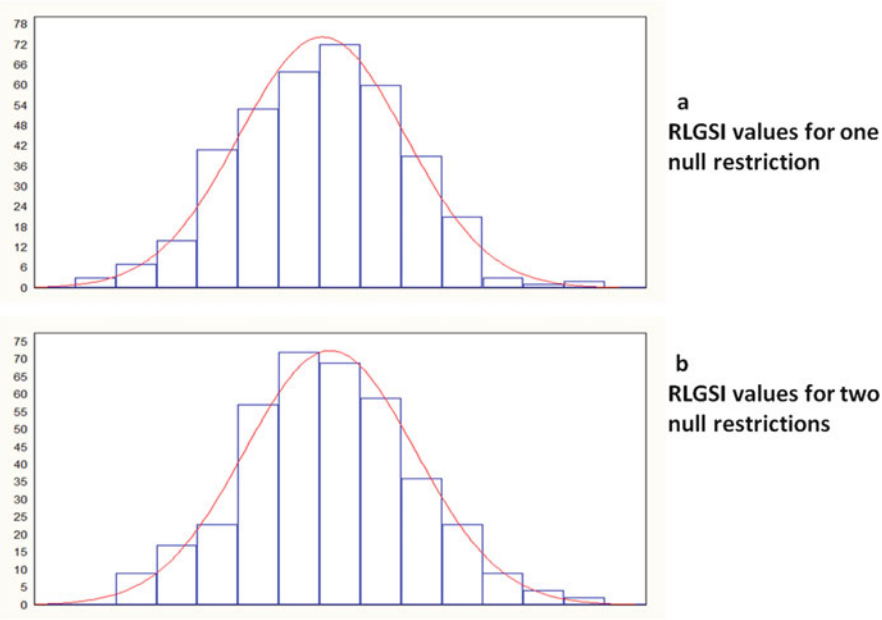

Fig. 6.1 Distribution of 380 estimated restricted linear genomic selection index (RLGSI) values with one (**a**) and two (**b**) null restrictions respectively obtained in a real testing population for one selection cycle in one environment

Now we use the simulated data set described in Chap. 2, Sect. 2.8.1, to compare RLPSI (restricted linear phenotypic selection index, Chap. 3 for details) efficiency versus RLGSI efficiency. Table 6.2 presents the estimated RLPSI and RLGSI selection response for one, two, and three null restrictions imposed by matrices $\mathbf{U}_1' = [1 \quad 0 \quad 0]$, $\mathbf{U}_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, and $\mathbf{U}_3' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ for five simulated selection cycles including and not including the interval between selection cycles. In each selection cycle, the sample size was equal to 500 genotypes, each with four repetitions and four traits, whereas the selection intensity was 10% ($k_I = 1.755$); the interval lengths for the RLPSI and RLGSI were 4 and 1.5 years (Beyene et al. 2015) respectively.

Table 6.2 was divided in two parts. The first part presents the estimated RLPSI whereas the second part presents the estimated RLGSI selection responses. Columns 2, 3, and 4 in Table 6.2 present the estimated RLPSI and RLGSI selection responses not including the interval length, whereas columns 5, 6, and 7 present the estimated RLPSI and RLGSI selection response, including the interval length. The averages of the estimated RLPSI selection response not including the interval length for one, two, and three restrictions were 7.04, 5.50, and 3.90, whereas when the interval length was included, the averages were 1.76, 1.38, and 0.98 respectively. The averages of the estimated RLGSI selection response not including the interval length

Table 6.2 Estimated restricted linear phenotypic selection index (RLPSI) and RLGSI selection responses for 1, 2, and 3 null restrictions for 5 simulated selection cycles including and not including the interval between selection cycles. The interval lengths for the RLPSI and the RLGSI were 4 and 1.5 years respectively

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

[a]The estimated RLPSI selection response was divided by 4
[b]The estimated RLGSI selection response was divided by 1.5

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

Table 6.3 presents the estimated RLPSI (first part) and RLGSI (second part) expected genetic gain per trait not including the interval between selection cycles for one, two, and three null restrictions in five simulated selection cycles. In this case, RLPSI efficiency is greater than RLGSI efficiency because the averages of the estimated RLPSI expected genetic gain per trait were −2.52, 2.26, and 2.26 for one null restriction; 2.84 and 2.65 for two null restrictions; and 3.90 for three null restrictions. For the same set of restrictions, the averages of the estimated RLGSI expected genetic gain per trait were: −1.85, 1.13, and 2.06 for one null restriction; 1.52 and 2.19 for two null restrictions, and 2.79 for three null restrictions. However, divided by the interval length (4 years in the RLPSI), the averages of the estimated RLPSI expected genetic gain per trait were −0.63, 0.57, and 0.57 for one null restriction; 0.71 and 0.66 for two null restrictions, and 0.98 for three null restrictions. In a similar manner, dividing by the interval length (1.5 years in this case), the averages of the estimated RLGSI expected genetic gain per trait were −1.23, 0.75,

Table 6.3 Estimated RLPSI and RLGSI expected genetic gain per trait for 1, 2, and 3 null restrictions for 5 simulated selection cycles (each with 4 traits) not including the interval length between selection cycles

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

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

Table 6.4 presents the estimated RLPSI heritability ($\widehat{h}^2_{I_R}$) values, the estimated restricted linear genomic selection index (RLGSI) accuracy $\widehat{\rho}_{HI_{RG}}$) values, the values of $W = \dfrac{\widehat{\rho}_{HI_{RG}}}{\widehat{h}_{I_R}} L_{RP}$ ($L_{RP} = 4$), and the values of $\widehat{p} = 100\left(\widehat{\lambda}_R - 1\right)$, where $\widehat{\lambda}_R = \widehat{\rho}_{HI_R}/\widehat{\rho}_{HI_{RG}}$ and $\widehat{\rho}_{HI_R}$ is the estimated RLPSI accuracy, for one, two, and three restrictions for five simulated selection cycles. The RLGSI interval length was $L_{RG} = 1.5$ whereas the averages of the values of $W = \dfrac{\widehat{\rho}_{HI_{RG}}}{\widehat{h}_{I_R}} L_{RP}$ for each restriction were 1.22, 0.85, and 0.60; this means that the estimated Technow inequality (Technow et al. 2013), $L_{RG} < \dfrac{\widehat{\rho}_{HI_{RG}}}{\widehat{h}_{I_R}} L_{RP}$ (Chap. 5, Eq. 5.18), was not true. Thus, according to the Technow inequality results, for this data set, RLGSI efficiency in terms of time was not greater than RLPSI efficiency. The inequality $L_{RG} < \dfrac{\widehat{\rho}_{HI_G}}{\widehat{h}_{I_R}} L_{I_R}$ was not true because the estimated RLGSI accuracy was very low, whereas RLPSI heritability was high. Thus, note that the averages of the estimated RLGSI accuracy for one, two, and three null restrictions were 0.25, 0.19, and 0.14 respectively, and the averages of the estimated RLPSI heritability values were 0.70, 0.78 and 0.88, respectively. Thus, according to these results, because the estimated RLGSI accuracy is very low and

Table 6.4 Estimated RLPSI heritability ($\widehat{h}^2_{I_R}$), estimated RLGSI accuracy ($\widehat{\rho}_{HI_{RG}}$), estimated values of $W = \dfrac{\widehat{\rho}_{HI_{RG}}}{\widehat{h}_{I_R}} L_{RP}$ ($L_{RP} = 4$), and values of $\widehat{p} = 100\left(\widehat{\lambda}_R - 1\right)$, where $\widehat{\lambda}_R = \widehat{\rho}_{HI_R}/\widehat{\rho}_{HI_{RG}}$, and $\widehat{\rho}_{HI_R}$ are the estimated RLPSI accuracy values, for 1, 2, and 3 restrictions for five simulated selection cycles

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

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

The last three columns of Table 6.4, from left to right, present the estimated p values, $\widehat{p} = 100\left(\widehat{\lambda}_R - 1\right)$, for one, two, and three null restrictions in five simulated selection cycles. The average of the $\widehat{p}$ values indicates that for each of the three restrictions the RLPSI efficiency was 65.05%, 78.73%, and 74.09%, greater than RLGSI efficiency at predicting the net genetic merit. Thus, for this data set, the RLPSI was a better predictor of the net genetic merit than the RLGSI in each cycle.

6.2 The Predetermined Proportional Gain Linear Genomic Selection Index

6.2.1 *Objective of the PPG-LGSI*

Let $\mathbf{d}' = [d_1 \quad d_2 \quad \ldots \quad d_r]$ be a vector $1 \times r$ (r is the number of predetermined proportional gains) of the predetermined proportional gains imposed by the breeder, and assume that μ_q is the population mean of the qth trait before selection. The objective of the predetermined proportional gain linear genomic selection index (PPG-LGSI) is to change μ_q to $\mu_q + d_q$ in the testing population, where d_q is a predetermined change in μ_q. It is possible to solve this problem minimizing the mean squared difference between $I_G = \boldsymbol{\beta}'\boldsymbol{\gamma}$ and $H = \mathbf{w}'\mathbf{g}$, $E[(H - I_G)^2]$, under the restriction $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \theta_G\mathbf{d}$, where θ_G is a proportionality constant, or under the restriction $\mathbf{D}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{D}' = \begin{bmatrix} d_r & 0 & \ldots & 0 & -d_1 \\ 0 & d_r & \ldots & 0 & -d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & d_r & -d_{r-1} \end{bmatrix}$ is a matrix $(r-1) \times r$ (see Chap. 3 for details), and d_q ($q = 1, 2\ldots, r$) is the q^{th} element of vector $\mathbf{d}' = [d_1 \quad d_2 \quad \ldots \quad d_r]$; $\mathbf{U}'$ is a matrix $(t-1) \times t$ of 1s and 0s, and $\boldsymbol{\Gamma} = \left\{\sigma_{\gamma_{qq'}}\right\}$ (q, $q' = 1, 2, \ldots, t$, $t =$ number of traits) is a covariance matrix of additive genomic breeding values, $\boldsymbol{\gamma}' = [\gamma_1\, \gamma_2 \ldots \gamma_t]$.

6.2.2 *The Maximized PPG-LGSI Parameters*

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

$$f_P(\boldsymbol{\beta}, \mathbf{v}) = \boldsymbol{\beta}'\boldsymbol{\Gamma}\boldsymbol{\beta} + \mathbf{w}'\mathbf{C}\mathbf{w} - 2\mathbf{w}'\boldsymbol{\Gamma}\boldsymbol{\beta} + 2\mathbf{v}'\mathbf{D}'\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} \tag{6.7}$$

with respect to $\boldsymbol{\beta}$ and $\mathrm{v}' = [\,v_1 \quad v_2 \quad \ldots \quad v_{r-1}\,]$, where $\mathbf{v}'$ is a vector of Lagrange multipliers. From a mathematical point of view, Eq. (6.7) is equal to Eq. (6.1); thus, the vector of coefficients $\boldsymbol{\beta}$ of the PPG-LGSI should be similar to the vector of coefficients of the RLGSI (Eq. 6.3), i.e., the PPG-LGSI vector of coefficients is equal to

$$\boldsymbol{\beta}_{PG} = \mathbf{K}_P\mathbf{w}, \tag{6.8}$$

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

Under the restriction $\mathbf{U}'\boldsymbol{\Gamma}\boldsymbol{\beta} = \theta_G\mathbf{d}$ (see Chap. 3 for details) the vector of coefficients of the PPG-LGSI can be written as

$$\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}_{RG} + \theta_G\mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{d}, \tag{6.9}$$

where $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$ (Eq. 6.3), $\mathbf{K}_G = [\mathbf{I} - \mathbf{Q}_G]$, $\mathbf{Q}_G = \mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\Gamma}$, and $\mathbf{d}' = [\,d_1 \quad d_2 \quad \ldots \quad d_r\,]$ is the vector of the predetermined proportional gains imposed by the breeder. It can be shown that θ_G, the proportionality constant, can be written as

$$\theta_G = \frac{\mathbf{d}'(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\Gamma}\mathbf{w}}{\mathbf{d}'(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{d}}. \tag{6.10}$$

When $\theta_G = 0$, $\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}_{RG}$, and when $\mathbf{U}'$ is a null matrix, $\boldsymbol{\beta}_{PG} = \mathbf{w}$. Equations (6.8) and (6.9) give the same results, that is, both equations express the same result in a different mathematical way.

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

$$R_{PG} = \frac{k_I}{L_G}\sqrt{\boldsymbol{\beta}'_{PG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{PG}} \tag{6.11}$$

and

$$\mathbf{E}_{\mathrm{PG}} = \frac{k_I}{L_G}\frac{\boldsymbol{\Gamma}\boldsymbol{\beta}_{PG}}{\sqrt{\boldsymbol{\beta}'_{PG}\boldsymbol{\Gamma}\boldsymbol{\beta}_{PG}}}, \tag{6.12}$$

respectively, where L_G is the time required to complete a selection cycle using the PPG-LGSI. Equations (6.11) and (6.12) depend only on GEBV information.

6.2.3 Statistical Properties of the PPG-LGSI

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

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

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

6.2.4 Numerical Example

To illustrate the PPG-LGSI theory, we use the estimated matrix $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.21 & 2.95 & 5.00 \\ 2.95 & 42.41 & 71.11 \\ 5.00 & 71.11 & 121.53 \end{bmatrix}$ and the GEBVs associated with the traits GY (ton ha^{-1}), EHT (cm), and PHT (cm), described in Sect. 6.1.3.

It is necessary to estimate the PPG-LGSI vector of coefficients $\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}_{RG} + \theta_g\mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{d}$ (Eqs. 6.9 and 6.10). In Sect. 6.1.3, we showed that the estimated vectors of coefficients of $\boldsymbol{\beta}_{RG} = \mathbf{K}_G\mathbf{w}$ for the null restrictions $\mathbf{U}'_1 = [1 \quad 0 \quad 0]$ and $\mathbf{U}'_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ were $\widehat{\boldsymbol{\beta}}'_{RG1} = \mathbf{w}'\widehat{\mathbf{K}}'_{G1} = [3.78 \quad -0.1 \quad -0.1]$ and $\widehat{\boldsymbol{\beta}}'_{RG2} = \mathbf{w}'\widehat{\mathbf{K}}'_{G2} = [1.12 \quad 0.09 \quad -0.1]$ respectively, where $\mathbf{w}' = [5 \quad -0.1 \quad -0.1]$. This means that to estimate $\boldsymbol{\beta}_{PG} = \boldsymbol{\beta}_{RG} + \theta_G\mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{d}$, we need only to estimate $\theta_G\mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{d}$ for both sets of restrictions.

Consider matrix $\mathbf{U}'_1 = [1 \quad 0 \quad 0]$ and let $d_1 = 7.0$ be the predetermined proportional gain restriction for trait 1. We can estimate θ_G and $\mathbf{U}(\mathbf{U}'\boldsymbol{\Gamma}\mathbf{U})^{-1}\mathbf{d}$ as $\widehat{\theta}_{G1} = \dfrac{7.0\left(\mathbf{U}'_1\widehat{\boldsymbol{\Gamma}}\mathbf{U}_1\right)^{-1}\mathbf{U}'_1\widehat{\boldsymbol{\Gamma}}\mathbf{w}}{7.0\left(\mathbf{U}'_1\widehat{\boldsymbol{\Gamma}}\mathbf{U}_1\right)^{-1}7.0} = 0.036$ and $\mathbf{U}_1\left(\mathbf{U}'_1\widehat{\boldsymbol{\Gamma}}\mathbf{U}_1\right)^{-1}7.0 = \begin{bmatrix} 33.333 \\ 0 \\ 0 \end{bmatrix}$, whence the PPG-LGSI vector of coefficients was $\widehat{\boldsymbol{\beta}}_{PG_1} = \widehat{\boldsymbol{\beta}}_{RG_1} + \widehat{\theta}_{G_1}\mathbf{U}_1\left(\mathbf{U}'_1\widehat{\boldsymbol{\Gamma}}\mathbf{U}_1\right)^{-1}7.0 = \begin{bmatrix} 5.0 \\ -0.1 \\ -0.1 \end{bmatrix}$, and the estimated PPG-LGSI

was $\widehat{I}_{PG_1} = 5.0\text{GEBV}_1 - 0.1\text{GEBV}_2 - 0.1\text{GEBV}_3$. In a similar manner, we can estimate the PPG-LGSI vector of coefficients under restrictions $\mathbf{U}_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ and $\mathbf{d}_2' = [7 \quad -3]$. In this case, $\widehat{\boldsymbol{\beta}}_{PG_2} = \widehat{\boldsymbol{\beta}}_{RG_2} + \widehat{\theta}_{G_2}\mathbf{U}_2\left(\mathbf{U}_2'\widehat{\boldsymbol{\Gamma}}\mathbf{U}_2\right)^{-1}\mathbf{d}_2 = \begin{bmatrix} 4.97 \\ -0.18 \\ -0.10 \end{bmatrix}$ and the estimated PPG-LGSI was $\widehat{I}_{PG_2} = 4.97\text{GEBV}_1 - 0.18\text{GEBV}_2 - 0.1\text{GEBV}_3$.

Figure 6.2 presents the frequency distribution of the estimated PPG-LGSI values for one (Fig. 6.2a) and two (Fig. 6.2b) predetermined restrictions, $d = 7$ and $\mathbf{d}' = [7 \quad -3]$ respectively, obtained in a real testing population for one selection cycle in one environment. For both restrictions, the frequency distribution of the estimated PPG-LGSI values approaches the normal distribution.

Assume a selection intensity of 10% ($k_{I_G} = 1.755$); then, for one predetermined restriction, where $\mathbf{U}_1' = [1 \quad 0 \quad 0]$ and $d_1 = 7.0$, the estimated PPG-LGSI selection response and expected genetic gain per trait, not including the interval length, were $\widehat{R}_{PG_1} = k_{I_G}\sqrt{\widehat{\boldsymbol{\beta}}_{PG_1}'\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{PG_1}} = 1.05$ and $\widehat{\mathbf{E}}_{PG_1}' = k_I\dfrac{\widehat{\boldsymbol{\beta}}_{PG_1}'\widehat{\boldsymbol{\Gamma}}}{\sqrt{\widehat{\boldsymbol{\beta}}_{PG_1}'\widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\beta}}_{PG_1}}} = [0.74 \quad 9.92 \quad 16.54]$

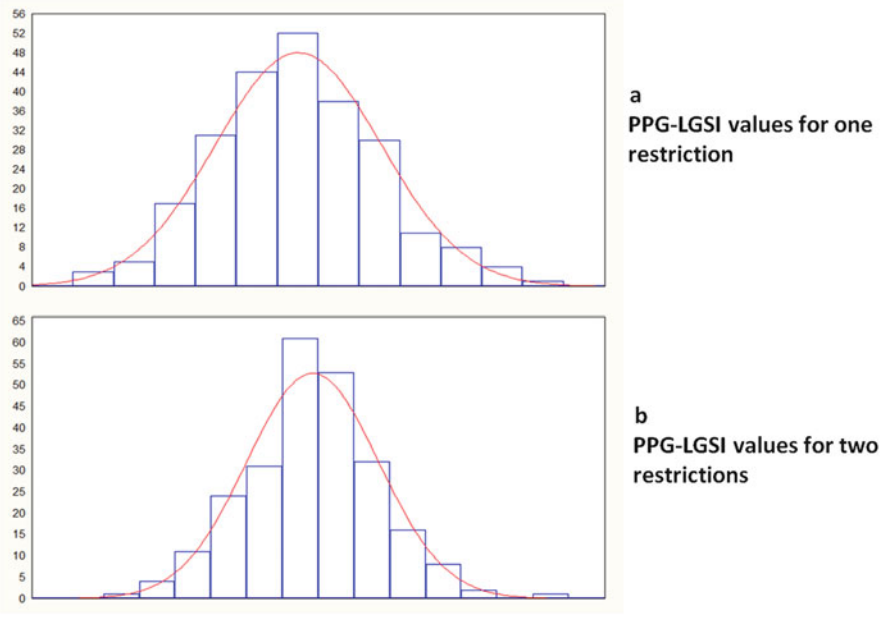

Fig. 6.2 Distribution of 380 estimated predetermined proportional gain linear genomic selection index (PPG-LGSI) values with one (**a**) and two (**b**) predetermined restrictions, $d = 7$ and $\mathbf{d}' = [7 \quad -3]$ respectively, obtained in a real testing population for one selection cycle in one environment

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

Now, we use the simulated data set described in Chap. 2, Sect. 2.8.1 to compare PPG-LGSI efficiency versus predetermined proportional gain linear phenotypic selection index (PPG-LPSI) efficiency. Let $\mathbf{U}_1' = [1 \quad 0 \quad 0]$, $\mathbf{U}_2' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, and $\mathbf{U}_3' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ be the matrices and $d_1 = 7$, $\mathbf{d}_2' = [7 \quad -3]$, and $\mathbf{d}_3' = [7 \quad -3 \quad 5]$ the vectors for one, two, and three predetermined restrictions respectively. Table 6.5 presents the estimated PPG-LPSI and PPG-LGSI selection response for each predetermined restriction in five simulated selection cycles including and not including the interval between selection cycles (4 years for the PPG-LPSI and 1.5 years for the PPG-LGSI); estimated PPG-LPSI and PPG-LGSI accuracy; and estimated variance of the predicted error (VPE). In each selection cycle, the sample size was equal to 500 genotypes, each with four repetitions and four traits. The selection intensity was 10% ($k_I = 1.755$).

The averages of the estimated PPG-LPSI selection response not including the interval length were 15.14, 14.87, and 13.30, whereas when the interval length was included, the average selection responses were 3.79, 3.72, and 3.33, for one, two, and three predetermined restrictions respectively (Table 6.5). The averages of the estimated PPG-LGSI selection responses not including the interval length for one, two, and three predetermined restrictions were 14.48, 13.47, and 11.26 respectively, and when the interval length was included, the selection responses were 9.65, 8.98, and 7.51 respectively (Table 6.5). These results indicate that when the interval length was included in the estimation of the PPG-LPSI and PPG-LGSI selection responses, PPG-LGSI efficiency was greater than PPG-LPSI efficiency, and vice versa, when the interval length was not included in the PPG-LPSI and PPG-LGSI selection responses, PPG-LPSI efficiency was higher than PPG-LGSI efficiency.

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

Table 6.5 Estimated predetermined proportional gain linear phenotypic and genomic selection index (PPG-LPSI and PPG-LGSI respectively) selection responses for 1, 2 and 3 predetermined restrictions for five simulated selection cycles including and not including the interval between selection cycles (4 years for the PPG-LPSI and 1.5 years for the PPG-LGSI); estimated PPG-LPSI and PPG-LGSI accuracy and estimated variance of the predicted error (VPE)

	Estimated PPG-LPSI selection response						Estimated PPG-LPSI accuracy			Estimated PPG-LPSI VPE		
	Not including interval length			Including interval length[a]								
Cycle	1	2	3	1	2	3	1	2	3	1	2	3
1	17.81	16.72	15.23	4.45	4.18	3.81	0.91	0.85	0.77	22.53	50.44	57.14
2	15.69	15.59	14.39	3.92	3.90	3.60	0.88	0.88	0.81	22.66	27.88	37.03
3	14.22	14.16	13.18	3.56	3.54	3.30	0.87	0.86	0.80	21.95	26.14	33.55
4	14.34	14.33	11.56	3.59	3.58	2.89	0.86	0.86	0.70	22.84	25.84	45.46
5	13.64	13.56	12.16	3.41	3.39	3.04	0.86	0.85	0.76	22.13	22.49	32.69
Average	15.14	14.87	13.30	3.79	3.72	3.33	0.88	0.86	0.77	22.42	30.56	41.17
	Estimated PPG-LGSI selection response						Estimated PPG-LGSI accuracy			Estimated PPG-LGSI VPE		
	Not interval length			Including interval length								
Cycle	1	2	3	1	2	3	1	2	3	1	2	3
1	21.22	17.83	16.66	14.15	11.89	11.10	0.65	0.91	0.85	74.75	22.25	35.43
2	13.90	13.52	10.65	9.27	9.01	7.10	0.72	0.70	0.55	58.66	62.06	84.62
3	13.59	13.15	10.70	9.06	8.77	7.13	0.70	0.67	0.55	63.49	67.33	86.30
4	12.30	11.84	9.36	8.20	7.89	6.24	0.61	0.59	0.46	83.26	86.91	103.93
5	11.38	11.03	8.96	7.59	7.35	5.97	0.56	0.54	0.44	93.62	96.21	109.63
Average	14.48	13.47	11.26	9.65	8.98	7.51	0.65	0.68	0.57	74.75	66.95	83.98

[a]The estimated LPSI selection response was divided by 4 and the estimated LGS selection response was divided by 1.5

Table 6.6 Estimated PPG-LPSI heritability ($\widehat{h}_P^2$), values of $W_P = \dfrac{\widehat{\rho}_{HI_G}}{\widehat{h}_P} L_P$ ($L_P = 4$), and the ratio of the estimated PPG-LPSI accuracy ($\widehat{\rho}_{HI_P}$) to the estimated PPG-LGSI accuracy ($\widehat{\rho}_{HI_{PG}}$): $\widehat{\lambda}_P = \widehat{\rho}_{HI_P}/\widehat{\rho}_{HI_{PG}}$, and values of $\widehat{p} = 100\left(\widehat{\lambda}_p - 1\right)$ for 1, 2 and 3 predetermined restrictions for five simulated selection cycles

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

Table 6.6 presents the estimated predetermined PPG-LPSI heritability ($\widehat{h}_P^2$) values, $W_P = \dfrac{\widehat{\rho}_{HI_G}}{\widehat{h}_P} L_P$ ($L_P = 4$) values, and ratio of the estimated PPG-LPSI accuracy ($\widehat{\rho}_{HI_P}$) to the estimated PPG-LGSI accuracy ($\widehat{\rho}_{HI_{PG}}$), i.e., $\widehat{\lambda}_P = \widehat{p}_{HI_P}/\widehat{p}_{HI_{PG}}$, and, finally, values of $\widehat{p} = 100\left(\widehat{\lambda}_P - 1\right)$ for one, two, and three null restrictions for five simulated selection cycles.

The averages of the W_P values for one, two, and three null restrictions were 3.29, 3.12, and 2.53, respectively, whereas the PPG-LGSI interval length was 1.5 ($L_G = 1.5$). This means that the estimated Technow inequality, $L_G < \dfrac{\widehat{\rho}_{HI_G}}{\widehat{h}_P} L_P$ (see Chap. 5, Eq. 5.18) was true. Thus, PPG-LGSI efficiency in terms of time was greater than PPG-LPSI efficiency for this data set. These results coincide with those obtained earlier in this chapter, when we compared PPG-LGSI efficiency versus PPG-LPSI efficiency in terms of interval length. However, the average values of $\widehat{p} = 100\left(\widehat{\lambda}_P - 1\right)$ (see Chap. 5, Eq. 5.15) were, in percentage terms, 16.80%, 20.76%, and 25.85% for each restriction. These latter results indicate that for this data set, the PPG-LPSI was a better predictor of the net genetic merit than the PPG-LGSI. This is because the estimated PPG-LPSI accuracies were higher than the estimated PPG-LPSI accuracies for this data set. We found similar results when we compared the PPG-LPSI VPE versus PPG-LGSI VPE (Table 6.5).

6.3 The Combined Restricted Linear Genomic Selection Index

The combined restricted linear genomic selection index (CRLGSI) is based on the RLPSI (Chap. 3) and combined linear genomic selection index (CLGSI, Chap. 5) theory. In the RLPSI, the breeder's objective is to improve only ($t - r$) of t ($r < t$)

traits, leaving r of them fixed; the same is true for the CRLGSI, but in the latter case, it is necessary to impose $2r$ restrictions, i.e., we need to fix r traits and their associated r GEBVs to obtain results similar to those obtained with the RLPSI. This is the main difference between the CRLGSI and the RLPSI.

It can be shown that $Cov(\mathrm{I}_C, \mathbf{a}_C) = \boldsymbol{\Psi}_C\boldsymbol{\beta}_C$ is the covariance between the breeding value vector ($\boldsymbol{a}'_C = [\mathbf{g}' \quad \boldsymbol{\gamma}']$) and the CLGSI, $I_C = \boldsymbol{\beta}'_C\mathbf{t}_C$ (see Chap. 5 for details), where $\mathbf{t}'_C = [\mathbf{y}' \quad \boldsymbol{\gamma}']$. In the CRLGSI, we want some covariances between the linear combinations of $\mathbf{a}_C$ ($\mathbf{U}'_C\mathbf{a}_C$) and CLGSI to be zero, i.e., $Cov(\mathrm{I}_C, \mathbf{U}'_C\mathbf{a}_C) = \mathbf{U}'_C\boldsymbol{\Psi}_C\boldsymbol{\beta}_C = \mathbf{0}$, where $\mathbf{U}'_C$ is a matrix $2(t-1) \times 2t$ of 1s and 0s (1 indicates that the trait and its associated GEBV are restricted, and 0 that the trait and its GEBV have no restrictions) and $\boldsymbol{\Psi}_C = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ is a block covariance matrix of $\mathbf{a}'_C = [\mathbf{g}' \quad \boldsymbol{\gamma}']$ where $\mathbf{C}$ and $\boldsymbol{\Gamma}$ are the covariance matrices of breeding ($\mathbf{g}$) and genomic ($\boldsymbol{\gamma}$) values respectively. This problem can be solved by minimizing the mean squared difference between the CLGSI and H ($E[(H - \mathrm{I}_C)^2]$) under the restriction $\mathbf{U}'_C\boldsymbol{\Psi}_C\boldsymbol{\beta}_C = \mathbf{0}$ similar to the RLGSI in Sect. 6.1.

6.3.1 The Maximized CRLGSI Parameters

Let $\mathbf{T}_C = \begin{bmatrix} \mathbf{P} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ be the block covariance matrix of $\mathbf{t}'_C = [\mathbf{y}' \quad \boldsymbol{\gamma}']$ where $\mathbf{P}$ and $\boldsymbol{\Gamma}$ are the covariance matrices of phenotypic ($\mathbf{y}$) and genomic ($\boldsymbol{\gamma}$) values respectively. Based on the Eq. (6.1) result, it can be shown that the CRLGSI vector of coefficients that minimizes $E[(H - \mathrm{I}_C)^2]$ under the restriction $\mathbf{U}'_C\boldsymbol{\Psi}_C\boldsymbol{\beta}_C = \mathbf{0}$ is

$$\boldsymbol{\beta}_{CR} = \mathbf{K}_C\boldsymbol{\beta}_C, \tag{6.13}$$

where $\mathbf{K}_C = [\mathbf{I} - \mathbf{Q}_C]$, $\mathbf{Q}_C = \mathbf{T}_C^{-1}\boldsymbol{\Phi}_C\left(\boldsymbol{\Phi}'_C\mathbf{T}_C^{-1}\boldsymbol{\Phi}_C\right)^{-1}\boldsymbol{\Phi}'_C$, $\boldsymbol{\Phi}_C = \mathbf{U}'_C\boldsymbol{\Psi}_C$, and $\boldsymbol{\beta}_C = \mathbf{T}_C^{-1}\boldsymbol{\Psi}_C\mathbf{a}_C$ (the vector of coefficients of the CLGSI, see Chap. 5 for details); $\mathbf{T}_C^{-1}$ is the inverse of matrix $\mathbf{T}_C$, and $\mathbf{I}$ is an identity matrix $2t \times 2t$. When no restrictions are imposed on any of the traits, $\mathbf{U}'_C$ is a null matrix and $\boldsymbol{\beta}_{CR} = \boldsymbol{\beta}_C$ (the vector of coefficients of the CLGSI). That is, the CRLGSI is more general than the CLGSI. Similar to the RLPSI and the RLGSI, matrices $\mathbf{K}_C$ and $\mathbf{Q}_C$ are idempotent ($\mathbf{K}_C = \mathbf{K}_C^2$ and $\mathbf{Q}_C = \mathbf{Q}_C^2$) and orthogonal ($\mathbf{K}_C\mathbf{Q}_C = \mathbf{Q}_C\mathbf{K}_C = \mathbf{0}$), that is, $\mathbf{K}_C$ and $\mathbf{Q}_C$ are projectors. Thus, we can assume that the CRLGSI has similar properties to those described for the RLPSI (see Chap. 3 for details) when matrices $\boldsymbol{\Psi}_C = \begin{bmatrix} \mathbf{C} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ and $\mathbf{T}_C = \begin{bmatrix} \mathbf{P} & \boldsymbol{\Gamma} \\ \boldsymbol{\Gamma} & \boldsymbol{\Gamma} \end{bmatrix}$ are known.

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

$$R_{CR} = \frac{k_I}{L_I}\sqrt{\boldsymbol{\beta}'_{CR}\mathbf{T}_C\boldsymbol{\beta}_{CR}} \tag{6.14}$$

and

$$\mathbf{E}_{CR} = \frac{k_I}{L_I}\frac{\boldsymbol{\Psi}\boldsymbol{\beta}_{CR}}{\sqrt{\boldsymbol{\beta}'_{CR}\mathbf{T}_C\boldsymbol{\beta}_{CR}}}, \tag{6.15}$$

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

6.3.2 Numerical Examples

To illustrate the CRLGSI theoretical results, we use a real training maize (*Zea mays*) F_2 population with 248 genotypes (each with two repetitions), 233 molecular markers, and three traits: GY (ton ha^{-1}), EHT (cm), and PHT (cm). Matrices $\mathbf{P}$ and $\mathbf{C}$ were estimated based on Eqs. (2.22) to (2.24) described in Chap. 2. The estimated matrices were $\widehat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}$ and $\widehat{\mathbf{C}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$. In a similar manner, we estimated matrix $\boldsymbol{\Gamma}$ using Eqs. (5.21) to (5.23) described in Chap. 5. The estimated matrix was $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}$.

To estimate the CRLGSI and its associated parameters (selection response, expected genetic gain per trait, etc.), we need to obtain matrices $\widehat{\mathbf{T}}_C = \begin{bmatrix} \widehat{\mathbf{P}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$ and $\widehat{\boldsymbol{\Psi}}_C = \begin{bmatrix} \widehat{\mathbf{C}} & \widehat{\boldsymbol{\Gamma}} \\ \widehat{\boldsymbol{\Gamma}} & \widehat{\boldsymbol{\Gamma}} \end{bmatrix}$ using phenotypic and genomic information and the estimated CRLGSI vector of coefficients $\widehat{\boldsymbol{\beta}}_{CR} = \widehat{\mathbf{K}}_C\widehat{\boldsymbol{\beta}}_C$, where $\widehat{\mathbf{K}}_C = \left[\mathbf{I} - \widehat{\mathbf{Q}}_C\right]$, $\widehat{\mathbf{Q}}_C = \widehat{\mathbf{T}}_C^{-1}\widehat{\boldsymbol{\Phi}}_C\left(\widehat{\boldsymbol{\Phi}}'_C\widehat{\mathbf{T}}_C^{-1}\widehat{\boldsymbol{\Phi}}_C\right)^{-1}\widehat{\boldsymbol{\Phi}}'_C$, $\widehat{\boldsymbol{\Phi}}_C = \mathbf{U}'_C\widehat{\boldsymbol{\Psi}}_C$, and $\widehat{\boldsymbol{\beta}}_C = \widehat{\mathbf{T}}_C^{-1}\widehat{\boldsymbol{\Psi}}_C\mathbf{a}_C$.

We have indicated that the main difference between the RLGSI and the CRLGSI is matrix $\mathbf{U}'_C$, on which we now need to impose two restrictions: one for the trait and another for its associated GEBV. Consider the (*Zea mays*) F_2 population described earlier and suppose that we restrict trait GY; then, matrix $\mathbf{U}'_C$ should be constructed as $\mathbf{U}'_{C_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$. If we restrict traits GY and EHT, matrix $\mathbf{U}'_C$ should

be constructed as $\mathbf{U}'_{C_2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$, etc. The procedure for obtaining matrices $\widehat{\mathbf{K}}_C = [\mathbf{I} - \widehat{\mathbf{Q}}_C]$, $\widehat{\mathbf{Q}}_C = \widehat{\mathbf{T}}_C^{-1}\widehat{\mathbf{\Phi}}_C\left(\widehat{\mathbf{\Phi}}'_C\widehat{\mathbf{T}}_C^{-1}\widehat{\mathbf{\Phi}}_C\right)^{-1}\widehat{\mathbf{\Phi}}'_C$, and $\widehat{\mathbf{\Phi}}_C = \mathbf{U}'_C\widehat{\mathbf{\Psi}}_C$ is similar to that described in Chap. 3.

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

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

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

Figure 6.3 presents the frequency distribution of the estimated CRLGSI values for one (Fig. 6.3a) and two null restrictions (Fig. 6.3b) using matrices $\mathbf{U}'_{C_1}$ and $\mathbf{U}'_{C_2}$, and the real data set of the F_2 population. For both restrictions, the frequency distribution of the estimated CRLGSI values approaches normal distribution.

Suppose a selection intensity of 10% ($k_I = 1.755$), matrix $\mathbf{U}'_{C_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ and that the vector of economic weights is $\mathbf{w}' = [5 \quad -0.1 \quad -0.1 \quad 0 \quad 0 \quad 0]$; then, according to the estimated matrices $\widehat{\mathbf{P}}$, $\widehat{\mathbf{C}}$, and $\widehat{\mathbf{\Gamma}}$ described earlier, the estimated CRLGSI selection response and the estimated CRLGSI expected genetic gain per trait were $\widehat{R}_{CR} = k_I\sqrt{\widehat{\boldsymbol{\beta}}'_{CR}\widehat{\mathbf{T}}_C\widehat{\boldsymbol{\beta}}_{CR}} = 0.96$ and $\widehat{\mathbf{E}}'_{CR} = k_I\dfrac{\widehat{\boldsymbol{\beta}}'_{CR}\widehat{\mathbf{\Psi}}}{\sqrt{\widehat{\boldsymbol{\beta}}'_{CR}\widehat{\mathbf{T}}_C\widehat{\boldsymbol{\beta}}_{CR}}} = [0 \quad -3.53 \quad -6.03 \quad 0 \quad -2.93 \quad -4.87]$ respectively, whereas the estimated CRLGSI accuracy was $\widehat{\rho}_{HI_{CR}} = \dfrac{\widehat{\sigma}_{I_{CR}}}{\widehat{\sigma}_H} = 0.51$ (see Chaps. 3 and 5 for details).

Now, we use the simulated data described in Chap. 2, Sect. 2.8.1 to compare CRLGSI efficiency versus RLGSI efficiency. The criteria for this comparison are the Technow inequality (Eq. 5.18, Chap. 5) and the ratio of the estimated CRLGSI accuracy ($\widehat{\rho}_{HI_{CR}}$) to the estimated RLGSI accuracy ($\widehat{\rho}_{HI_R}$) expressed as percentages (Eq. 5.17, Chap. 5), i.e., $\widehat{p} = 100\left(\widehat{\lambda}_{CR} - 1\right)$, where $\widehat{\lambda}_P = \widehat{\rho}_{HI_{CR}}/\widehat{\rho}_{HI_R}$, for one, two, and three null restrictions for five simulated selection cycles.

Table 6.7 presents the estimated CRLGSI heritability ($\widehat{h}^2_C$), the estimated RLGSI accuracy ($\widehat{\rho}_{HI_R}$), the values of $W_C = \dfrac{\widehat{\rho}_{HI_R}}{\widehat{h}_I}L_I$ ($L_I = 4$), and the values of

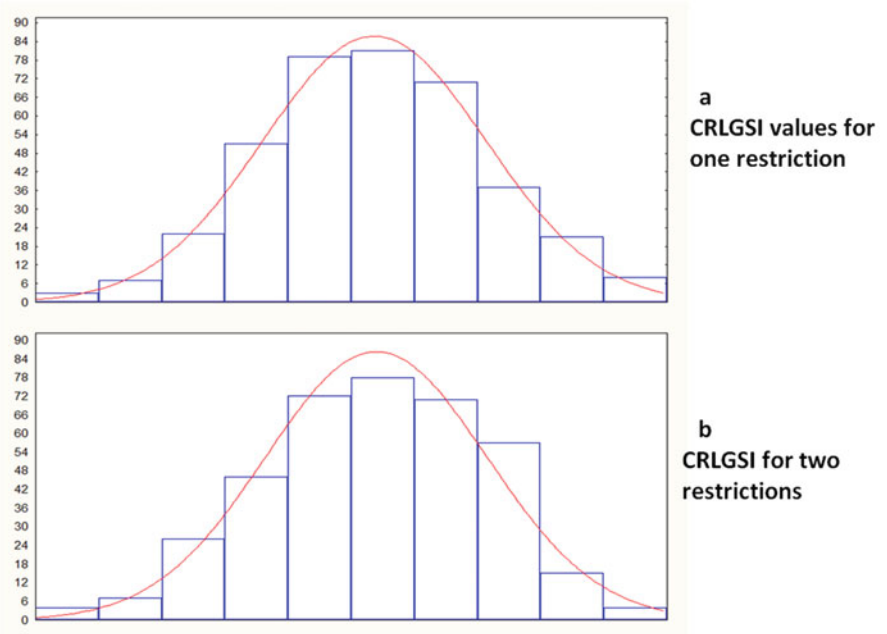

Fig. 6.3 Distribution of 244 estimated combined restricted linear genomic selection index (CRLGSI) values with one (**a**) and two (**b**) null restrictions respectively obtained in a real training population for one selection cycle in one environment

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

Table 6.7 Estimated combined restricted linear genomic selection index (CRLGSI) heritability ($\widehat{h}_I^2$), estimated RLGSI accuracy ($\widehat{\rho}_{HI_R}$), values of $W_C = \dfrac{\widehat{\rho}_{HI_R}}{\widehat{h}_I} L_I$ ($L_I = 4$), and values of $\widehat{p} = 100\left(\widehat{\lambda}_{CR} - 1\right)$, where $\widehat{\lambda}_{CR} = \widehat{\rho}_{HI_{CR}} / \widehat{\rho}_{HI_R}$ and $\widehat{\rho}_{HI_{CR}}$ is the estimated CRLGSI accuracy, for 1, 2, and 3 null restrictions for five simulated selection cycles

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

The last three columns of Table 6.7, from left to right, present the average of the values of $\widehat{p} = 100\left(\widehat{\lambda}_{CR} - 1\right)$, for one, two, and three null restrictions of five simulated selection cycles. According to these results, CRLGSI efficiency was 53.78%, 78.25%, and 61.25% higher than RLGSI efficiency. Thus, for this data set, the CRLGSI was a better predictor of the net genetic merit than the RLGSI.

6.4 The Combined Predetermined Proportional Gains Linear Genomic Selection Index

In the PPG-LPSI described in Chap. 3, the vector of the PPG (predetermined proportional gains) was $\mathbf{d}' = [d_1 \quad d_2 \quad \ldots \quad d_r]$. However, because the combined predetermined proportional gains LGSI (CPPG-LGSI) uses phenotypic and GEBV information jointly to predict the net genetic merit, the vector of the PPG ($\mathbf{d}_C$) should be twice the standard vector $\mathbf{d}'$, that is, $\mathbf{d}'_C = [d_1 \quad d_2 \quad \cdots \quad d_r \quad d_{r+1} \quad d_{r+2} \quad \cdots \quad d_{2r}]$, where we would expect that if d_1 is the PPG imposed on trait 1, then d_{r+1} should be the PPG imposed on the GEBV associated with trait 1, etc. In addition, in the CPPG-LGSI, we have three possible options for determining (for each trait and GEBV) the PPG, e.g., for trait 1, $d_1 = d_{r+1}$, $d_1 > d_{r+1}$, or $d_1 < d_{r+1}$. This is the main difference between the standard PPG-LPSI described in Chap. 3 and the CPPG-LGSI.

6.4.1 The Maximized CPPG-LGSI Parameters

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

$$\boldsymbol{\beta}_{CP} = \boldsymbol{\beta}_{CR} + \theta_{CP}\boldsymbol{\delta}_{CP}, \tag{6.16}$$

where

$$\theta_{\mathrm{CP}} = \frac{\boldsymbol{\beta}'_C\boldsymbol{\Phi}_C\left(\boldsymbol{\Phi}'_C\widehat{\mathbf{T}}_C^{-1}\boldsymbol{\Phi}_C\right)^{-1}\mathbf{d}_C}{\mathbf{d}'_C\left(\boldsymbol{\Phi}'_C\widehat{\mathbf{T}}_C^{-1}\boldsymbol{\Phi}_C\right)^{-1}\mathbf{d}_C} \tag{6.17}$$

is a proportionality constant. In addition, in Eq. (6.16), $\boldsymbol{\beta}_{CR} = \mathbf{K}_C\boldsymbol{\beta}_C$ is the vector of coefficients of the CRLGSI (Eq. 6.13), $\boldsymbol{\delta}_{CP} = \mathbf{T}_C^{-1}\boldsymbol{\Phi}_C\left(\boldsymbol{\Phi}'_C\widehat{\mathbf{T}}_C^{-1}\boldsymbol{\Phi}_C\right)^{-1}\mathbf{d}_C$, $\boldsymbol{\Phi}'_C = \mathbf{U}'_C\boldsymbol{\Psi}_C$, and $\boldsymbol{\beta}_C = \mathbf{T}_C^{-1}\boldsymbol{\Psi}_C\mathbf{a}_C$ (the vector of coefficients of the CLGSI). When $\theta_{\mathrm{CP}} = 0$, $\boldsymbol{\beta}_{CP} = \boldsymbol{\beta}_{CR}$, and if $\theta = 0$ and $\mathbf{U}'_C$ is the null matrix, then $\boldsymbol{\beta}_{CR} = \boldsymbol{\beta}_C$.

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

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

$$R_{CP} = \frac{k_I}{L_I}\sqrt{\boldsymbol{\beta}'_{CP}\mathbf{T}_C\boldsymbol{\beta}_{CP}} \tag{6.18}$$

and

$$\mathbf{E}_{CP} = \frac{k_I}{L_I}\frac{\boldsymbol{\Psi}\boldsymbol{\beta}_{CP}}{\sqrt{\boldsymbol{\beta}'_{CP}\mathbf{T}_C\boldsymbol{\beta}_{CP}}}, \tag{6.19}$$

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

6.4.2 Numerical Examples

Similar to the CRLGSI, to illustrate the CPPG-LGSI results we use the real training maize (*Zea mays*) F_2 population with 248 genotypes, 233 molecular markers, and three traits—GY (ton ha^{-1}), EHT (cm), and PHT (cm)—where $\widehat{\mathbf{P}} = \begin{bmatrix} 0.45 & 1.33 & 2.33 \\ 1.33 & 65.07 & 83.71 \\ 2.33 & 83.71 & 165.99 \end{bmatrix}$, $\widehat{\mathbf{C}} = \begin{bmatrix} 0.07 & 0.61 & 1.06 \\ 0.61 & 17.93 & 22.75 \\ 1.06 & 22.75 & 44.53 \end{bmatrix}$, and $\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} 0.07 & 0.65 & 1.05 \\ 0.65 & 10.62 & 14.25 \\ 1.05 & 14.25 & 26.37 \end{bmatrix}$ were the estimated matrices of $\mathbf{P}$, $\mathbf{C}$, and $\boldsymbol{\Gamma}$ respectively.

We can obtain the estimated CPPG-LGSI vector of coefficients as $\widehat{\boldsymbol{\beta}}_{CP} = \widehat{\boldsymbol{\beta}}_{CR} + \widehat{\theta}_{CP}\widehat{\boldsymbol{\delta}}_{CP}$ (Eq. 6.16). Suppose that we restrict trait GY and its associated GEBV with matrix $\mathbf{U}'_{C_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ and the vector of predetermined restriction $\mathbf{d}'_C = [7 \quad 3.5]$. In Sect. 6.3.2, we showed that the estimated CRLGSI vector of coefficients was $\widehat{\boldsymbol{\beta}}'_{CR} = [0.076 \quad -0.004 \quad -0.018 \quad 2.353 \quad -0.096 \quad -0.082]$; then, we only need to calculate $\widehat{\theta}_{CP}$ and $\widehat{\boldsymbol{\delta}}_{CP}$ to obtain the vector of coefficients $\widehat{\boldsymbol{\beta}}_{CP}$.

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

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

where GEBV_{GY}, GEBV_{EHT}, and GEBV_{PHT} are the GEBVs associated with traits GY, EHT, and PHT respectively. The same procedure is valid for two or more restrictions. Note that because $\widehat{\theta}_{CP} = 0.0003$ is very small, the estimated CPPG-LGSI and CRLGSI values were very similar.

Figure 6.4 presents the frequency distribution of the estimated CPPG-LGSI values for one (Fig. 6.4a) and two predetermined restrictions (Fig. 6.4b) using matrices $\mathbf{U}'_{C_1}$ and $\mathbf{U}'_{C_2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$, the vectors of the PPG

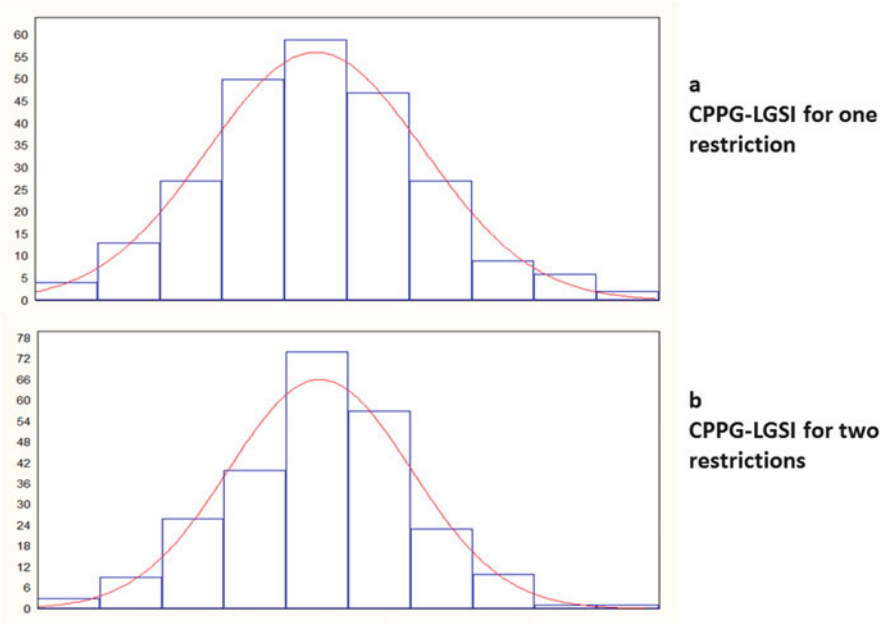

Fig. 6.4 Distribution of 244 estimated combined predetermined proportional gain linear genomic selection index (CPPG-LGSI) values with one (**a**) and two (**b**) predetermined restrictions, $d = 7$ and $\mathbf{d}' = [7 \quad -3]$ respectively, obtained in a real training population for one selection cycle in one environment

$\mathbf{d}'_{C1} = [7 \quad 3.5]$ and $\mathbf{d}'_{C2} = [7 \quad -3 \quad 3.5 \quad -1.5]$, and the real data set F_2. For both restrictions, the frequency distribution of the estimated CPPG-LGSI values approaches normal distribution.

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

Now, we use the simulated data described in Chap. 2, Sect. 2.8.1, to compare CPPG-LGSI efficiency versus PPG-LGSI efficiency. The criteria for this comparison are the Technow inequality (Chap. 5, Eq. 5.18) and the ratio of CPPG-LGSI accuracy ($\rho_{HI_{CP}}$) to PPG-LGSI accuracy (ρ_{HI_P}) expressed as percentages (Chap. 5, Eq. 5.17), $\widehat{p} = 100\left(\widehat{\lambda}_{CP} - 1\right)$, where $\widehat{\lambda}_{CP} = \widehat{\rho}_{HI_{CP}}/\widehat{\rho}_{HI_P}$ for one, two, and three null restrictions in five simulated selection cycles.

Table 6.8 presents the estimated CPPG-LGSI heritability ($\widehat{h}^2_I$), the estimated PPG-LGSI accuracy ($\widehat{\rho}_{HI_{CP}}$), values of $W_{CP} = \dfrac{\widehat{\rho}_{HI_G}}{\widehat{h}_I}L_I$ ($L_I = 4$) and $\widehat{p} = 100\left(\widehat{\lambda}_{CP} - 1\right)$, where $\widehat{\lambda}_P = \widehat{\rho}_{HI_{CP}}/\widehat{\rho}_{HI_P}$ and $\widehat{\rho}_{HI_P}$ is the estimated CPPG-LGSI accuracy, for one, two, and three null restrictions in five simulated selection cycles. The averages of the estimated W_{CP} values for one, two, and three predetermined restrictions were 3.60, 3.31, and 2.50 respectively, whereas the PPG-LGSI interval length was 1.5 ($L_G = 1.5$). This means that the estimated Technow inequality, $L_G < \dfrac{\widehat{\rho}_{HI_G}}{\widehat{h}_I}L_I$, was true. Thus, for this data set, PPG-LGSI efficiency is greater than CPPG-LGSI efficiency in terms of time.

The last three columns of Table 6.8, from left to right, present the values of $\widehat{p} = 100\left(\widehat{\lambda}_{CP} - 1\right)$, for one, two, and three null restrictions in five simulated selection cycles. The average values of $\widehat{p} = 100\left(\widehat{\lambda}_{CP} - 1\right)$ for each of the three restrictions, in percentage terms, were 37.19%, 32.82%, and 37.08% respectively. This means that the CPPG-LGSI efficiency was greater than PPG-LGSI efficiency at predicting the net genetic merit.

Table 6.8 Estimated combined predetermined proportional gain linear genomic selection index (CPPG-LGSI) heritability ($\widehat{h}_I^2$), estimated PPG-LGSI accuracy ($\widehat{\rho}_{HI_{CP}}$), values of $W_{CP} = \dfrac{\widehat{\rho}_{HI_G}}{\widehat{h}_I} L_I$ ($L_I = 4$), and $\widehat{p} = 100\left(\widehat{\lambda}_{CP} - 1\right)$, where $\widehat{\lambda}_P = \widehat{\rho}_{HI_{CP}} / \rho_{HI_P}$ and $\widehat{\rho}_{HI_P}$ is the estimated CPPG-LGSI accuracy, for one, two, and three null restrictions for five simulated selection cycles

	CPPG-LGSI heritability			PPG-LGSI accuracy			Values of W_{CP}			Values of $\widehat{p}$		
Cycle	1	2	3	1	2	3	1	2	3	1	2	3
1	0.41	0.41	0.85	0.65	0.91	0.85	6.25	5.68	3.69	24.62	−3.30	−12.94
2	0.75	0.86	0.84	0.72	0.70	0.55	3.33	3.02	2.40	25.00	24.29	38.18
3	0.78	0.85	0.81	0.70	0.67	0.55	3.17	2.91	2.44	30.00	32.84	40.00
4	0.78	0.84	0.82	0.61	0.59	0.46	2.76	2.57	2.03	49.18	49.15	56.52
5	0.80	0.83	0.82	0.56	0.54	0.44	2.50	2.37	1.94	57.14	61.11	63.64
Average	0.70	0.76	0.83	0.65	0.68	0.57	3.60	3.31	2.50	37.19	32.82	37.08

References

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

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