Chapter 2
# The Linear Phenotypic Selection Index Theory

Abstract The main distinction in the linear phenotypic selection index (LPSI) theory is between the net genetic merit and the LPSI. The net genetic merit is a linear combination of the true unobservable breeding values of the traits weighted by their respective economic values, whereas the LPSI is a linear combination of several observable and optimally weighted phenotypic trait values. It is assumed that the net genetic merit and the LPSI have bivariate normal distribution; thus, the regression of the net genetic merit on the LPSI is linear. The aims of the LPSI theory are to predict the net genetic merit, maximize the selection response and the expected genetic gains per trait (or multi-trait selection response), and provide the breeder with an objective rule for evaluating and selecting parents for the next selection cycle based on several traits. The selection response is the mean of the progeny of the selected parents, whereas the expected genetic gain per trait, or multi-trait selection response, is the population means of each trait under selection of the progeny of the selected parents. The LPSI allows extra merit in one trait to offset slight defects in another; thus, with its use, individuals with very high merit in one trait are saved for breeding even when they are slightly inferior in other traits. This chapter describes the LPSI theory and practice. We illustrate the theoretical results of the LPSI using real and simulated data. We end this chapter with a brief description of the quadratic selection index and its relationship with the LPSI.

2.1 Bases for Construction of the Linear Phenotypic Selection Index

The study of quantitative traits (QTs) in plants and animals is based on the mean and variance of phenotypic values of QTs. Quantitative traits are phenotypic expressions of plant and animal characteristics that show continuous variability and are the result of many gene effects interacting among them and with the environment. That is, QTs are the result of unobservable gene effects distributed across plant or animal genomes that interact among themselves and with the environment to produce the observable characteristic plant and animal phenotypes (Mather and Jinks 1971; Falconer and Mackay 1996).
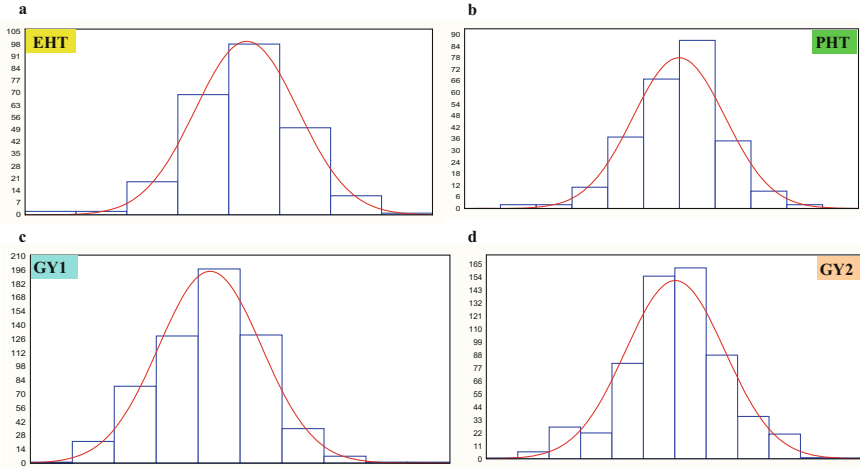
Fig. 2.1 Distribution of 252 phenotypic means of two maize (*Zea mays*) F$_2$ population traits: plant height (*PHT*, cm; **a**) and ear height (*EHT*, cm; **b**), evaluated in one environment, and of 599 - phenotypic means of the grain yield (*GY1* and *GY2*, ton ha^{-1}; **c** and **d** respectively) of one double haploid wheat (*Triticum aestivum* L.) population evaluated in two environments

The QTs are the traits that concern plant and animal breeders the most. They are particularly difficult to analyze because heritable variations of QTs are masked by larger nonheritable variations that make it difficult to determine the genotypic values of individual plants or animals (Smith 1936). However, as QTs usually have normal distribution (Fig. 2.1), it is possible to apply normal distribution theory when analyzing this type of data.

Any phenotypic value of QTs (*y*) can be divided into two main parts: one related to the genes and the interactions (*g*) among them (called genotype), and the other related to the environmental conditions (*e*) that affect genetic expression (called environment effects). Thus, the genotype is the particular assemblage of genes possessed by the plant or animal, whereas the environment consists of all the nongenetic circumstances that influence the phenotypic value of the plant or animal (Cochran 1951; Bulmer 1980; Falconer and Mackay 1996). In the context of only one environment, the phenotypic value of QTs (*y*) can be written as

$$y = g + e, \tag{2.1}$$

where *g* denotes the genotypic values that include all types of gene and interaction values, and *e* denotes the deviations from the mean of *g* values. For two or more environments, Eq. (2.1) can be written as *y* = *g* + *e* + *ge*, where *ge* denotes the interaction between genotype and environment. Assumptions regarding Eq. (2.1) are:

1. The expectation of e is zero, $E(e) = 0$.
2. Across several environments, the expectation of y is equal to the expectation of g, i.e., $E(g) = \mu_g = E(y) = \mu_y$.
3. The covariance between g and e is equal to 0.

The g value can be partitioned into three additional components: additive genetic (a) effects (or intra-locus additive allelic interaction), dominant genetic (d) effects (or intra-locus dominance allelic interaction), and epistasis (ι) effects (or inter-loci allelic interaction) such that $g = a + d + \iota$. In this book, we have assumed that $g = a$.

According to Kempthorne and Nordskog (1959), the following four theoretical conditions are necessary to construct a valid LPSI:

1. The phenotypic value (Eq. 2.1) shall be additively made up of two parts: a genotypic value (g) (defined as the average of the phenotypic values possible across a population of environments), and an environmental contribution (e).
2. The genotypic value g is composed entirely of the additive effects of genes and is thus the individual breeding value.
3. The genotypic economic value of an individual is its net genetic merit.
4. The phenotypic values and the net genetic merit are such that the regression of the net genetic merit on any linear function of the phenotypic values is linear.

Under assumptions 1 to 4, the offspring of a mating will have a genotypic value equal to the average of the breeding values of the parents (Kempthorne and Nordskog 1959). Additional conditions for practical objectives are:

5. Selection is practiced at only one stage of the life cycle.
6. The generations do not overlap.
7. All individuals below a certain level of desirability are culled without exception.
8. Selected individuals have equal opportunity to have offspring (Hazel and Lush 1942).
9. The LPSI values in the ith selection cycle and the LPSI values in the $(i + 1)$th selection cycle do not correlate.
10. The correlation between the LPSI and the net genetic merit should be at its maximum in each selection cycle.

Conditions 5 to 10 indicate that the LPSI is applying in a single stage context.

2.2   The Net Genetic Merit and the LPSI

Not all the individual traits under selection are equally important from an economic perspective; thus, the economic value of a trait determines how important that trait is for selection. Economic value is defined as the increase in profit achieved by improving a particular trait by one unit (Tomar 1983; Cartuche et al. 2014). This means that for several traits, the total economic value is a linear combination of the breeding values of the traits weighted by their respective economic values (Smith 1936; Hazel and Lush 1942; Hazel 1943; Kempthorne and Nordskog 1959); this is called *the net genetic merit of one individual* and can be written as

$$H = \mathbf{w}'\mathbf{g}, \tag{2.2}$$

where $\mathbf{g}' = [g_1 \ g_2 \ \ldots \ g_t]$ is a vector of true unobservable breeding values and $\mathbf{w}' = [w_1 \ w_2 \ \ldots \ w_t]$ is a vector of known and fixed economic weights. Equation (2.2) has several names, e.g., linear aggregate genotype (Hazel 1943), genotypic economic value (Kempthorne and Nordskog 1959), net genetic merit (Akbar et al. 1984; Cotterill and Jackson 1985), breeding objective (Mac Neil et al. 1997), and total economic merit (Cunningham and Tauebert 2009), among others. In this book, we call Eq. (2.2) *net genetic merit* only. The values of $H = \mathbf{w}'\mathbf{g}$ are unobservable but they can be simulated for specific studies, as is seen in the examples included in this chapter and in Chap. 10, where four indices have been simulated for many selection cycles.

In practice, the net genetic merit of an individual is not observable; thus, to select an individual as parent of the next generation, it is necessary to consider its overall merit based on several observable traits; that is, we need to construct an LPSI of observable phenotypic values such that the correlation between the LPSI and $H = \mathbf{w}'\mathbf{g}$ is at a maximum. The LPSI should be a good predictor of $H = \mathbf{w}'\mathbf{g}$ and should be useful for ranking and selecting among individuals with different net genetic merits. The LPSI for one individual can be written as

$$I = \mathbf{b}'\mathbf{y}, \tag{2.3}$$

where $\mathbf{b}' = [b_1 \ b_2 \ \cdots \ b_t]$ is the I vector of coefficients, t is the number of traits on I, and $\mathbf{y}' = [y_1 \ y_2 \ \cdots \ y_t]$ is a vector of observable trait phenotypic values usually centered with respect to its mean. The LPSI allows extra merit in one trait to offset slight defects in another. With its use, individuals with very high merit in some traits are saved for breeding, even when they are slightly inferior in other traits (Hazel and Lush 1942). Only one combination of $\mathbf{b}$ values allows the correlation of the LPSI with $H = \mathbf{w}'\mathbf{g}$ for a particular set of traits to be maximized.

Figure 2.2 indicates that the regression of the net genetic merit on the LPSI is lineal and that the correlation between the LPSI and the net genetic merit is maximal in each selection cycle. Also, note that the true correlations between the LPSI and the net genetic merit, and the true regression coefficients of the net genetic merit over the LPSI are the same, but the estimated correlation values between the LPSI and the net genetic merit are lower than the true correlation (Fig. 2.2). Table 2.1 indicates that the LPSI in the ith selection cycle and the LPSI in the $(i + 1)$th selection cycle do not correlate. However, in practice, the correlation values between any pair of LPSIs could be different from zero in successive selection cycles.

One fundamental assumption of the LPSI is that $I = \mathbf{b}'\mathbf{y}$ has normal distribution. This assumption is illustrated in Fig. 2.3 for two real datasets: a maize (*Zea mays*) F_2 population with 252 lines and three traits—grain yield (ton ha^{-1}); plant height (cm) and ear height (cm)—evaluated in one environment; and a double haploid wheat (*Triticum aestivum* L.) population with 599 lines and one trait—grain yield (ton ha^{-1})—evaluated in three environments. Figure 2.3 indicates that, in effect, the LPSI values approach normal distribution when the number of lines is very large.
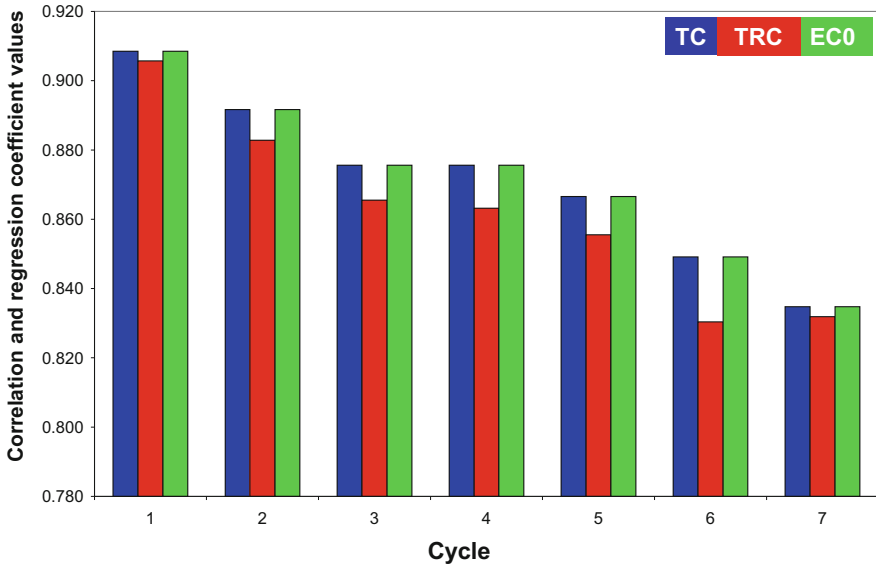
Fig. 2.2 True correlation (*TC*) and estimated correlation (*ECO*) values between the linear phenotypic selection index (*LPSI*) and the net genetic merit for seven selection cycles, and true regression coefficient (*TRC*) of the net genetic merit over the LPSI for four traits and 500 genotypes in one environment simulated for seven selection cycles

Table 2.1 Estimated correlation values between the linear phenotypic selection index (LPSI) values in seven simulated selection cycles

1	2	3	4	5	6	7
1.000	0.199	0.256	0.220	0.168	0.225	0.123
0.199	1.000	0.225	0.252	0.284	0.292	0.362
0.256	0.225	1.000	0.198	0.276	0.267	0.213
0.220	0.252	0.198	1.000	0.258	0.224	0.240
0.168	0.284	0.276	0.258	1.000	0.269	0.195
0.225	0.292	0.267	0.224	0.269	1.000	0.325
0.123	0.362	0.213	0.240	0.195	0.325	1.000

2.3   Fundamental Parameters of the LPSI

There are two fundamental parameters associated with the LPSI theory: the selection response (*R*) and the expected genetic gain per trait (**E**). In general terms, the selection response is the difference between the mean phenotypic values of the offspring (μ_O) of the selected parents and the mean of the entire parental generation (μ_P) before selection, i.e., $R = \mu_O - \mu_P$ (Hazel and Lush 1942; Falconer and Mackay 1996). The expected genetic gain per trait (or multi-trait selection response) is the covariance between the breeding value vector and the LPSI (*I*) values weighted by the standard deviation of the variance of $I(\sigma_I)$, i.e., $\frac{Cov(I,\mathbf{g})}{\sigma_I} = \frac{\mathbf{Gb}}{\sigma_I}$, multiplied by the
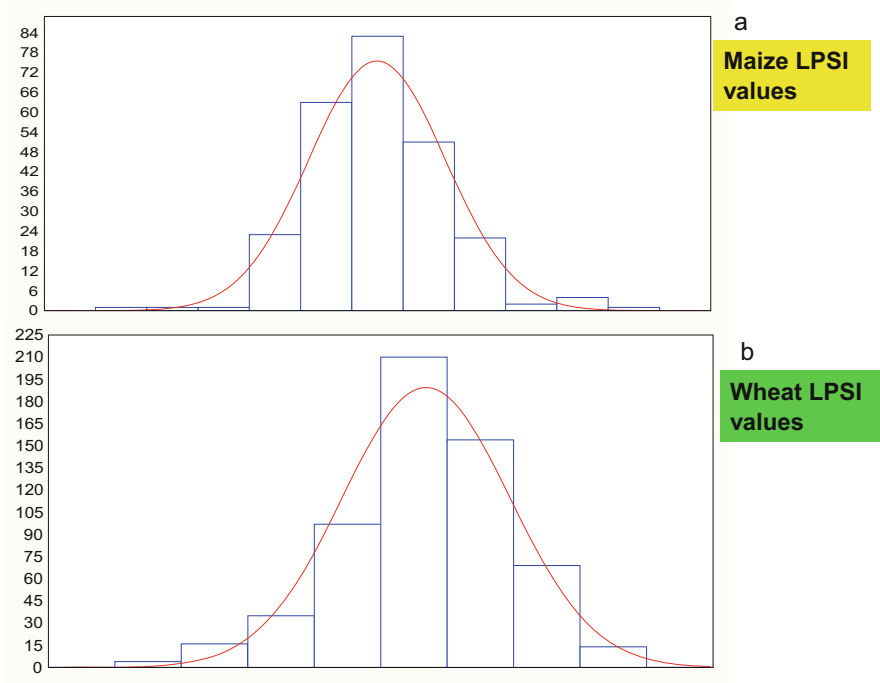
Fig. 2.3 *Maize LPSI* (Fig. 2.3a) is the distribution of 252 values of the LPSI constructed with the phenotypic means of three maize (*Zea mays*) F_2 population traits: grain yield (ton ha^{-1}), PHT (cm) and EHT (cm), evaluated in one environment. *Wheat LPSI* (Fig. 2.3b) is the distribution of 599 LPSI values constructed with the phenotypic means of the grain yield (ton ha^{-1}) of a double haploid wheat (*Triticum aestivum* L.) population evaluated in three environments

selection intensity. This is one form of the LPSI multi-trait selection response. In the univariate context, the expected genetic gain per trait is the same as the selection response.

One additional way of defining the selection response is based on the selection differential (D). The selection differential is the mean phenotypic value of the individuals selected as parents (μ_S) expressed as a deviation from the population mean (μ_P) or parental generation before the selection was made (Falconer and Mackay 1996); that is, $D = \mu_S - \mu_P$. Thus, another way of defining R is as the part of the expected differential of selection ($D = \mu_S - \mu_P$) that is gained when selection is applied (Kempthorne and Nordskog 1959); that is

$$R = \frac{Cov(g, y)}{\sigma_y^2} D = k\sigma_y h^2, \qquad (2.4)$$

where $Cov(g, y) = \sigma_g^2$ is the covariance between g and y, g is the individual breeding value associated with trait y, σ_y^2 is the variance of y, $k = \frac{D}{\sigma_y}$ is the standardized

selection differential or selection intensity, and $h^2 = \dfrac{\sigma_g^2}{\sigma_y^2}$ is the heritability of trait y in the base population. Heritability (h^2) appears in Eq. (2.4) as a measure of the accuracy with which animals or plants having the highest genetic values can be chosen by selecting directly for phenotype (Hazel and Lush 1942).

The selection response (Eq. 2.4) is the mean of the progeny of the selected parents or the future population mean of the trait under selection (Cochran 1951). Thus, the selection response enables breeders to estimate the expected progress of the selection before carrying it out. This information gives improvement programs a clearer orientation and helps to predict the success of the selection method adopted and choose the option that is technically most effective on a scientific base (Costa et al. 2008). Equation (2.4) is very powerful but its application requires strong assumptions. For example, Eq. (2.4) assumes that the trait of interest does not correlate with other traits having causal effects on fitness and, in its multivariate form the validity of predicted change rests on the assumption that all such correlated traits have been measured and incorporated into the analysis (Morrissey et al. 2010).

2.3.1 The LPSI Selection Response

The univariate selection response (Eq. 2.4) can also be rewritten as

$$R = k\sigma_y h^2 = k\sigma_g \rho_{gy}, \qquad (2.5)$$

where σ_g was defined in Eq. (2.4) and ρ_{gy} is the correlation between g and y. Thus, as $H = \mathbf{w}'\mathbf{g}$ and $I = \mathbf{b}'\mathbf{y}$ are univariate random variables, the selection response of the LPSI (R_I) can be written in a similar form as Eq. (2.5), i.e.,

$$R_I = k_I \sigma_H \rho_{HI}, \qquad (2.6)$$

where σ_H and σ_I are the standard deviation and ρ_{HI} the correlation between $H = \mathbf{w}'\mathbf{g}$ and $I = \mathbf{b}'\mathbf{y}$ respectively; $k_I = \frac{\mu_{IA} - \mu_{IB}}{\sigma_I}$ is the standardized selection differential or the selection intensity associated with the LPSI; μ_{IA} and μ_{IB} are the means of the LPSI values after and before selection respectively. The second part of Eq. (2.6) ($k_I \sigma_H \rho_{HI}$) indicates that the genetic change due to selection is proportional to k_I, σ_H, and ρ_{HI} (Kempthorne and Nordskog 1959). Thus, the genetic gain that can be achieved by selecting for several traits simultaneously within a population of animals or plants is the product of the selection differential (k_I), the standard deviation of $H = \mathbf{w}'\mathbf{g}$ (σ_H), and the correlation between $H = \mathbf{w}'\mathbf{g}$ and $I = \mathbf{b}'\mathbf{p}$ (ρ_{HI}). Selection intensity k_I is limited by the rate of reproduction of each species, whereas σ_H is relatively beyond man's control; hence, the greatest opportunity for increasing selection progress is by ensuring that ρ_{HI} is as large as possible (Hazel 1943). In general, it is assumed that k_I

and σ_H are fixed and **w** known and fixed; hence, R_I is maximized when ρ_{HI} is maximized only with respect to the LPSI vector of coefficients **b**.

Equation (2.6) is the mean of $H = \mathbf{w}'\mathbf{g}$, whereas $\sigma_H^2 \rho_{HI}^2 (1 - v)$ is its variance and $\rho_{HI}^* = \rho_{HI} \sqrt{\dfrac{1 - v}{1 - v\rho_{HI}^2}}$ the correlation between $H = \mathbf{w}'\mathbf{g}$ and $I = \mathbf{b}'\mathbf{p}$ after selection was carried out (Cochran 1951), where $v = k_I(k_I - \tau)$ and τ is the truncation point. For example, if the selection intensity is 5%, $k_I = 2.063$, $\tau = 1.645$, and $v = 0.862$ (Falconer and Mackay 1996, Table A). In R (in this case R denotes a platform for data analysis, see Kabakoff 2011 for details), the truncation point and selection intensity can be obtained as $v <- qnorm(1 - q)$ and $k <- dnorm(v)/q$, respectively, where q is the proportion retained. Both the variance and the correlation (ρ_{HI}^*) are reduced by selection. If $H = \mathbf{w}'\mathbf{g}$ could be selected directly, the gain in $H = \mathbf{w}'\mathbf{g}$ would be k_I. Thus, the gain due to indirect selection using $I = \mathbf{b}'\mathbf{p}$ is a fraction ρ_{HI} of that due to direct selection using $H = \mathbf{w}'\mathbf{g}$. As k_I increases, R_I increases (Eq. 2.6), $\sqrt{\sigma_H^2 \rho_{HI}^2 (1 - v)}$ and ρ_{HI}^* decrease, and the effects are in the same direction as ρ_{HI}^* increases (Cochran 1951). These results should be valid for all selection indices described in this book.

Smith (1936) gave an additional method to obtain Eq. (2.6). Suppose that we have a large number of plant lines and we select one proportion q for further propagation. In addition, assume that the values of I for each line are normally distributed with variance $\sigma_I^2 = \mathbf{b}'\mathbf{Pb}$; let I be transformed into a variable u, with unit variance and mean at zero, that is, $u = \frac{I - \mu_I}{\sigma_I}$, where μ_I is the mean of I. Assume that all I values higher than I' value are selected; then the value of $u' = \frac{I' - \mu_I}{\sigma_I}$ corresponding to any given value of q may be ascertained from a table of the standard normal probability integral (Fig. 2.4).

Assuming that the expectations of H and I are $E(H) = 0$ and $E(I) = \mu_I$, the conditional expectation of H given I can be written as
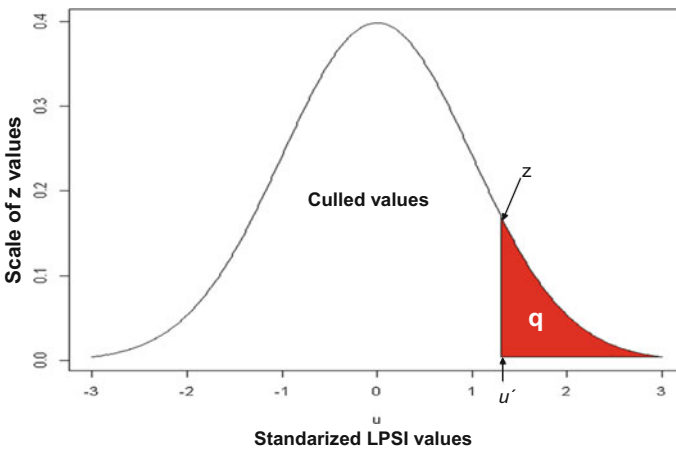
Fig. 2.4 Graph of standardized LPSI values showing how a population can be separated sharply at a given point (u') into a selected fraction (q), denoted by the *red area*, and a remainder that is culled, denoted by the *white area*

$E(H/I) = \frac{\sigma_{HI}}{\sigma_I^2}[I - \mu_I] = \frac{\sigma_{HI}}{\sigma_I^2}\sigma_I u = B\sigma_I u$, where $B = \frac{\sigma_{HI}}{\sigma_I^2}$, $\sigma_{HI} = \mathbf{w}'\mathbf{Gb}$ is the covariance between H and I, and $\sigma_I^2 = \mathbf{b}'\mathbf{Pb}$ is the variance of I. Therefore, if σ_I^2 and σ_{HI} are fixed, the LPSI selection response (R_I) can be obtained as the expectation of the selected population, which has univariate left truncated normal distribution. 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 LPSI context, a truncation distribution occurs when a sample of individuals from the parent distribution is selected as parents for the next selection cycle, thus creating a new population of individuals that follow a truncated normal distribution. Thus, we need to find $E[E(H/I)] = q^{-1}B\sigma_I E(u)$, or, using integral calculus,

$$E[E(H/I)] = \frac{B\sigma_I}{q} \int_{u=u'}^{\infty} \frac{u}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}u^2\right\} du = \frac{z}{q}\sigma_H\rho_{HI}, \qquad (2.7)$$

where $z = \frac{\exp\{-0.5u'^2\}}{\sqrt{2\pi}}$ is the height of the ordinate of the normal curve at the lowest value of u' retained and q is the proportion of the population of animal or plant lines that is selected (Fig. 2.4). The proportion q that must be saved depends on the reproductive rate and longevity of the species under consideration and on whether the population is expanding, stationary or declining in numbers. The ordinate (z) of the normal curve is determined by the proportion selected (q) (Fig. 2.4). The amount of progress is expected to be larger as q becomes smaller; that is, as selection becomes more intense (Hazel and Lush 1942). Kempthorne and Nordskog (1959) showed that $\frac{z}{q} = k_I$. Thus, Eqs. (2.6) and (2.7) are the same, that is, $E[E(H/I)] = R_I$.

2.3.2   The Maximized Selection Response

The main objective of the LPSI is to maximize the mean of $H = \mathbf{w}'\mathbf{g}$ (Eq. 2.7). Assuming that $\mathbf{P}$, $\mathbf{G}$, $\mathbf{w}$, and k_I are known, to maximize R_I we can either maximize ρ_{HI} or minimize the mean squared difference between I and H, $E[(H - I)^2] = \mathbf{w}'\mathbf{Gw} + \mathbf{b}'\mathbf{Pb} - 2\mathbf{w}'\mathbf{Gb}$ with respect to $\mathbf{b}$, that is, $\frac{\partial}{\partial \mathbf{b}}E\left[(H - I)^2\right] = 2\mathbf{Pb} - 2\mathbf{Gw} = \mathbf{0}$, from where

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw} \qquad (2.8)$$

is the vector that simultaneously minimizes $E[(H - I)^2]$ and maximizes ρ_{HI}, and then $R_I = k_I\sigma_H\rho_{HI}$.

By Eq. (2.8), the maximized LPSI selection response can be written as

$$R_I = k_I \sqrt{\mathbf{b'Pb}}. \tag{2.9}$$

The maximized LPSI selection response predicts the mean improvement in H due to indirect selection on I only when $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$ (Harris 1964) and is proportional to the standard deviation of the LPSI variance (σ_I) and the standardized selection differential or the selection intensity (k_I).

The maximized LPSI selection response (Eq. 2.9) it related to the Cauchy–Schwarz inequality (Rao 2002; Cerón-Rojas et al. 2006), which establishes that for any pair of vectors $\mathbf{u}$ and $\mathbf{v}$, if $\mathbf{A}$ is a positive definite matrix, then the inequality $(\mathbf{u'v})^2 \leq (\mathbf{v'Av})(\mathbf{u'A}^{-1}\mathbf{u})$ holds. Kempthorne and Nordskog (1959) proved that maximizing $\rho_{HI}^2 = \dfrac{(\mathbf{w'Gb})^2}{(\mathbf{w'Gw})(\mathbf{b'Pb})}$ also maximizes R_I. According to Eqs. (2.6) and (2.7), R_I^2 can be written as $R_I^2 = k_I^2 \dfrac{(\mathbf{w'Gb})^2}{(\mathbf{b'Pb})}$, such that maximizing R_I^2 is equivalent to maximizing $\dfrac{(\mathbf{w'Gb})^2}{(\mathbf{b'Pb})}$. Let $\mathbf{Gw} = \mathbf{u}$, $\mathbf{b} = \mathbf{v}$, and $\mathbf{A} = \mathbf{P}$, by the Cauchy–Schwarz inequality $\dfrac{(\mathbf{w'Gb})^2}{(\mathbf{b'Pb})} \leq \mathbf{w'GP}^{-1}\mathbf{Gw}$. This implies that the maximum is reached when $\dfrac{(\mathbf{w'Gb})^2}{(\mathbf{b'Pb})} = \mathbf{w'GP}^{-1}\mathbf{Gw}$, at which point $R_I = k_I \sqrt{\mathbf{w'GP}^{-1}\mathbf{Gw}}$. This latter result is the same as Eq. (2.9) when $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$.

Result $R_I = k_I \sqrt{\mathbf{w'GP}^{-1}\mathbf{Gw}}$ obtained using the Cauchy–Schwarz inequality corroborates that $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$ (Eq. 2.8) is a global minimum when the mean squared difference between I and H ($E[(H - I)^2]$) is minimized, and a global maximum when the correlation ρ_{HI} between I and H is maximized because $R_I = k_I \sqrt{\mathbf{b'Pb}} = k_I \sqrt{\mathbf{w'GP}^{-1}\mathbf{Gw}}$ only when $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$.

2.3.3 The LPSI Expected Genetic Gain Per Trait

Whereas $R = \dfrac{Cov(g, y)}{\sigma_y^2} D$ (Eq. 2.4) denotes the selection response in the univariate case, $\mathbf{E} = \dfrac{Cov(I, \mathbf{g})}{\sigma_I}$ denotes the LPSI expected genetic gain per trait. Also, except by $\dfrac{D}{\sigma_y}$, $\dfrac{Cov(g, y)}{\sigma_y}$ and $\dfrac{Cov(I, \mathbf{g})}{\sigma_I}$ are mathematically equivalent and whereas $\dfrac{Cov(g, y)}{\sigma_y}$ is the covariance between g and y weighted by the standard deviation of the variance of y, $\dfrac{Cov(I, \mathbf{g})}{\sigma_I}$ is the covariance between the breeding value vector and the LPSI values weighted by the standard deviation of the variance of LPSI. This means that in effect, $\mathbf{E}$ is the LPSI multi-trait selection response and can be written as

$$\mathbf{E} = k_I \frac{\mathbf{Gb}}{\sigma_I}, \tag{2.10}$$

where $\mathbf{G}$, σ_I and k_I were defined earlier. As Eq. (2.10) is the covariance between $I = \mathbf{b}'\mathbf{p}$ and $\mathbf{g}' = [g_1 \quad g_2 \quad \dots \quad g_t]$ divided by σ_I, considering g_j and $I = \sum_{j=1}^{t} b_j y_j$, the genetic gain in the jth index trait due to selection on I will be

$$\frac{k_I}{\sigma_I} Cov(I, g_j) = \frac{k_I}{\sigma_I}\left[b_1\sigma_{1j} + b_2\sigma_{2j} + \cdots + b_j\sigma_j^2 + \cdots + b_t\sigma_{tj}\right] = k_I \frac{\mathbf{b}'\boldsymbol{\sigma}_j}{\sigma_I}, \quad (2.11)$$

where $\boldsymbol{\sigma}_j' = \begin{bmatrix} \sigma_{1j} & \cdots & \sigma_j^2 & \cdots & \sigma_{tj} \end{bmatrix}$ is a vector of genotypic covariances of the jth index trait with all the index traits (Lin 1978; Brascamp 1984).

If Eq. (2.11) is multiplied by its economic weight, we obtain a measure of the economic value of each trait included in the net genetic merit (Cunningham and Tauebert 2009). In percentage terms, the economic value attributable to genetic change in the jth trait can be written as

$$w_j \frac{\mathbf{b}'\boldsymbol{\sigma}_j}{\sigma_I^2} 100. \quad (2.12)$$

In addition, the percentage reduction in the net genetic merit of overall genetic gain if the jth trait is omitted from the LPSI (Cunningham and Tauebert 2009) is

$$\left[1 - \sqrt{1 - \frac{b_j^2}{\sigma_I^2 \varphi_j^{-2}}}\right] 100, \quad (2.13)$$

where φ_j^{-2} is the jth diagonal element of the inverse of the phenotypic covariance matrix $\mathbf{P}^{-1}$ and b_j^2 the square of the jth coefficient of the LPSI. Equations (2.12) and (2.13) are measures of the importance of each trait included in the LPSI when makes selection.

2.3.4 Heritability of the LPSI

As the variance of $I = \mathbf{b}'\mathbf{y}$ is equal to $\sigma_I^2 = \mathbf{b}'\mathbf{P}\mathbf{b} = \mathbf{b}'\mathbf{G}\mathbf{b} + \mathbf{b}'\mathbf{R}\mathbf{b}$, where $\mathbf{P} = \mathbf{G} + \mathbf{R}$, $\mathbf{G}$ and $\mathbf{R}$ are the phenotypic, genetic, and residual covariance matrices respectively, then the LPSI heritability (Lin and Allaire 1977; Nordskog 1978) can be written as

$$h_I^2 = \frac{\mathbf{b}'\mathbf{G}\mathbf{b}}{\mathbf{b}'\mathbf{P}\mathbf{b}}. \quad (2.14)$$

When selecting a trait, the correlation between the phenotypic and genotypic values is equal to the square root of the trait's heritability ($\rho_{gy} = h$); however, in the LPSI context, when $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{w}$, the maximized correlation between H and I is $\rho_{HI} = \sqrt{\frac{\mathbf{b}'\mathbf{P}\mathbf{b}}{\mathbf{w}'\mathbf{G}\mathbf{w}}} = \frac{\sigma_I}{\sigma_H}$, whereas $h_I = \sqrt{\frac{\mathbf{b}'\mathbf{G}\mathbf{b}}{\mathbf{b}'\mathbf{P}\mathbf{b}}}$ is the square root of I heritability; that is,
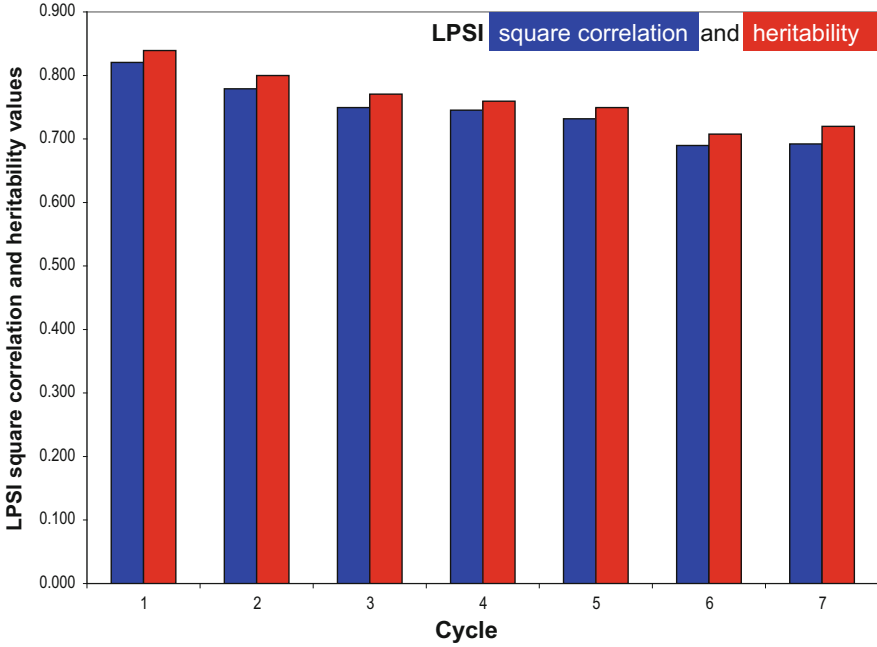
Fig. 2.5 Estimated values of the square correlation between the LPSI and the net genetic merit $(H = \mathbf{w}'\mathbf{g})$ and the LPSI heritability for four traits and 500 genotypes in one environment simulated for seven selection cycles

from a mathematical point of view, $\rho_{HI} \neq h_I$. In practice, h_I^2 and ρ_{HI}^2 give similar results (Fig. 2.5).

2.4   Statistical LPSI Properties

Assuming that H and I have joint bivariate normal distribution, $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$, and $\mathbf{P}$, $\mathbf{G}$ and $\mathbf{w}$ are known, the statistical LPSI properties (Henderson 1963) are the following:

1. The variance of I (σ_I^2) and the covariance between H and I (σ_{HI}) are equal, i.e., $\sigma_I^2 = \sigma_{HI}$. We can demonstrate this property noting that as $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$, $\sigma_I^2 = \mathbf{b}'\mathbf{Pb}$, and $\sigma_{HI} = \mathbf{w}'\mathbf{Gb}$, then $\sigma_I^2 = (\mathbf{w}'\mathbf{GP}^{-1})\mathbf{PP}^{-1}\mathbf{Gw} = \mathbf{w}'\mathbf{GP}^{-1}\mathbf{Gw}$, and $\sigma_{HI} = \mathbf{w}'\mathbf{GP}^{-1}\mathbf{Gw}$; i.e., $\sigma_I^2 = \sigma_{HI}$. This last result implies that when $\mu_I = 0$, $E(H/I) = I$.

2. The maximized correlation between H and I is equal to $\rho_{HI} = \frac{\sigma_I}{\sigma_H}$. That is,

$$\rho_{HI} = \frac{\mathbf{w}'\mathbf{Gb}}{\sqrt{\mathbf{w}'\mathbf{Gw}}\sqrt{\mathbf{b}'\mathbf{Pb}}} = \frac{\mathbf{w}'\mathbf{GP}^{-1}\mathbf{Gw}}{\sqrt{\mathbf{w}'\mathbf{Gw}}\sqrt{\mathbf{w}'\mathbf{GP}^{-1}\mathbf{Gw}}} = \sqrt{\frac{\mathbf{w}'\mathbf{GP}^{-1}\mathbf{Gw}}{\mathbf{w}'\mathbf{Gw}}} = \frac{\sigma_I}{\sigma_H}, \text{ thus, } \rho_{HI} = \frac{\sigma_I}{\sigma_H}.$$

3. The variance of the predicted error, $Var(H-I) = \left(1 - \rho_{HI}^2\right)\sigma_H^2$, is minimal. Note that $Var(H-I) = E\left[(H-I)^2\right] = \sigma_I^2 + \sigma_H^2 - 2\sigma_{HI}$, and when $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$, $\sigma_I^2 = \sigma_{HI}$, from where $Var(H-I) = \sigma_H^2 - \sigma_I^2 = \left(1 - \rho_{HI}^2\right)\sigma_H^2$ is minimal because by Eq. (2.8), $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$ minimizes $Var(H-I) = \left(1 - \rho_{HI}^2\right)\sigma_H^2$. Thus, the larger ρ_{HI}, the smaller $E[(H-I)^2]$ and the more similar I and H are. If $\rho_{HI} > 0$, I and H tend to be positively related; if $\rho_{HI} < 0$, they tend to be negatively related; and if $\rho_{HI} = 0$, I and H are independent (Anderson 2003).

4. The total variance of H explained by I is $\sigma_I^2 = \rho_{HI}^2\sigma_H^2$. It is evident that if $\rho_{HI} = 1$, $\sigma_I^2 = \sigma_H^2$, and if $\rho_{HI} = 0$, $\sigma_I^2 = 0$. That is, the variance of H explained by I is proportional to ρ_{HI}, and when ρ_{HI} is close to 1, σ_I^2 is close to σ_H^2, and if ρ_{HI} is close to 0, σ_I^2 is close to 0.

2.5 Particular Cases of the LPSI

2.5.1 The Base LPSI

To derive the LPSI theory, we assumed that the phenotypic ($\mathbf{P}$) and the genotypic ($\mathbf{G}$) covariance matrix, and the vector of economic values ($\mathbf{w}$) are known. However, $\mathbf{P}$, $\mathbf{G}$, and $\mathbf{w}$ are generally unknown and it is necessary to estimate them. There are many methods for estimating $\mathbf{P}$ and $\mathbf{G}$ (Lynch and Walsh 1998) and $\mathbf{w}$ (Cotterill and Jackson 1985; Magnussen 1990). However, when the estimator of $\mathbf{P}(\widehat{\mathbf{P}})$ is not positive definite (all eigenvalues positive) or the estimator of $\mathbf{G}(\widehat{\mathbf{G}})$ is not positive semidefinite (no negative eigenvalues), the estimator of $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$ ($\widehat{\mathbf{b}} = \widehat{\mathbf{P}}^{-1}\widehat{\mathbf{G}}\mathbf{w}$) could be biased. In this case, the base linear phenotypic selection index (BLPSI):

$$I_B = \mathbf{w}'\mathbf{y} \tag{2.15}$$

may be a better predictor of $H = \mathbf{w}'\mathbf{g}$ than the *estimated* LPSI $\widehat{I} = \widehat{\mathbf{b}}'\mathbf{y}$ (Williams 1962a; Lin 1978) if the vector of economic values $\mathbf{w}$ is indeed known. Many authors (Williams 1962b; Harris 1964; Hayes and Hill 1980, 1981) have investigated the influence of parameter estimation errors on LPSI accuracy and concluded that those errors affect the accuracy of $\widehat{I} = \widehat{\mathbf{b}}'\mathbf{y}$ when the accuracy of $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{G}}$ is low. If vector $\mathbf{w}$ values are known, the BLPSI has certain advantages because of its simplicity and its freedom from parameter estimation errors (Lin 1978). Williams (1962a) pointed out that the BLPSI is superior to $\widehat{I} = \widehat{\mathbf{b}}'\mathbf{y}$ unless a large amount of data is available for estimating $\mathbf{P}$ and $\mathbf{G}$.

There are some problems associated with the BLPSI. For example, what is the BLPSI selection response and the BLPSI expected genetic gains per trait when no data are available for estimating $\mathbf{P}$ and $\mathbf{G}$? The BLPSI is a better selection index than the standard LPSI only if the correlation between the BLPSI and the net genetic merit is higher than that between the LPSI and the net genetic merit (Hazel 1943).

However, if estimations of $\mathbf{P}$ and $\mathbf{G}$ are not available, how can the correlation between the base index and the net genetic merit be obtained? Williams (1962b) pointed out that the correlation between the BLPSI and $H = \mathbf{w}'\mathbf{g}$ can be written as

$$\rho_{HI_B} = \sqrt{\frac{\mathbf{w}'\mathbf{G}\mathbf{w}}{\mathbf{w}'\mathbf{P}\mathbf{w}}} \tag{2.16}$$

and indicated that the ratio ρ_{HI_B}/ρ_{HI} can be used to compare LPSI efficiency versus BLPSI efficiency; however, in the latter case, at least the estimates of $\mathbf{P}$ and $\mathbf{G}$, i.e., $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{G}}$, need to be known.

In addition, Eq. (2.15) is only an assumption, not a result, and implies that $\mathbf{P}$ and $\mathbf{G}$ are the same. That is, $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{w} = \mathbf{w}$ only when $\mathbf{P} = \mathbf{G}$, which indicates that the BLPSI is a special case of the LPSI. Thus, to obtain the selection response and the expected genetic gains per trait of the BLPSI, we need some information about $\mathbf{P}$ and $\mathbf{G}$. Assuming that the BLPSI is indeed a particular case of the LPSI, the BLPSI selection response and the BLPSI expected genetic gains per trait could be written as

$$R_B = k_I \sqrt{\mathbf{w}'\mathbf{P}\mathbf{w}}, \tag{2.17}$$

and

$$\mathbf{E}_B = k_I \frac{\mathbf{G}\mathbf{w}}{\sqrt{\mathbf{w}'\mathbf{P}\mathbf{w}}}, \tag{2.18}$$

respectively. The parameters of Eqs. (2.17) and (2.18) were defined earlier.

There are additional implications if $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{w} = \mathbf{w}$. For example, if $\mathbf{P} = \mathbf{G}$, then $\rho_{HI_B} = \sqrt{\frac{\mathbf{w}'\mathbf{G}\mathbf{w}}{\mathbf{w}'\mathbf{P}\mathbf{w}}}$ and BLPSI heritability $h^2_{I_B} = \frac{\mathbf{w}'\mathbf{G}\mathbf{w}}{\mathbf{w}'\mathbf{P}\mathbf{w}}$ are equal to 1. However, in practice, the estimated values of the $\rho_{HI_B}(\widehat{\rho}_{HI_B})$ are usually lower than the estimated values of the $\rho_{HI}(\widehat{\rho}_{HI})$ (Fig. 2.6).

2.5.2 The LPSI for Independent Traits

Suppose that the traits under selection are independent, then $\mathbf{P}$ and $\mathbf{G}$ are diagonal matrices and $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{w}$ is a vector of single-trait heritabilities multiplied by the economic weights, because $\mathbf{P}^{-1}\mathbf{G}$ is the matrix of multi-trait heritabilities (Xu and Muir 1992). Based on this result, Hazel and Lush (1942) and Smith et al. (1981) used trait heritabilities multiplied by the economic weights (or heritabilities only) as coefficients of the LPSI. Thus, when the traits are independent and the economic weights are known, the LPSI can be constructed as
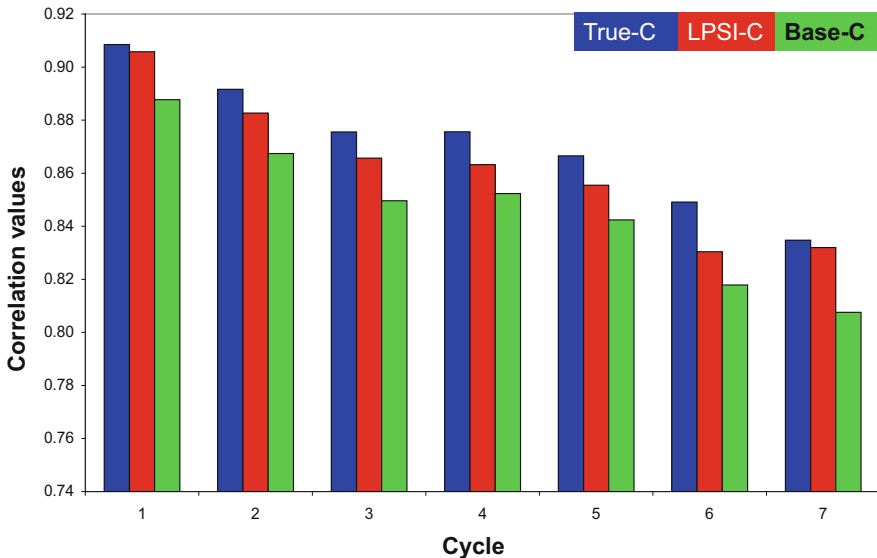
Fig. 2.6 Values of the true correlation between the LPSI and the net genetic merit ($H = \mathbf{w}'\mathbf{g}$) (*True-C*), the estimated correlation between the LPSI and H (*LPSI-C*), and the estimated correlation between the base index and H (*Base-C*) for four traits and 500 genotypes in one environment simulated for seven selection cycles

$$I = \sum_{i=1}^{t} w_i h_i^2 y_i, \tag{2.19}$$

and when the economic weights are unknown, the LPSI can be constructed as

$$I = \sum_{i=1}^{t} h_i^2 y_i. \tag{2.20}$$

The selection response of Eq. (2.19) and (2.20) can be seen in Hazel and Lush (1942).

2.6   Criteria for Comparing LPSI Efficiency

Assuming that the intensity of selection is the same in both indices, we can compare BLPSI ($I_B = \mathbf{w}'\mathbf{y}$) efficiency versus LPSI efficiency to predict the net genetic merit in percentage terms as

$$p = 100(\lambda - 1), \qquad\qquad (2.21)$$

where $\lambda = \frac{\rho_{HI}}{\rho_{HI_B}}$ (Williams [1962b](#); Bulmer [1980](#)). Therefore, when $p = 0$, the efficiency of both indices is the same; when $p > 0$, the efficiency of the LPSI is higher than the base index efficiency, and when $p < 0$, the base index efficiency is higher than LPSI efficiency (Fig. [2.6](#)). Equation ([2.21](#)) is useful for comparing the efficiency of any linear selection index, as we shall see in this book.

2.7   Estimating Matrices G and P

To derive the LPSI theory we assumed that matrices **P** and **G** are known. In practice, we have to estimate them. Matrices **P** and **G** can be estimated by analysis of variance (ANOVA), maximum likelihood or restricted maximum likelihood (REML) (Baker [1986](#); Lynch and Walsh [1998](#); Searle et al. [2006](#); Hallauer et al. [2010](#)). Equation ([2.1](#)) is the simplest model because we only need to estimate two variance components: the genotypic variance (σ_g^2) and the residual variance (σ_e^2), from where the phenotypic variance for trait y is the sum of σ_g^2 and σ_e^2, that is, $\sigma_y^2 = \sigma_g^2 + \sigma_e^2$. However, to construct matrices **P** and **G**, we also need the covariance between any two traits. Thus, if y_i and y_j ($i, j = 1, 2, \cdots, t$) are any two traits, then the covariance between y_i and y_j ($\sigma_{y_{ij}}$) can be written as $\sigma_{y_{ij}} = \sigma_{g_{ij}} + \sigma_{e_{ij}}$, where $\sigma_{g_{ij}}$ and $\sigma_{e_{ij}}$ denote the genotypic and residual covariance respectively of traits y_i and y_j.

Several authors (Baker [1986](#); Lynch and Walsh [1998](#); Hallauer et al. [2010](#)) have described ANOVA methods for estimating matrix **G** using specific design data, for example, half-sib, full-sib, etc., when the sample sizes are well balanced. In the ANOVA method, observed mean squares are equal to their expected values; the expected values are linear functions of the unknown variance components; thus the resulting equations are a set of simultaneous linear equations in the variance components. The expected values of mean squares in the ANOVA method do not need assumptions of normality because the variance component estimators do not depend on normality assumptions (Lynch and Walsh [1998](#); Hallauer et al. [2010](#)).

In cases where the sample sizes are not well balanced, Lynch and Walsh ([1998](#)) and Fry ([2004](#)) proposed using the REML method to estimate matrix **G**. The REML estimation method does not require a specific design or balanced data and can be used to estimate genetic and residual variance and covariance in any arbitrary pedigree of individuals. The REML method is based on projecting the data in a subspace free of fixed effects and maximizing the likelihood function in this subspace, and has the advantage of producing the same results as the ANOVA in balanced designs (Blasco [2001](#)).

In the context of the linear mixed model, Lynch and Walsh ([1998](#)) have given formulas for estimating variances σ_g^2 and σ_e^2 that can be adapted to estimate covariances $\sigma_{g_{ij}}$ and $\sigma_{e_{ij}}$. Suppose that we want to estimate σ_g^2 and σ_e^2 for the qth trait ($q = 1, 2\cdots, t =$ number of traits) in the absence of dominance and epistatic

effects using the model $\mathbf{y}_q = \mathbf{1}\mu_q + \mathbf{Zg}_q + \mathbf{e}_q$, where the vector of averages $\mathbf{y}_q \sim$ NMV $(\mathbf{1}\mu_q, \mathbf{V}_q)$ is $g \times 1$ (g = number of genotypes in the population) and has multivariate normal distribution; $\mathbf{1}$ is a $g \times 1$ vector of ones, μ_q is the mean of the qth trait, $\mathbf{Z}$ is an identity matrix $g \times g$, $\mathbf{g}_q \sim$ NMV$(\mathbf{0}, \mathbf{A}\sigma_{gq}^2)$ is a vector of true breeding values, and $\mathbf{e}_q \sim$ NMV$(\mathbf{0}, \mathbf{I}\sigma_{e_q}^2)$ is a $g \times 1$ vector of residuals, where NMV stands for normal multivariate distribution. Matrix $\mathbf{A}$ denotes the numerical relationship matrix between individuals (Lynch and Walsh 1998; Mrode 2005) and $\mathbf{V}_q = \mathbf{A}\sigma_{gq}^2 + \mathbf{I}\sigma_{e_q}^2$.

The expectation–maximization algorithm allows the REML to be computed for the variance components $\sigma_{g_q}^2$ and $\sigma_{e_q}^2$ by iterating the following equations:

$$\sigma_{g_q}^{2(n+1)} = \sigma_{g_q}^{2(n)} + \frac{\left(\sigma_{g_q}^{2(n)}\right)^2}{g} \left[\mathbf{y}_q'\left(\mathbf{T}^{(n)}\mathbf{A}\mathbf{T}^{(n)}\right)\mathbf{y}_q - tr\left(\mathbf{T}^{(n)}\mathbf{A}\right)\right] \quad (2.22)$$

and

$$\sigma_{e_q}^{2(n+1)} = \sigma_{e_q}^{2(n)} + \frac{\left(\sigma_{e_q}^{2(n)}\right)^2}{g} \left[\mathbf{y}_q'\left(\mathbf{T}^{(n)}\mathbf{T}^{(n)}\right)\mathbf{y}_q - tr\left(\mathbf{T}^{(n)}\right)\right], \quad (2.23)$$

where, after n iterations, $\sigma_{g_q}^{2(n+1)}$ and $\sigma_{e_q}^{2(n+1)}$ are the estimated variance components of $\sigma_{g_q}^2$ and $\sigma_{e_q}^2$ respectively; $tr(.)$ denotes the trace of the matrices within brackets; $\mathbf{T} = \mathbf{V}_q^{-1} - \mathbf{V}_q^{-1}\mathbf{1}\left(\mathbf{1}'\mathbf{V}_q^{-1}\mathbf{1}\right)\mathbf{1}'\mathbf{V}_q^{-1}$ and $\mathbf{V}_q^{-1}$ is the inverse of matrix $\mathbf{V}_q = \mathbf{A}\sigma_{gq}^2 + \mathbf{I}\sigma_{e_q}^2$. In $\mathbf{T}^{(n)}$, $\mathbf{V}_q^{-1(n)}$ is the inverse of matrix $\mathbf{V}_q^{(n)} = \mathbf{A}\sigma_{\gamma q}^{2(n)} + \mathbf{I}\sigma_{e_q}^{2(n)}$.

The additive genetic and residual covariances between the observations of the qth and ith traits, $\mathbf{y}_q$ and $\mathbf{y}_i$ ($\sigma_{g_{q,i}}$ and $\sigma_{e_{q,i}}$, $q, i = 1, 2, \ldots, t$), can be estimated using REML by adapting Eqs. (2.22) and (2.23). Note that the variance of the sum of $\mathbf{y}_q$ and $\mathbf{y}_i$ can be written as $Var(\mathbf{y}_i + \mathbf{y}_q) = \mathbf{V}_i + \mathbf{V}_q + 2\mathbf{C}_{iq}$, where $\mathbf{V}_i = \mathbf{A}\sigma_{gi}^2 + \mathbf{I}\sigma_{e_i}^2$ is the variance of $\mathbf{y}_i$ and $\mathbf{V}_q = \mathbf{A}\sigma_{gq}^2 + \mathbf{I}\sigma_{e_q}^2$ is the variance of $\mathbf{y}_q$; in addition, $2\mathbf{C}_{iq} = 2\mathbf{A}\sigma_{giq} + 2\mathbf{I}\sigma_{eiq} = 2Cov(\mathbf{y}_i, \mathbf{y}_q)$ is the covariance of $\mathbf{y}_q$ and $\mathbf{y}_i$, and σ_{giq} and σ_{eiq} are the additive and residual covariances respectively associated with the covariance of $\mathbf{y}_q$ and $\mathbf{y}_i$. Thus, one way of estimating σ_{giq} and σ_{eiq} is by using the following equation:

$$0.5Var(\mathbf{y}_i + \mathbf{y}_q) - 0.5Var(\mathbf{y}_i) - 0.5Var(\mathbf{y}_q), \quad (2.24)$$

for which Eqs. (2.22) and (2.23) can be used. Equations (2.22) to (2.24) are used to estimate $\mathbf{P}$ and $\mathbf{G}$ in the illustrative examples of this book.

2.8 Numerical Examples

2.8.1 Simulated Data

This data set was simulated by Ceron-Rojas et al. (2015) and can be obtained at http://hdl.handle.net/11529/10199. The data were simulated for eight phenotypic selection cycles (C0 to C7), each with four traits (T_1, T_2, T_3 and T_4), 500 genotypes, and four replicates for each genotype (Fig. 2.7). The LPSI economic weights for T_1,
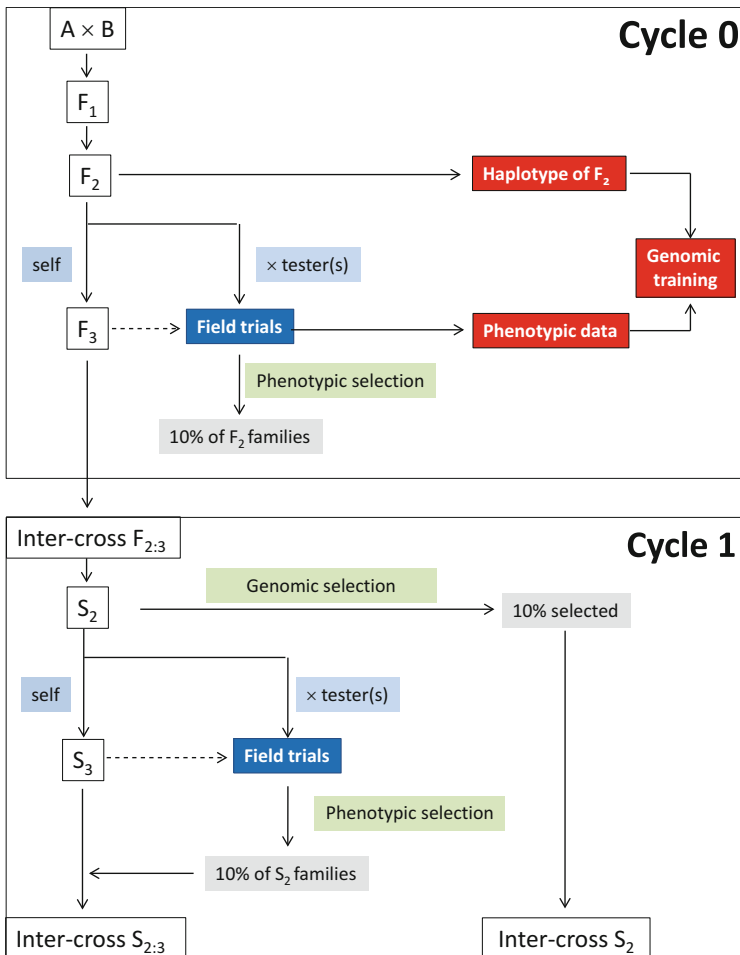
Fig. 2.7 Schematic illustration of the steps followed to generate data sets 1 and 2 for the seven selection cycles using the linear phenotypic selection index and the linear genomic selection index. Dotted lines indicate the process used to simulate the phenotypic data (according to Ceron-Rojas et al. 2015)

T_2, T_3 and T_4 were 1, -1, 1, and 1 respectively. Each of the four traits was affected by a different number of quantitative trait loci (QTLs): 300, 100, 60, and 40, respectively. The common QTLs affecting the traits generated genotypic correlations of -0.5, 0.4, 0.3, -0.3, -0.2, and 0.1 between T_1 and T_2, T_1 and T_3, T_1 and T_4, T_2 and T_3, T_2 and T_4, and T_3 and T_4 respectively. The genotypic value of each plant was generated based on its haplotypes and the QTL effects for each trait.

Simulated data were generated using QU-GENE software (Podlich and Cooper 1998; Wang et al. 2003). A total of 2500 molecular markers were distributed uniformly across 10 chromosomes, whereas 315 QTLs were randomly allocated over the ten chromosomes to simulate one maize (*Zea mays* L.) population. Each QTL and molecular marker was biallelic and the QTL additive values ranged from 0 to 0.5. As QU-GENE uses recombination fraction rather than map distance to calculate the probability of crossover events, recombination between adjacent pairs of markers was set at 0.0906; for two flanking markers, the QTL was either on the first (recombination between the first marker and QTL was equal to 0.0) or the second (recombination between the first marker and QTL was equal to 0.0906) marker; excluding the recombination fraction between 15 random QTLs and their flanking markers, which was set at 0.5, i.e., complete independence (Haldane 1919), to simulate linkage equilibrium between 5% of the QTLs and their flanking markers. In addition, in every case, two adjacent QTLs were in complete linkage. For each trait, the phenotypic value for each of four replications of each plant was obtained from QU-GENE by setting the per-plot heritability of T_1, T_2, T_3, and T_4 at 0.4, 0.6, 0.6, and 0.8 respectively.

2.8.2   Estimated Matrices, LPSI, and Its Parameters

For this example, we used only cycle C1 data and traits T_1, T_2, and T_3. The phenotypic and genotypic estimated covariance matrices for traits T_1, T_2, and T_3 were $\widehat{\mathbf{P}} = \begin{bmatrix} 62.50 & -12.74 & 8.53 \\ -12.74 & 17.52 & -3.38 \\ 8.53 & -3.38 & 12.31 \end{bmatrix}$ and $\widehat{\mathbf{G}} = \begin{bmatrix} 36.21 & -12.93 & 8.35 \\ -12.93 & 13.04 & -3.40 \\ 8.35 & -3.40 & 9.96 \end{bmatrix}$ respectively, whereas the inverse of matrix $\widehat{\mathbf{P}}$ was $\widehat{\mathbf{P}}^{-1} = \begin{bmatrix} 0.01997 & 0.01251 & -0.01040 \\ 0.01251 & 0.06809 & 0.01005 \\ -0.01040 & 0.01005 & 0.09123 \end{bmatrix}$. The estimated heritabilities for T_1, T_2, and T_3 were $\widehat{h}_1^2 = 0.579$, $\widehat{h}_2^2 = 0.744$, and $\widehat{h}_2^2 = 0.809$ respectively.

According to matrices $\widehat{\mathbf{P}}^{-1}$ and $\widehat{\mathbf{G}}$, and because $\mathbf{w}' = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$, the estimated vector of coefficients was $\widehat{\mathbf{b}}' = \mathbf{w}'\widehat{\mathbf{G}}\widehat{\mathbf{P}}^{-1} = \begin{bmatrix} 0.555 & -1.063 & 1.087 \end{bmatrix}$, from which the estimated LPSI can be written as $\widehat{I} = 0.555T_1 - 1.063T_2 + 1.087T_3$. Table 2.2 presents the first 20 genotypes, the means of the three traits (T1, T2 and T3) and the first 20 estimated unranked LPSI values of the 500 simulated genotypes for cycle C1.

Table 2.2 Number of genotypes, means of the trait (T1, T2 and T3) values, and unranked values of the LPSI for part of a simulated data set

Number of genotypes	Means of the trait values			Unranked LPSI values
	T1	T2	T3	
1	164.46	39.63	34.66	86.81
2	144.39	50.77	34.65	63.82
3	157.48	48.04	37.9	77.52
4	167.30	47.98	30.49	74.97
5	164.11	49.89	32.03	72.85
6	166.26	40.44	29.93	81.81
7	154.59	52.22	30.31	63.22
8	160.00	42.91	31.23	77.12
9	158.51	46.32	34.52	76.25
10	163.63	45.43	35.73	81.35
11	156.16	46.75	35.58	75.62
12	171.38	41.17	35.13	89.52
13	153.17	54.18	36.23	66.79
14	149.89	52.33	31.13	61.39
15	159.63	49.01	31.72	70.96
16	160.70	42.51	32.99	79.85
17	157.07	45.49	28.4	69.68
18	167.50	41.69	36.73	88.55
19	159.17	50.6	36.25	73.93
20	161.80	46.58	37.33	80.84

According to the means of the three traits, the first estimated LPSI value was obtained as

$$\widehat{I}_1 = 0.555(164.46) - 1.063(39.63) + 1.087(34.66) = 86.81;$$

the second estimated LPSI value was obtained as

$$\widehat{I}_2 = 0.555(144.39) - 1.063(144.39) + 1.087(34.65) = 63.82, \text{etc.};$$

and the 20th estimated LPSI value was obtained as

$$\widehat{I}_{20} = 0.555(161.80) - 1.063(46.58) + 1.087(37.33) = 80.84.$$

This estimation procedure is valid for any number of genotypes. Table 2.3 presents the 20 genotypes ranked by the estimated LPSI values. Note that if we use 20% selection intensity for Table 2.2 data, we should select genotypes 12, 18, 1, 6, and 10, because their estimated LPSI values are higher than the remaining LPSI values for that set of genotypes. Using the idea described in Fig. 2.4, genotypes 12, 18, 1, 6, and 10 should be in the red zone, whereas the rest of the genotypes are in the white zone and should be culled. Here, the proportion selected is $q = 0.2$ and

Table 2.3 Number of genotypes, means of the trait (T1, T2 and T3) values and ranked values of the LPSI for part of a simulated data set

Number of genotypes	Means of the trait values			Ranked
	T1	T2	T3	LPSI values
12	171.38	41.17	35.13	89.52
18	167.50	41.69	36.73	88.55
1	164.46	39.63	34.66	86.81
6	166.26	40.44	29.93	81.81
10	163.63	45.43	35.73	81.35
20	161.80	46.58	37.33	80.84
16	160.70	42.51	32.99	79.85
3	157.48	48.04	37.9	77.52
8	160.00	42.91	31.23	77.12
9	158.51	46.32	34.52	76.25
11	156.16	46.75	35.58	75.62
4	167.30	47.98	30.49	74.97
19	159.17	50.6	36.25	73.93
5	164.11	49.89	32.03	72.85
15	159.63	49.01	31.72	70.96
17	157.07	45.49	28.4	69.68
13	153.17	54.18	36.23	66.79
2	144.39	50.77	34.65	63.82
7	154.59	52.22	30.31	63.22
14	149.89	52.33	31.13	61.39

$z = \frac{\exp\{-0.5u'^2\}}{\sqrt{2\pi}} = 0.31$, where $u' = \frac{81.35 - 75.64}{8.11} = 0.704$, 81.35 is the estimated LPSI value or the genotype number 10, 75.64 is the mean of the 20 LPSI values, and 8.11 is the standard deviation of the estimated LPSI values of the 20 genotypes presented in Tables 2.2 and 2.3.

Table 2.4 presents 25 genotypes and the means of the three traits obtained from the 500 simulated genotypes for cycle C1 and ranked by the estimated LPSI values. In this case, we used 5% selection intensity ($k_I = 2.063$). Also, the last four rows in Table 2.4 give:

1. The means of traits T_1, T_2, and T_3 (175.46, 39.26, and 38.83 respectively) of the selected individuals and the mean of the selected LPSI values (97.84).
2. The means of the three traits in the base population (161.88, 45.19, and 34.39) and the mean of the LPSI values in the base population (79.18)
3. The selection differentials for the three traits (13.58, −5.92, and 4.44) and the selection differential for the LPSI (18.66)
4. The LPSI expected genetic gain per trait (9.51, −5.48, and 4.22) and the LPSI selection response (19.21).

The variance of the estimated selection index for the 500 genotypes was $\widehat{V}(\widehat{I}) = \widehat{\mathbf{b}}'\widehat{\mathbf{P}}\widehat{\mathbf{b}} = 86.72$, from which the standard deviation of $\widehat{I}$ was 9.312. The

Table 2.4 Number of selected genotypes, selected means of the trait (T1, T2 and T3) values and ranked selected values of the LPSI from one simulated set of 500 genotypes with four repetitions

Number of genotypes	Means of the trait values			Ranked
	T1	T2	T3	LPSI values
353	189.68	38.16	36.13	103.97
370	178.27	34.38	37.79	103.45
480	174.84	42.72	45.12	100.66
300	177.38	39.15	40.34	100.65
273	181.18	35.94	35.14	100.52
275	167.94	36.82	42.2	99.92
148	173.37	37.07	39.62	99.86
137	185.48	46.48	42.55	99.77
351	173.79	38.38	40.52	99.68
236	182.85	37.88	34.96	99.2
217	175.13	38.48	39.16	98.84
356	171.09	39.6	41.98	98.47
167	175.39	38.73	37.73	97.17
230	169.73	37.1	38.69	96.8
243	171.9	41.53	41.45	96.29
55	170.02	36.92	37.76	96.15
68	172.56	37.18	36.7	96.13
36	175.8	38.86	36.34	95.75
164	173.61	38.37	36.42	95.14
140	170.53	42.52	41.97	95.05
146	177.4	39.64	35.5	94.89
432	174.01	40.73	38.26	94.84
378	176.62	42.69	38.47	94.44
288	172.14	39.31	37.26	94.23
386	175.77	42.89	38.81	94.13
Mean of selected individuals	175.46	39.26	38.83	97.84
Mean of all individuals	161.88	45.19	34.39	79.18
Selection differential	13.58	−5.92	4.44	18.66
Expected genetic gain for 5%	9.51	−5.48	4.22	19.21

The selection intensity was 5%

estimated standardized selection differentials for the LPSI can be obtained from Table A in Falconer and Mackay (1996), where, for 5% selection intensity, $k_I = 2.063$. This means that the estimated LPSI selection response was $\widehat{R} = 2.063(9.312) = 19.21$, whereas the expected genetic gain per trait, or multi-trait selection response, was $\widehat{\mathbf{E}}' = 2.063 \left[\dfrac{\mathbf{b}'\widehat{\mathbf{G}}}{9.312} \right] = [\, 9.51 \quad -5.48 \quad 4.22 \,]$.

2.8.3   LPSI Efficiency Versus Base Index Efficiency

The estimated correlation between the LPSI and the net genetic merit was $\widehat{\rho}_{HI} = \dfrac{\widehat{\sigma}_I}{\widehat{\sigma}_H} = 0.894$, whereas the estimated correlation between the base index and the net genetic merit was $\widehat{\rho}_{HI_B} = 0.875$, thus $\widehat{\lambda} = \dfrac{\widehat{\rho}_{HI}}{\widehat{\rho}_{HI_B}} = 1.0217$ and, by Eq. (2.21), $\widehat{p} = 100(\widehat{\lambda} - 1) = 2.171$. This means that LPSI efficiency was only 2.2% higher than the base index efficiency for this data set.

   Using the same data set described in Sect. 2.8.1 of this chapter, we conducted seven selection cycles (C1 to C7) for the four traits (T_1, T_2, T_3, and T_4) using the LPSI and the BLPSI. These results are presented in Table 2.5. To compare the LPSI efficiency versus BLPSI efficiency, we obtained the true selection response of the simulated data (second column in Table 2.5) and we estimated the LPSI and BLPSI selection response for each selection cycle (third column in Table 2.5); in addition, we estimated the LPSI and BLPSI expected genetic gain per trait for each selection cycle (columns 4 to 7 in Table 2.5). The first part of Table 2.5 shows the true selection response and the estimated values of the LPSI selection response and expected genetic gain per trait. In a similar manner, the second part of Table 2.5 shows the true selection response, the estimated values of the BLPSI selection

Table 2.5 The LPSI and BLPSI responses (true and estimated) and estimated expected genetic gain per trait for seven simulated selection cycles

Cycle	Selection response		Estimated expected genetic gain per trait			
	True	Estimated	T1	T2	T3	T4
LPSI						
1	17.84	17.81	7.90	−4.67	3.33	1.92
2	15.66	15.69	7.06	−3.59	3.17	1.86
3	14.44	14.22	6.67	−3.21	2.82	1.52
4	14.29	14.34	7.53	−3.45	2.07	1.29
5	13.86	13.64	7.14	−2.66	2.51	1.33
6	12.47	12.04	6.23	−2.62	1.98	1.21
7	12.44	11.61	5.38	−2.55	2.47	1.22
Average	14.43	14.19	6.85	−3.25	2.62	1.48
BLPSI						
1	17.84	22.15	8.38	−4.40	3.04	1.64
2	15.66	20.49	7.74	−3.33	2.82	1.53
3	14.44	19.33	7.29	−3.00	2.44	1.22
4	14.29	19.49	8.05	−3.17	1.89	1.05
5	13.86	18.93	7.64	−2.53	2.19	1.07
6	12.47	17.72	6.81	−2.40	1.72	0.93
7	12.44	17.28	5.89	−2.35	2.11	0.93
Average	14.43	19.34	7.40	−3.02	2.32	1.19

The selection intensity was 10% ($k_I = 1.755$)

response, and the expected genetic gain per trait. The average value of the true selection response was equal to 14.43, whereas the average values of the estimated LPSI and BLPSI selection response were 14.19 and 19.34 respectively. Note that $14.43–14.19 = 0.24$, but $19.34–14.43 = 4.91$. According to this result, the BLPSI over-estimated the true selection response of the simulated data by 34.7%. Thus, based on the Table 2.5 results and those presented in Fig. 2.6, we can conclude that the LPSI was more efficient than the BLPSI for this data set.

Finally, additional results can be seen in Chap. 10, where the LPSI was simulated for many selection cycles. Chapter 11 describes RIndSel: a program that uses R and the selection index theory to make selection.

2.9 The LPSI and Its Relationship with the Quadratic Phenotypic Selection Index

In the nonlinear selection index theory, the net genetic merit and the index are both nonlinear. There are many types of nonlinear indices; Goddard (1983) and Weller et al. (1996) have reviewed the general theory of nonlinear selection indices. In this chapter, we describe only the simplest of them: the quadratic index developed mainly by Wilton et al. (1968), Wilton (1968), and Wilton and Van Vleck (1969), which is related to the LPSI.

2.9.1 The Quadratic Nonlinear Net Genetic Merit

The most common form of writing the quadratic net genetic merit is

$$H_q = \alpha + \mathbf{w}'(\boldsymbol{\mu} + \mathbf{g}) + (\boldsymbol{\mu} + \mathbf{g})'\mathbf{A}(\boldsymbol{\mu} + \mathbf{g}), \tag{2.25}$$

where α is a constant, $\mathbf{g}$ is the vector of breeding values, which has normal distribution with zero mean and covariance matrix $\mathbf{G}$, $\boldsymbol{\mu}$ is the vector of population means, and $\mathbf{w}$ is a vector of economic weights. In addition, matrix $\mathbf{A}$ can be written as $\mathbf{A} = \begin{bmatrix} w_1 & 0.5w_{12} & \cdots & 0.5w_{1t} \\ 0.5w_{12} & w_2 & \cdots & 0.5w_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ 0.5w_{1t} & 0.5w_{2t} & \ldots & w_t \end{bmatrix}$, where the diagonal ith values w_i ($i = 1, 2$, ..., t) is the relative economic weight of the genetic value of the squared trait i and w_{ij} ($i, j = 1, 2, \ldots, t$) is the economic weight of the cross products between the genetic values of traits i and j. The main difference between the linear net genetic merit (Eq. 2.2) and the net quadratic merit (Eq. 2.25) is that the latter depends on $\boldsymbol{\mu}$ and $(\boldsymbol{\mu} + \mathbf{g})'\mathbf{A}(\boldsymbol{\mu} + \mathbf{g})$.

2.9.2 The Quadratic Index

The quadratic phenotypic selection index is

$$I_q = \beta + \mathbf{b}'\mathbf{y} + \mathbf{y}'\mathbf{B}\mathbf{y} \tag{2.26}$$

where β is a constant, $\mathbf{y}$ is the vector of phenotypic values that has multivariate normal distribution with zero mean and covariance matrix $\mathbf{P}$, $\mathbf{b}' = \begin{bmatrix} b_1 & b_2 & \cdots & b_t \end{bmatrix}$ is a vector of coefficients, and $\mathbf{B} = \begin{bmatrix} b_1 & 0.5b_{12} & \cdots & 0.5b_{1t} \\ 0.5b_{12} & b_2 & \cdots & 0.5b_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ 0.5b_{1t} & 0.5b_{2t} & \ldots & b_t \end{bmatrix}$. In matrix $\mathbf{B}$, the diagonal ith values b_i $(i = 1,2, \ldots, t\,)$ is the index weight for the square of the phenotypic i and b_{ij} $(i,j = 1,2, \ldots, t\,)$ is the index weight for the cross products between the phenotype of the traits i and j.

2.9.3 The Vector and the Matrix of Coefficients of the Quadratic Index

As we saw in Sect. 2.3.2 of this chapter, to obtain the vector ($\mathbf{b}$) and the matrix ($\mathbf{B}$) of coefficients of the quadratic index that maximized the selection response, we can minimize the expectation of the square difference between the quadratic index (I_q) and the quadratic net genetic merit (H_q): $\Phi = E\{[I_q - E(I_q)] - [H_q - E(H_q)]\}^2$, or we can maximize the correlation between I_q and H_q, i.e., $\rho_{H_q I_q} = \dfrac{Cov(H_q, I_q)}{\sqrt{Var(I_q)}\sqrt{Var(H_q)}}$, where $Cov(H_q, I_q)$ is the covariance between I_q and H_q, $\sqrt{Var(I_q)}$ is the standard deviation of the variance of I_q, and $\sqrt{Var(H_q)}$ is the standard deviation of the variance of H_q. In this context, it is easier to maximize $\rho_{H_q I_q}$ than to minimize Φ. Vandepitte (1972) minimized Φ, but in this section we shall maximize $\rho_{H_q I_q}$.

Suppose that $\boldsymbol{\mu} = \mathbf{0}$, since α and β are constants that do not affect $\rho_{H_q I_q}$, we can write I_q and H_q as $I_q = \mathbf{b}'\mathbf{y} + \mathbf{y}'\mathbf{B}\mathbf{y}$ and $H_q = \mathbf{w}'\mathbf{g} + \mathbf{g}'\mathbf{A}\mathbf{g}$. Thus, under the assumption that $\mathbf{y}$ and $\mathbf{g}$ have multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{P}$ and $\mathbf{G}$, respectively, $E(I_q) = tr(\mathbf{B}\mathbf{P})$ and $E(H_q) = tr(\mathbf{A}\mathbf{G})$ are the expectations of I_q and H_q, whereas $Var(I_q) = \mathbf{b}'\mathbf{P}\mathbf{b} + 2tr[(\mathbf{B}\mathbf{P})^2]$ and $Var(H_q) = \mathbf{w}'\mathbf{G}\mathbf{w} + 2tr[(\mathbf{A}\mathbf{G})^2]$ are the variances of I_q and H_q, respectively. The covariance between I_q and H_q is $Cov(H_q, I_q) = \mathbf{w}'\mathbf{G}\mathbf{b} + 2tr(\mathbf{B}\mathbf{G}\mathbf{A}\mathbf{G})$ (Vandepitte 1972), where $tr(\circ)$ denotes the trace function of matrices.

According to the foregoing results, we can maximize the natural logarithm of $\rho_{H_q I_q}$ [$\ln\left(\rho_{H_q I_q}\right)$] with respect to vector $\mathbf{b}$ and matrix $\mathbf{B}$ assuming that $\mathbf{w}, \mathbf{A}, \mathbf{P}$, and $\mathbf{G}$ are known. Hence, except for two proportional constants that do not affect the

maximum value of $\rho_{H_q I_q}$ because this is invariant to the scale change, the results of the derivatives of $\ln\left(\rho_{H_q I_q}\right)$ with respect to $\mathbf{b}$ and $\mathbf{B}$ are

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw} \text{ and } \mathbf{B} = \mathbf{P}^{-1}\mathbf{GAGP}^{-1}, \tag{2.27}$$

respectively. In this case, $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$ is the same as the LPSI vector of coefficients (see Eq. 2.8 for details); however, when $\boldsymbol{\mu} \neq \mathbf{0}$, $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}(\mathbf{w} + 2\mathbf{A}\boldsymbol{\mu}) = \mathbf{P}^{-1}\mathbf{Gw} + 2\mathbf{P}^{-1}\mathbf{GA}\boldsymbol{\mu}$. In the latter case, $\mathbf{b}$ has the additional term $2\mathbf{P}^{-1}\mathbf{GA}\boldsymbol{\mu}$, which is null when $\boldsymbol{\mu} = \mathbf{0}$ or $\mathbf{A} = \mathbf{0}$. Hence, when $\boldsymbol{\mu} \neq \mathbf{0}$ the quadratic index vector $\mathbf{b}$ shall have two components: $\mathbf{P}^{-1}\mathbf{Gw}$, which is the LPSI vector of coefficients, and $2\mathbf{P}^{-1}\mathbf{GA}\boldsymbol{\mu}$, which is a function of the current population mean $\boldsymbol{\mu}$ multiplied by matrix $\mathbf{A}$. Therefore, when $\boldsymbol{\mu} \neq \mathbf{0}$ and $\mathbf{A} \neq \mathbf{0}$, the quadratic index vector $\mathbf{b}$ will change when the $\boldsymbol{\mu}$ values change. However, $\boldsymbol{\mu}$ does not affect matrix $\mathbf{B}$.

2.9.4   The Accuracy and Maximized Selection Response of the Quadratic Index

According to Eq. (2.27) results, $Var(I_q) = Cov(H_q, I_q) = \mathbf{b}'\mathbf{Pb} + 2tr[(\mathbf{BP})^2]$, which means that the quadratic index accuracy and the maximized selection response can be written as:

$$\rho_{H_q I_q} = \frac{\sqrt{\mathbf{w}'\mathbf{GP}^{-1}\mathbf{Gw} + 2tr\left[\left(\mathbf{P}^{-1}\mathbf{GAG}\right)^2\right]}}{\sqrt{\mathbf{w}'\mathbf{Gw} + 2tr\left[(\mathbf{AG})^2\right]}} \tag{2.28}$$

and

$$R_q = k\sqrt{\mathbf{w}'\mathbf{GP}^{-1}\mathbf{Gw} + 2tr\left[\left(\mathbf{P}^{-1}\mathbf{GAG}\right)^2\right]}, \tag{2.29}$$

respectively, where k is the selection intensity of the quadratic index. Equations (2.27) to (2.29) indicate that the LPSI and the quadratic index are related, and the only difference between them is the quadratic terms. Wilton et al. (1968) wrote Eq. (2.29) as: $R_q = k\sqrt{\mathbf{b}'\mathbf{Pb}} + k2tr\left[(\mathbf{BP})^2\right]$.

References

Akbar MK, Lin CY, Gyles NR, Gavora JS, Brown CJ (1984) Some aspects of selection indices with constraints. Poult Sci 63:1899–1905

Anderson TW (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, Hoboken, NJ

Baker RJ (1986) Selection indices in plant breeding. CRC Press, Boca Raton, FL

Blasco A (2001) The Bayesian controversy in animal breeding. J Anim Sci 79:2023–2046

Brascamp EW (1984) Selection indices with constraints. Anim Breed Abstr 52(9):645–654

Bulmer MG (1980) The mathematical theory of quantitative genetics. Lectures in biomathematics. University of Oxford, Clarendon Press, Oxford

Cartuche L, Pascual M, Gómez EA, Blasco A (2014) Economic weights in rabbit meat production. World Rabbit Sci 22:165–177

Cerón-Rojas JJ, Crossa J, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A (2006) A selection index method based on eigen analysis. Crop Sci 46:1711–1721

Ceron-Rojas JJ, Crossa J, Arief VN, Basford K, Rutkoski J, Jarquín D, Alvarado G, Beyene Y, Semagn K, DeLacy I (2015) A genomic selection index applied to simulated and real data. Genes/Genomes/Genetics 5:2155–2164

Cochran WG (1951) Improvement by means of selection. In: Neyman J (ed) Proceedings of the second Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley, CA, pp 449–470

Costa MM, Di Mauro AO, Unêda-Trevisoli SH, Castro Arriel NH, Bárbaro IM, Dias da Silveira G, Silva Muniz FR (2008) Analysis of direct and indirect selection and indices in soybean segregating populations. Crop Breed Appl Biotechnol 8:47–55

Cotterill PP, Jackson N (1985) On index selection. I. Methods of determining economic weight. Silvae Genet 34:56–63

Cunningham EP, Tauebert H (2009) Measuring the effect of change in selection indices. J Dairy Sci 92:6192–6196

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longman, New York

Fry JD (2004) Estimation of genetic variances and covariance by restricted maximum likelihood using PROC MIXED. In: Saxto AM (ed) Genetics analysis of complex traits using SAS. SAS Institute, Cary, NC, pp 11–34

Goddard ME (1983) Selection indices for non-linear profit functions. Theor Appl Genet 64:339–344

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Hallauer AR, Carena MJ, Miranda Filho JB (2010) Quantitative genetics in Maize breeding. Springer, New York

Harris DL (1964) Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20(1):46–72

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

Hayes JF, Hill WG (1980) A reparameterization of a genetic selection index to locate its sampling properties. Biometrics 36(2):237–248

Hayes JF, Hill WG (1981) Modification of estimates of parameters in the construction of genetic selection indices ('Bending'). Biometrics 37(3):483–493

Hazel LN (1943) The genetic basis for constructing selection indexes. Genetics 8:476–490

Hazel LN, Lush JL (1942) The efficiency of three methods of selection. J Hered 33:393–399

Henderson CR (1963) Selection index and expected genetic advance. In: Statistical genetics and plant breeding. National Academy of Science-National Research Council, Washington, DC, pp 141–163

Kabakoff RI (2011) R in action: data analysis and graphics with R. Manning Publications Co., Shelter Island, NY

Kempthorne O, Nordskog AW (1959) Restricted selection indices. Biometrics 15:10–19

Lin CY (1978) Index selection for genetic improvement of quantitative characters. Theor Appl Genet 52:49–56

Lin CY, Allaire FR (1977) Heritability of a linear combination of traits. Theor Appl Genet 51:1–3

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA

MacNeil MD, Nugent RA, Snelling WM (1997) Breeding for profit: an Introduction to selection index concepts. Range Beef Cow Symposium. Paper 142

Magnussen S (1990) Selection index: economic weights for maximum simultaneous genetic gains. Theor Appl Genet 79:289–293

Mather K, Jinks JL (1971) Biometrical genetics: the study of continuous variation. Chapman and Hall, London

Morrissey MB, Kruuk LEB, Wilson AJ (2010) The danger of applying the breeder's equation in observational studies of natural populations. J Evol Biol 23:2277–2288

Mrode RA (2005) Linear models for the prediction of animal breeding values, 2nd edn. CABI Publishing, Cambridge, MA

Nordskog AW (1978) Some statistical properties of an index of multiple traits. Theor Appl Genet 52:91–94

Podlich DW, Cooper M (1998) QU-GENE: a simulation platform for quantitative analysis of genetic models. Bioinformatics 14:632–653

Rao CR (2002) Linear statistical inference and its applications, 2nd edn. Wiley, New York

Searle S, Casella G, McCulloch CE (2006) Variance components. Wiley, Hoboken, NJ

Smith HF (1936) A discriminant function for plant selection. In: Papers on quantitative genetics and related topics. Department of Genetics, North Carolina State College, Raleigh, NC, pp 466–476

Smith O, Hallauer AR, Russell WA (1981) Use of index selection in recurrent selection programs in maize. Euphytica 30:611–618

Tomar SS (1983) Restricted selection index in animal system: a review. Agric Rev 4(2):109–118

Vandepitte WM (1972) Factors affecting the accuracy of selection indexes for the genetic improvement of pigs. Retrospective Theses and Dissertations. Paper 6130

Wang J, van Ginkel M, Podlich DW, Ye G, Trethowan R, Pfeiffer W, DeLacy IH, Cooper M, Rajaram S (2003) Comparison of two breeding strategies by computer simulation. Crop Sci 43:1764–1773

Weller JI, Pasternak H, Groen AF (1996) Selection indices for non-linear breeding objectives, selection for optima. In: Proceedings international workshop on genetic improvement of functional traits in Cattle, Gembloux, January 1996. INTERBULL bulletin no. 12, pp 206–214

Williams JS (1962a) The evaluation of a selection index. Biometrics 18:375–393

Williams JS (1962b) Some statistical properties of a genetic selection. Biometrika 49(3):325–337

Wilton JW, Evans DA, Van Vleck LD (1968) Selection indices for quadratic models of total merit. Biometrics 24:937–949

Wilton JW (1968) Selection of dairy cows for economic merit. Faculty Papers and Publications in Animal Science, Paper 420

Wilton JW, Van Vleck LD (1969) Sire evaluation for economic merit. Faculty Papers and Publications in Animal Science, Paper 425

Xu S, Muir WM (1992) Selection index updating. Theor Appl Genet 83:451–458