Chapter 4

Deterministic Models for Estimated Breeding Values

The previous chapter established the main factors that affect response to selection, i.e. intensity of selection ($i$), accuracy of selection ($r$), genetic standard deviation ($\sigma_g$), and generation interval ($L$). The objective of this chapter is to develop methods to model and evaluate accuracy of selection, and to evaluate the main factors that determine this parameter. The latter will help us with the design of breeding programs.

Accuracy of selection is defined as the correlation between the criterion on which selection is based ($I$) and the objective of selection. For the moment, we will consider the breeding value of a single trait to be the selection objective but this will be extended to more complicated economic selection objectives in Chapter 6.

The previous chapter showed that when selection is on the individual’s own phenotype, the accuracy of selection is equal to the correlation between phenotype and breeding value, which is equal to the square root of heritability ($h$). In practical animal breeding, selection is often not solely on own phenotype but on estimates of breeding values (EBV) that are derived from records on the animal itself as well as its relatives using Best Linear Unbiased Prediction (BLUP) for an animal model (Lynch and Walsh, 1998). An important property of EBV derived from an animal model is that all records that are available on the individual and its relatives are optimally used, while simultaneously adjusting for systematic environmental effects (e.g. herd-year-season), such that the accuracy of the EBV is maximized. Given the equation for predicting genetic superiority of selected animals, i.e. $S = ir\sigma_g$, it is clear that maximizing accuracy is crucial to maximizing genetic gain.

<table>
<thead>
<tr>
<th>Impact of Accuracy on Genetic Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta G/\text{year} = \frac{\sum S}{\sum L}$</td>
</tr>
<tr>
<td>$S = i r \sigma_g$</td>
</tr>
</tbody>
</table>

- $r$ = accuracy of selection
- $= \text{corr}(\text{selection criterion, true BV})$
- $= \text{accuracy of EBV for single trait selection}$
- $\cdot$ maximize to increase gain
- $\cdot$ need to be able to model accuracy to be able to predict genetic gain and to determine the value of information

<table>
<thead>
<tr>
<th>Process of Genetic Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data recording</strong></td>
</tr>
<tr>
<td>- Phenotypic records</td>
</tr>
<tr>
<td>- Pedigree records</td>
</tr>
<tr>
<td><strong>Central data base</strong></td>
</tr>
<tr>
<td><strong>Statistical Analysis</strong></td>
</tr>
<tr>
<td>EBV + accuracy</td>
</tr>
<tr>
<td><strong>Use for selection</strong></td>
</tr>
</tbody>
</table>

Stochastic simulation models of breeding programs can directly incorporate genetic evaluations based on animal models because the data that provide the input for such models are individually
simulated. This is not possible for deterministic models. Thus, when developing deterministic models for genetic improvement, other methods to model selection and accuracy of EBV from BLUP animal models must be used. In addition to allowing deterministic modeling of selection on EBV, these methods are also required to develop a basic understanding of factors that affect accuracy of selection, which are important for the design of breeding programs, including the contribution that different types of records make to accuracy of EBV.

In our development of methods to model accuracy of EBV, we will slowly build our methodology up using the following steps:

1. EBV from own records – simple regression
2. EBV from records on a single type of relatives – simple regression
3. EBV from multiple sources of information – multiple regression – selection index theory
4. EBV from BLUP animal models (module B)

As noted above, the common theme through these methods is the use of linear regression for the prediction of EBV from phenotypic records.

Before going into these developments, we will first describe some general properties of EBV. These properties hold regardless which of the methods listed above is used to estimate the EBV, provided the model used for evaluation is correct and systematic environmental factors are properly accounted for.

### 4.1 Some general properties of EBV

As indicated above, all methods for prediction of breeding values are based on the principles of linear regression: regression of breeding values on phenotypic records. As a result, properties of linear regression can be used to derive general properties of EBV.

One important property of EBV is unbiasedness. This means that the expected magnitude of the true breeding value of an animal is equal to its estimated breeding value:

$$ E(g_i | \hat{g}_i) = \hat{g}_i $$

This implies that selection on \( \hat{g} \) will maximize the expected value of \( g \) for the group of selected individuals. A related property is that the regression of true on estimated breeding values is equal to 1:

$$ b_{g,\hat{g}} = 1 $$

Given unbiasedness, the accuracy of EBV can be derived as the correlation between true and estimated BV as:

$$ r = r_{g,\hat{g}} = b_{g,\hat{g}} \cdot \frac{\sigma_{\hat{g}}}{\sigma_g} = \frac{\sigma_{\hat{g}}}{\sigma_g} $$  \hspace{1cm} (4.1) $$

and the covariance between true and estimated BV as:

$$ \sigma_{g,\hat{g}} = r_{g,\hat{g}} \cdot \sigma_g \cdot \sigma_{\hat{g}} = \sigma_{\hat{g}}^2 $$  \hspace{1cm} (4.2) $$
The variance of EBV is then equal to:  \[ \sigma_{g}^2 = r^2 \sigma_{s}^2 \] (4.3)

Thus, the variance of EBV is equal to the square of accuracy (also referred to as ‘reliability’) multiplied by genetic variance. This shows the importance of accuracy: the larger the accuracy, the larger the variance and spread of EBV of animals in the population, the better we will able to distinguish between genetically superior and average or inferior animals, and the greater the genetic superiority of selected animals will be. This is illustrated in Figure 4.1.

Like any prediction, EBV also have a prediction error, which is the deviation of true BV from the EBV:  \[ \varepsilon_i = g_i - \hat{g}_i \]

The variance of prediction errors (prediction error variance, PEV) can be derived as:

\[
\sigma^2 = \sigma_{g}^2 + \sigma_{e}^2 - 2 \sigma_{g,e} = \sigma_{g}^2 + \sigma_{e}^2 - 2 \sigma_{g,e} = \sigma_{g}^2 - \sigma_{e}^2 = \sigma_{g}^2 - r^2 \sigma_{s}^2 \\
= (1-r^2) \sigma_{s}^2
\] (4.4)

Note that  \[ \sigma_{g}^2 = \sigma_{g}^2 + \sigma_{e}^2 \]

Thus, additive genetic variance is partitioned into variance that is explained by the EBV and unexplained error variance. The higher the accuracy is, the greater the proportion of genetic variance that is explained by the EBV. Also note that the covariance between EBV and prediction errors is equal to zero:  \[ \sigma_{g,e} = \sigma_{g,e} = \sigma_{g}^2 - \sigma_{g}^2 = \sigma_{g}^2 - \sigma_{g}^2 = 0 \]

This makes sense because a non-zero covariance would imply that the prediction error contains some information that can be used to improve the EBV.

Given an animal’s EBV and assuming normality, its true BV is expected to follow a Normal distribution with mean equal to the EBV and variance equal to \((1-r^2) \sigma_{g}^2\):

\[ g_i | \hat{g}_i \sim N(\hat{g}_i, (1-r^2) \sigma_{g}^2) \] (4.5)
This distribution is illustrated in Figure 4.2.

Prediction errors are expected to follow a Normal distribution with mean zero:

$$
\varepsilon_i \sim N(0,(1-r^2)\sigma_g^2)
$$

(4.6)

### 4.2 EBV from own records

In the derivations below, we will assume that phenotypic records, $x_i$, are adjusted for systematic environmental effects and deviated from the mean.

#### 4.2.1 Phenotypic Selection

The simplest form of selection is based on EBV derived from a single record of the phenotype of the individual itself. In this case, the EBV can be derived from regression of BV on phenotype as:

$$
\hat{g}_i = b_{g,x}x_i = b_{g,x} (\text{phenotype of individual})
$$

(4.7)

The regression coefficient can be derived as:

$$
b_{g,x} = \frac{\sigma^2_{g|x}}{\sigma^2_p} = \frac{\sigma^2_{g|x} + \sigma^2_{e|x}}{\sigma^2_p} = \frac{\sigma^2_p}{\sigma^2_p} = h^2
$$

(4.8)

Thus the prediction of an individual's additive genetic value, expressed as a deviation from the population mean, is given by

$$
\hat{g}_i = h^2 x_i
$$

(4.9)

where $x_i$ is the phenotype of individual $i$ expressed as a deviation from the population mean.

The accuracy of selection is:

$$
r = r_{g,g} = \frac{\sigma^2_{g|x} h^2 x_i}{\sigma^2_g h^2 x_i} = h^2 \frac{\sigma^2_g}{h \sigma^2_g} = h
$$

(4.10)
As an example, growth rate in pigs and cattle often has a heritability of around 0.5. Thus with phenotypic selection for growth rate, the EBV of individual $i$ is $\hat{g}_i = 0.5 x_i$ and the accuracy of evaluation is $r = \sqrt{0.5} = 0.707$.

Alternatively, if we were selecting on a single record for milk yield in cows with a heritability of $0.25$, our EBV would be $\hat{g}_i = 0.25 x_i$ and accuracy would be $r = 0.5$

### 4.2.2 Selection on the Mean of Two or more Phenotypic Records on a Single Trait

#### Definition of Repeatability

We can increase accuracy of selection by increasing the number of records collected on each individual. This can be done for traits that are expressed several times during the lifetime of an animal. For example, having two lactation records on a cow should give more information than having only one lactation. For traits with repeated observations, such as milk production, the environmental and/or non-additive genetic component of the phenotype can then be separated in a permanent component that affects the animal for its lifetime and a temporary component, which changes over time. Thus the phenotype for record $j$ on animal $i$ can be written as:

$$x_{ij} = g_i + pe_i + te_{ij}$$

where $pe_i$ is a permanent environment effect specific to animal $i$ and $te_{ij}$ a temporary environment effect that is specific to record $j$ on animal $i$. The genetic and permanent environment effects are the same for all observations on the same individual. On the other hand, the temporary environment effects for different observations on the same individual are uncorrelated. This implies that all observations on the same individual are genetically the same trait. This leads to the concept of repeatability. Repeatability, $t$, is defined as the proportion of
the total phenotypic variance which is due to permanent effects (environment and genetic) associated with each animal. Thus, assuming no correlations between the genetic, permanent environment, and temporary environment effects, affecting a single observation,

\[
t = \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_p^2} \quad \text{or} \quad \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_g^2 + \sigma_{pe}^2 + \sigma_{te}^2}
\]  

(4.12)

Imagine that a cow, \( i \), has two lactation records, \( x_{i1} \) and \( x_{i2} \), which can be denoted as

\[
x_{i1} = g_i + pe_i + te_i1
\]

\[
x_{i2} = g_i + pe_i + te_i2
\]

The correlation between two records on an individual is

\[
r_{x_{i1}x_{i2}} = \frac{\sigma_{x_{i1}x_{i2}}}{\sqrt{\sigma_{x_{i1}}^2 \sigma_{x_{i2}}^2}}
\]

where

\[
\sigma_{x_{i1}x_{i2}} = \sigma(g_i + pe_i + te_{i1}, g_i + pe_i + te_{i2})
\]

\[
= \sigma_g^2 + \sigma_{pe}^2
\]

Hence,

\[
r_{x_{i1}x_{i2}} = \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_p^2} = t
\]

Thus, the repeatability of a trait is also the correlation between two records for that trait on the same individual; literally a measure of how "repeatable" that trait is over several records.

**EBV from Repeated Records on a Single Trait**

Imagine a situation where \( m \) records are collected on each individual and we wish to select on the mean of those \( m \) records. Then,

\[
\hat{g}_i = b_{gx} \bar{x}_i
\]  

(4.13)

where

\[
\bar{x}_i = \frac{\sum_{j=1}^{m} x_{ij}}{m}
\]  

(4.14)

and \( x_{ij} \) is the \( j^{th} \) record for the chosen trait on individual \( i \). Thus

\[
\bar{x}_i = \frac{\sum_{j=1}^{m} (g_i + pe_i + te_{ij})}{m}
\]  

(4.15)

Then,

\[
b_{gx} = \sigma_{gx} / \sigma_x^2
\]

The variance of \( \bar{x}_i \) is:

\[
\sigma_{\bar{x}_i}^2 = \sigma_g^2 + \frac{\sigma_{pe}^2 + \sigma_{te}^2}{m} = \sigma_g^2 + \frac{(1-t)\sigma_p^2}{m} = \frac{(mt + (1-t))\sigma_p^2}{m}
\]  

(4.16)

The covariance is:

\[
\sigma_{gx} = \sigma_g^2
\]  

(4.17)
Thus,
\[ b_{g^2} = \frac{m\sigma_g^2}{\sigma_p^2((m-1)t+1)} = \frac{mh^2}{(m-1)t+1} \]  \hspace{1cm} (4.18)

And accuracy of selection is given by:
\[ r = \sqrt{\frac{mh^2\sigma_g^2}{(m-1)t+1\sigma_g^2}} = \sqrt{\frac{mh^2}{(m-1)t+1}} \]  \hspace{1cm} (4.19)

Note that when \( t=1 \) there is no value in recording a trait more than once on an individual. Repeated measurements only add additional information when they allow separation of temporary and permanent effects acting on an observation.

### Numerical Example of EBV Based on the Mean of Two or More Phenotypic Records

Consider selection for milk yield with a heritability of 0.25 and a repeatability of 0.5. Assume the observation is the mean of 1, 2, 5 or 10 lactation records. Substituting \( h^2 = 0.25, t = 0.5 \) and \( m = 1, 2, 5 \) or 10 into (4.18) and (4.19) we obtain regression coefficients of
\[ b_{g^2} = 0.25, 0.333, 0.42 \text{ or } 0.45\]
and accuracies of
\[ r = 0.5, 0.58, 0.65 \text{ or } 0.67.\]

### 4.3 EBV from One Type of Relatives’ Records

The simple regression methods for estimation of BV described in the previous section for own records can be extended to one or more records on a single type of relatives.

Imagine a situation where 1 record is collected on each of \( m \) relatives of individual \( i \) for which we want to estimate the breeding value. Each relative \( j \) has the same additive genetic relationship \( a_{ij} \) with individual \( j \). Also, the relatives have the same additive genetic relationship to each other, \( a_{jj} \).
Then, the BV of individual $i$ can be predicted from the average of the records of its relatives based on:

$$\hat{g}_i = b_{g\bar{x}} \bar{x}_i$$

where

$$\bar{x}_i = \frac{1}{m} \sum_{j=1}^{m} x_{ij}$$

and $x_{ij}$ is the record on the $j^{th}$ relative of $i$.

Then,

$$b_{g\bar{x}} = \frac{\sigma_{g\bar{x}}}{\sigma_g^2}$$

To derive $\sigma_{g\bar{x}}$, let $t$ be the (intra-class) correlation between phenotypic records on relatives $j$ and $j'$:

$$t = r_{x_{ij}x_{ij'}} = \frac{\sigma_{x_{ij}x_{ij'}}}{\sigma_p^2} = \frac{\sigma_{g_{ij}+e_{ij}, g_{ij'}+e_{ij'}}}{\sigma_p^2}$$

$$= \frac{(a_{ij}\sigma_g^2 + c^2 \sigma_e^2)}{\sigma_p^2}$$

$$= a_{ij} \sigma_g^2 + c^2$$  \hspace{1cm} (4.20)

Here $c^2$ is the **common environment correlation** between records. This parameter quantifies the extent to which relatives are exposed to the same environment (e.g., litter mates):

$$c^2 = \frac{\sigma_{e_{ij}e_{ij'}}}{\sigma_e^2}$$  \hspace{1cm} (4.21)

As an aside, note that this equation for the intra-class correlation also holds for repeated own records. In that case, $a_{ij'} = 1$, $c^2 = \sigma_p^2 / \sigma_p^2$, and thus $t = h^2 + \sigma_p^2 / \sigma_p^2 = (\sigma_g^2 + \sigma_p^2) / \sigma_p^2$, which is equal to repeatability (see equation 4.12).

The variance of the mean of $m$ records with intra-class correlation $t$ can be derived as:

$$\sigma^2_{\bar{x}} = \text{Var}\left( \frac{1}{m} \sum_{j=1}^{m} x_{ij} / m \right) = \frac{m \sigma_p^2 + m(m-1)t \sigma_p^2}{m^2} = \frac{1 + (m-1)t}{m} \sigma_p^2$$  \hspace{1cm} (4.22)

The covariance is:

$$\sigma_{g\bar{x}} = a_{ij} \sigma_g^2$$  \hspace{1cm} (4.23)
Thus, 
\[
b_{\beta x} = \frac{ma_{ij}\sigma^2_g}{\sigma^2_p((m-1)t+1)} = a_{ij} \frac{mh^2}{(m-1)t+1}
\]  
(4.24)

And, accuracy of selection is given by, 
\[
r = a_{ij} \sqrt{\frac{mh^2}{(m-1)t+1}}
\]  
(4.25)

Note that for repeated own records \(a_{ij}=1\) and equations (4.24) and (4.25) simplify to equation (4.18) and (4.19).

### 2) EBV from single relative record

**Effect of degree of relationship**

- **Own** \(a_{ij}=1\)
- **Progeny** \(a_{ij}=0.5\)
- **Halfsibs** \(a_{ij}=0.25\)

<table>
<thead>
<tr>
<th>Heritability</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### 2) EBV from records on one type of relative

**Progeny**

- \(dtr_1\rightarrow -150\)
- \(a_{ij}=0.25\)
- \(t=0.25\)
- \(h^2=0.075\)
- \(b=0.24\)
- \(g=(0.69)(125)+30.6\)
- \(r=0.49\)

<table>
<thead>
<tr>
<th># records</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.2</td>
</tr>
<tr>
<td>400</td>
<td>0.4</td>
</tr>
<tr>
<td>600</td>
<td>0.6</td>
</tr>
<tr>
<td>800</td>
<td>0.8</td>
</tr>
<tr>
<td>1000</td>
<td>1.0</td>
</tr>
<tr>
<td>1200</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Sire**

- \(dtr_2\rightarrow +200\)
- \(a_{ij}=0.5\)
- \(t=0.5\)
- \(h^2=0.3\)
- \(b=0.24\)
- \(g=(0.69)(125)+30.6\)
- \(r=0.49\)

<table>
<thead>
<tr>
<th># records</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.2</td>
</tr>
<tr>
<td>400</td>
<td>0.4</td>
</tr>
<tr>
<td>600</td>
<td>0.6</td>
</tr>
<tr>
<td>800</td>
<td>0.8</td>
</tr>
<tr>
<td>1000</td>
<td>1.0</td>
</tr>
<tr>
<td>1200</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average (P*=+125)</th>
<th># records</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.2</td>
</tr>
<tr>
<td>400</td>
<td>0.4</td>
</tr>
<tr>
<td>600</td>
<td>0.6</td>
</tr>
<tr>
<td>800</td>
<td>0.8</td>
</tr>
<tr>
<td>1000</td>
<td>1.0</td>
</tr>
<tr>
<td>1200</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* each from a different dam

### 2) EBV from records on one type of relative

**Effect of # records and relationship**

- **Clones**
- **Progeny**
- **Fullsibs**
- **Halfsibs**

<table>
<thead>
<tr>
<th># records</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.1</td>
</tr>
<tr>
<td>20</td>
<td>0.2</td>
</tr>
<tr>
<td>30</td>
<td>0.3</td>
</tr>
<tr>
<td>40</td>
<td>0.4</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
</tr>
<tr>
<td>60</td>
<td>0.6</td>
</tr>
<tr>
<td>70</td>
<td>0.7</td>
</tr>
<tr>
<td>80</td>
<td>0.8</td>
</tr>
<tr>
<td>90</td>
<td>0.9</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### 2) EBV from records on one type of relative

**Effect of heritability on progeny test accuracy**

<table>
<thead>
<tr>
<th>Heritability</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>0.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### 4.4 EBV from Multiple Sources - Selection Index

When records are available from multiple sources, e.g. records on the animal itself, its dam, half-sibs, progeny, etc., it will obviously be most beneficial to use all records to estimate the breeding value. This can be achieved by extending the simple regression methods described in the previous to a multiple regression setting:

\[
g_i = b_1 x_1 + b_2 x_2 + ... + b_m x_m
\]  
(4.26)
where \( x_i \) represents the \( i^{th} \) source of records, which could be an individual record or the mean of records on a given type of relative, and \( b_i \) are partial regression coefficients. Equation (4.26) is called a selection index and the coefficients \( b_i \) are called index weights. The methodology that is used to derive the optimal index weights, i.e. those that maximize the accuracy of the EBV, is called selection index theory.

The selection index was first proposed by Smith (1936) for use in plant breeding for simultaneous selection on multiple traits, and seven years later, but apparently independently, by Hazel (1943) for animal breeding. In this Chapter we shall first discuss the basic problem, then go on to derive selection index equations, and then illustrate their use with some examples.

Selection index theory deals with the general problem of combining information from a variety of sources in such a way that the most accurate predictor of the overall genetic merit for a pre-defined combination of traits is obtained. Two separate types of selection indexes can be distinguished: 1) the economic selection index, where information from several recorded traits is used to predict genetic merit for overall economic value, and 2) the family selection index, where information from a single trait on various relatives is combined to predict the genetic merit of an individual for that trait.

3) Estimating EBV from Multiple Sources of Information

Maximizing Accuracy of EBV

Optimize weight given to each record

\[
E_{\text{select}} = b_1 x_{\text{own}} + b_2 x_{\text{HS}} + b_3 x_{\text{dam}} + b_4 x_{\text{prog}} + \ldots
\]

Weights by selection index theory = Multiple regression

Optimal weights depend on

- genetic relationships
- genetic parameters of trait
- availability of other records

The economic selection index and family selection index are special cases of the general selection index, where the selection index is defined as a linear function of a series of observations which when selected upon maximizes response of an aggregate genotype, which is a linear function of the additive genetic values of a defined set of traits. Although the focus in this Chapter is prediction of breeding values for a single trait, we will develop the theory of selection indexes within the context of the economic index because it is more general. We will then discuss the family index as a special case of the economic selection index and go into more detail into family indexes and their extension to modeling BLUP EBV. We will come back to various applications related to economic indexes in Chapter 5.

4.4.1 Selection Index theory

In economically oriented breeding programs, the trait that we want to improve could be called economic merit. The breeding objective of our program is then to maximize improvement of
economic merit. Economic merit might be defined in different ways, e.g. as profit per animal, profit per enterprise, economic efficiency, or something else. We will return to this problem in later Chapters. For the present, it is only necessary to recognize that the breeding objective is a general statement of the economic genetic goal of the breeding program.

For a given definition of the breeding objective, there will likely be several or many traits, which would contribute to the objective. The aggregate genotype is then defined as a function of the additive genetic values of the traits of interest of an individual, which if selected upon would achieve the breeding objective. The function need not necessarily be linear, but in many cases an approximate linear relationship can be found that adequately defines aggregate genotype over the range of genetic values encountered (see later chapters). If the function is a linear function, then the aggregate genotype, $H$, can be written as

$$H = v_1 g_1 + v_2 g_2 + \ldots + v_n g_n = v'g$$

(4.27)

where $g_i$ is the additive genetic value of trait $i$, expressed as a deviation from the population mean, and $v_i$ is a weighting factor (usually, but not necessarily, an economic weight) for trait $i$. In vector notation, $v' = [v_1, v_2, \ldots, v_n]$ and $g' = [g_1, g_2, \ldots, g_n]$.

In practice, the additive genetic values (i.e. true BV) of the various traits for an individual are not known. However we can record each individual’s performance for a number of traits. The observations on these traits can then be combined into a selection index, $I$ of the form,

$$I = b_1 x_1 + b_2 x_2 + \ldots + b_m x_m = b'x$$

(4.28)

where $x_j$ is the $j^{th}$ phenotypic observation, as a deviation from the population mean, and $b_i$ is a selection index coefficient (weight) for that observation. In vector notation, $b' = [b_1, b_2, \ldots, b_m]$ and $x' = [x_1, x_2, \ldots, x_m]$. In principle, observations $x_j$ do not necessarily have to be on the traits that are in the aggregate genotype or on the animal that is being evaluated; observations can be on any trait and from the animal itself or its relatives.

The problem is then to estimate the selection index weights, $b_i$, such that selection of individuals on their selection index value, $I$, maximizes response in the aggregate genotype, $H$. Equivalently, we want to find $b_i$ such that the correlation between $I$ and $H$ is maximized, or that the variance of prediction errors ($\text{Var}(H-I)$) is minimized.

With family selection indexes, the problem is to combine information from different types of relatives to provide the most accurate estimate of the additive genetic value of a given trait ($g$) for a given individual. In this case, the aggregate genotype is given by $H = g$ and, thus $v = [1]$. In this case the selection index is equal to the EBV for the trait evaluated:

$$I = \hat{g} = b_1 x_1 + b_2 x_2 + \ldots + b_m x_m$$

(4.29)

Similar to an economic index, a family index can include information on the animal itself and its relatives for the trait being evaluated, as well as records on other traits. Thus, the derivations that follow for an economic index also apply to family indexes by setting $H = g$ and $v = [1]$. 

48
4.4.1.1 Derivation of index coefficients

We wish to define $I$ such that selection of animals on $I$ maximizes response in $H$. From standard regression theory (see also Chapter 3) expected response (genetic superiority) of selected individuals in $H$, $S_H$, is given by

$$S_H = b_{HI} I^2 (I - I_0)$$ (4.30)

where $b_{HI}$ is the regression of aggregate genotype on index values, $I$ is the index value of the selected animal or group of animals, and $I_0$ is the mean index value of all selection candidates. Since $I - I_0$ can be written as $i\sigma_I$, where $i$ is the intensity of selection (see Chapter 3),

$$S_H = b_{HI} i \sigma_I = \frac{\sigma_{HI}}{\sigma_I} i \sigma_I / \sigma_I$$ (4.31)

Thus for any given intensity of selection, $i$, response in $H$ is maximized when $\sigma_{HI}/\sigma_I$ is maximized.

Apart from maximizing response in $H$ to selection on $I$, it would also be useful if the index value, $I$, was an unbiased predictor of the aggregate genotypic value $H$. This means that the true aggregate genotype of an individual is, on average, no more likely to be greater than its index value than it is to be less than its index value, or

$$E(H - \bar{H}) = I - \bar{I}$$ (4.32)

Under the assumption of multivariate normality, this is achieved when the regression of $H$ on $I$, $b_{HI} = 1$. Thus we wish to find the index coefficients $b_1, b_2 \ldots b_n$ that maximize $\sigma_{HI}/\sigma_I$, subject to $b_{HI} = 1$.

Considering first the maximization of $\sigma_{HI}/\sigma_I$. Let $\sigma_{g_{ki}}$ be the genetic covariance between the $k^{th}$ observation in the index and the $i^{th}$ trait in the aggregate genotype. Similarly, let $\sigma_{p_{ki}}$ be the phenotypic covariance between the $k^{th}$ and $l^{th}$ observations in the selection index. Recalling the definition of $I$ given by equation (4.28), it follows that

$$\sigma_i^2 = b_1^2 \sigma_{p_{11}} + b_2^2 \sigma_{p_{22}} + \ldots + 2b_1b_2\sigma_{p_{12}} + 2b_1b_3\sigma_{p_{13}} \ldots = \sum_{k=1}^m \sum_{l=1}^m b_k b_l \sigma_{p_{ki}}$$ (4.33)

Similarly, the covariance between $H$ and $I$, recalling the definitions given at (4.27) and (4.28), is

$$\sigma_{HI} = b_1 v_1 \sigma_{g_{11}} + b_1 v_2 \sigma_{g_{12}} + \ldots + b_m v_n \sigma_{g_{mn}} = \sum_{k=1}^m \sum_{l=1}^m b_k v_l \sigma_{g_{ki}}$$ (4.34)

If we write the term to be maximized as, $M = \sigma_{HI}/\sigma_I$ then

$$\log M = \log \sigma_{HI} - \log \sigma_I$$
or \[ \log M = \log \sigma_{HI} - \frac{1}{2} \log \sigma_i^2 \]

and substituting from (4.33) and (4.34):

\[ \log M = \log(\sum b_k v_i \sigma_{g_i}) - \frac{1}{2} \log (\sum b_k b_i \sigma_{p_{ki}}^2) \]

(4.35)

Since \( M \) will be maximal when \( \log M \) is maximal, we can maximize \( M \) by differentiating \( \log M \) with respect to each of the \( b \) in turn and setting each partial differential to zero:

\[ \frac{\delta \log M}{\delta b_k} = 0 \quad \text{for } k = 1 \text{ to } m. \]

From standard differential algebra, with \( \log M \) defined at (4.35), it follows that

\[ \frac{\delta \log M}{\delta b_k} = \sum_{i=1}^n v_i \sigma_{g_i}^2 \frac{\sigma_{HI}}{\sigma_{HI}} - \sum_{i=1}^m b_i \sigma_{p_{ki}} \frac{\sigma_i^2}{\sigma_i^2} \]

Hence, \( M \) is maximal when

\[ \sum_{i=1}^m b_i \sigma_{p_{ki}} = \frac{\sigma_{HI}^2}{\sigma_{HI}} \sum_{i=1}^n v_i \sigma_{g_i} \]

(4.36)

But from standard regression theory:

\[ \frac{\sigma_i^2}{\sigma_{HI}} = \frac{1}{b_{HI}} \]

and if the index \( I \) is to give unbiased estimates of the aggregate genotype \( H \), we recall that \( b_{HI} \) must equal 1. Hence (4.36) becomes,

\[ \sum_{i=1}^m b_i \sigma_{p_{ki}} = \sum_{i=1}^n v_i \sigma_{g_i} \]

(4.37)

Since there are \( m \) observations in the index, there are \( m \) equations of the general form of (4.37), i.e.

\[ \sum_{i=1}^m b_i \sigma_{p_{ki}} = \sum_{i=1}^n v_i \sigma_{g_i} \]

\[ \sum_{i=1}^m b_i \sigma_{p_{k2}} = \sum_{i=1}^n v_i \sigma_{g_{z2}} \]

\[ \vdots \]

\[ \sum_{i=1}^m b_i \sigma_{p_{kw}} = \sum_{i=1}^n v_i \sigma_{g_{zw}} \]

If we write these equations in their expanded form, i.e.
\[
\begin{align*}
&b_1 \sigma_{p_{1,1}} + b_2 \sigma_{p_{1,2}} \ldots + b_m \sigma_{p_{1,m}} = v_1 \sigma_{g_{1,1}} + v_2 \sigma_{g_{1,2}} \ldots + v_n \sigma_{g_{1,n}} \\
&b_1 \sigma_{p_{2,1}} + b_2 \sigma_{p_{2,2}} \ldots + b_m \sigma_{p_{2,m}} = v_1 \sigma_{g_{2,1}} + v_2 \sigma_{g_{2,2}} \ldots + v_n \sigma_{g_{2,n}} \\
&\vdots \\
&b_1 \sigma_{p_{m,1}} + b_2 \sigma_{p_{m,2}} \ldots + b_m \sigma_{p_{m,m}} = v_1 \sigma_{g_{m,1}} + v_2 \sigma_{g_{m,2}} \ldots + v_n \sigma_{g_{m,n}}
\end{align*}
\]

it is clear that they can be written in matrix notation as:

\[
Pb = Gv
\]  
(4.38)

where \(b\) is a column vector of \(m\) selection index coefficients, \(P\) is an \(m \times m\) matrix of phenotypic covariances among the observations in the selection index, \(G\) is an \(m \times n\) matrix of genetic covariances among the \(m\) index observations and the \(n\) traits in the aggregate genotype, and \(v\) is a column vector of economic weights of the \(n\) traits in the aggregate genotype.

Recalling that pre-multiplying a matrix by itself yields an identity matrix, i.e. that, \(P^{-1} P = I\), the solution to obtaining \(b\) can be obtained by pre-multiplying both sides of (4.38) by \(P^{-1}\) to obtain,

\[
b = P^{-1}Gv
\]  
(4.39)

These are the so-called selection index equations that must be solved to find the optimal index weights.

4.4.1.2 Alternative derivation using matrix notation

The object is to minimize the variance of the difference between the predicted value, \(I\), and the true value, \(H\), i.e. minimize \(\text{Var}(H - I)\). Thus we wish to minimize

\[
E(H - I)^2 = E[I - H]'(I - H)]
\]

\[
= E[I - H]'(I - H)]
\]

\[
= E[(b'x - v'\tilde{g})(x'b - g'v)]
\]

\[
= E[(b'xxb - b'xg'v - v'gx'b + v'gg'v]
\]

where \(x\) = column vector of observations and \(g\) = column vector of genetic values. Each of the terms in the above equality can be found as:

\[
E(b'xxb) = b'E(xx')b = b'Pb,
\]

\[
E(b'xg'v) = b'E(xg')v = b'Gv,
\]

\[
E(v'gx'b) = v'G'b = b'Gv \quad \text{since } v'G'b \text{ is a scalar}
\]

and

\[
E(v'gg'v) = v'E(gg')v = v'Cv
\]
Therefore, to minimize \( M = b'Pb - 2b'Gv + v'Cv \)

we must find the values which correspond to
\[
\frac{\delta M}{\delta b} = 0 = 2Pb - 2Gv + 0
\]

Therefore
\[
Pb = Gv
\]

Hence,
\[
b = P^{-1}Gv \quad \text{which is identical to equation (4.39).}
\]

**4.4.1.2 Accuracy of the index**

The accuracy of the selection index can be computed as the correlation between \( I \) and \( H \):
\[
r_{HI} = \frac{\sigma_{HI}}{\sigma_I \sigma_H} \tag{4.40}
\]

The variance of the index, \( \sigma^2_I \), is easily found as
\[
\sigma^2_I = \text{Var}(b_1x_1 + b_2x_2 \ldots b_mx_m)
= b_1^2 \sigma_{p_1}^2 + b_2^2 \sigma_{p_2}^2 + \ldots + 2b_1b_2 \sigma_{p_1p_2} + 2b_1b_3 \sigma_{p_1p_3}
\]
or in matrix notation:
\[
\sigma^2_I = \text{Var}(b'x) = b' \text{Var}(x)b = b'Pb \tag{4.41}
\]

Following the same argument as for \( \sigma^2_I \),
\[
\sigma^2_H = \text{Var}(v'g) = v' \text{Var}(g)v = v'Cv \tag{4.42}
\]

where \( C \) is an \( n \times n \) matrix of genetic covariances among the traits in the aggregate genotype.

Similarly, it follows that
\[
\sigma_{HI} = \text{Cov}(b'x, v'g) = b' \text{Cov}(x,g)v = b'Gv \tag{4.43}
\]

Hence,
\[
r_{HI} = \frac{\sigma_{HI}}{\sigma_I \sigma_H} = \frac{b'Gv}{\sqrt{b'Pb} \sqrt{v'Cv}} \tag{4.44}
\]

Note that because the index was constrained such that \( b_{HI} = 1 \) and \( b_{HI} = \sigma_{HI} / \sigma_I^2 \), thus \( \sigma_{HI} = \sigma^2_I \)

and from equations (4.41) and (4.43),
\[
b'Pb = b'Gv \tag{4.45}
\]

Thus, for the optimal index, equation (4.44) for accuracy simplifies to:
\[
r_{HI} = \frac{\sigma_I}{\sigma_H} = \frac{b'Pb}{\sqrt{v'Cv}} = \frac{b'Gv}{\sqrt{v'Cv}} \tag{4.46}
\]

Note, however, that equations (4.45) and (4.46) only hold for the optimal index, whereas equation (4.44) holds for any arbitrary index.
4.4.2 Family Selection Indexes

With family selection indexes, the problem is to combine information from different types of relatives to provide the most accurate estimate of the additive genetic value of a given trait \( g \) for a given individual. As indicated previously, in this case \( H = g, \ v = [1], \) and \( \sigma_H^2 = \sigma_g^2 \). This simplifies derivations to:

from equation (4.39)  \[ b = P^{-1}G \]  
(4.47)

and from equation (4.46)  \[ r_{HI} = r_{g,g} = \frac{b^*G}{\sigma_G^2} \]  
(4.48)

4.4.2.1 Examples of family selection indexes

Single source of information

The simplest form of a family index are the cases discussed in sections 4.2 and 4.3, where only a single source of observations is used, i.e. a single record or the mean of \( m \) records of the same type. The simplest case is a single record of the phenotype of the individual itself. In this case, the selection index is \( I = \gamma = b_1x_1 \) and the aggregate genotype is \( H = g \)

where \( x_1 \) and \( g \) are both expressed as deviations from their population mean.

In this case,  \[ P = \sigma_x^2 \] and  \[ G = \sigma_g^2 \]

Hence,  \[ b = b = P^{-1}G = (\sigma_x^2)^{-1}\sigma_g^2 = \sigma_g^2 / \sigma_x^2 \]

The accuracy of selection, given by (4.48), is  \[ r_{HI} = r_{g,g} = \frac{bG}{\sigma_g^2} = \frac{bG}{C} = \frac{\sigma_g^2}{\sigma_x^2\sigma_g^2} \]

These results are equivalent to those obtained in section 4.4.2.

More than one observation in the index

For the previous example, when there was only one source of information in the index, algebraic expectations for \( b \) and \( r_{HI} \) were derived directly in terms of basic population parameters. Appropriate formulae can be derived for a wide range of situations, including some situations with two or more sources for a single trait. A few more examples are given in Table 4.1, and a more extensive list is given by Van Vleck, 1993. Once there is more than one source of information in the index, it is often more useful to derive the expectations for the elements of \( P \) and \( G \) and then solve for \( b, b_{HI}, \) etc. using a computer package for matrix programming, rather than attempting to derive an algebraic solution directly.
Table 4.1 Selection index coefficients, $b$, and accuracies, $r_{HI}$, for some common sources of information in family indexes to predict additive genetic value for a single trait.

<table>
<thead>
<tr>
<th>Information Source</th>
<th>$b$</th>
<th>$r_{HI} = r_{g,b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single record on individual</td>
<td>$h^2$</td>
<td>$\sqrt{h^2}$</td>
</tr>
<tr>
<td>$m$ records on individual</td>
<td>$\frac{m h^2}{(m-1)t+1}$</td>
<td>$\sqrt{\frac{m h^2}{(m-1)t+1}}$</td>
</tr>
<tr>
<td>Single record on one parent</td>
<td>$\frac{1}{2} h^2$</td>
<td>$\frac{1}{2} \sqrt{h^2}$</td>
</tr>
<tr>
<td>$m$ records on one parent</td>
<td>$\frac{m h^2}{2((m-1)r+1)}$</td>
<td>$\frac{1}{2} \sqrt{\frac{m h^2}{(m-1)r+1}}$</td>
</tr>
<tr>
<td>Single record on both parents</td>
<td>$\frac{1}{2} h^2, \frac{1}{2} h^2$</td>
<td>$0.71 \sqrt{h^2}$</td>
</tr>
<tr>
<td>$m$ records on both parents</td>
<td>$\frac{2mh^2}{2((m-1)t+1)} + \frac{mh^2}{2((m-1)r+1)}$</td>
<td>$0.71 \sqrt{\frac{m h^2}{(m-1)t+1}}$</td>
</tr>
<tr>
<td>Mean of $n$ half-sib progeny with one record</td>
<td>$\frac{2nh^2}{(n-1)h^2 + 4}$</td>
<td>$\sqrt{\frac{mh^2}{(n-1)h^2 + 4}}$</td>
</tr>
</tbody>
</table>

4.4.2.2 General equations to derive elements of selection index matrices

This section describes general equations that can be used to derive elements of the $P$, $G$, and $C$ matrices that are needed for selection index calculations. Possible sources of information in the index are individual records and the mean of $m$ records on a group of individuals or of $m$ own records. Records on different traits can be included in the index and the aggregate genotype can consist of a single trait or of multiple traits.

It must be noted that these equations assume no selection or inbreeding. The impact of selection and inbreeding on index derivations will be discussed in a later chapter.

Notation:
- $m$ = number of records within a group
- $c^2$ = common environment component within a group of individuals that contribute to a mean
- $\sigma_{p_k}$ = phenotypic standard deviation of trait $k$
- $\sigma_{g_k}$ = additive genetic standard deviation of trait $k$
- $r_{pkl}$ = phenotypic correlation between traits $k$ and $l$
- $r_{gkl}$ = genetic correlation between traits $k$ and $l$
- $a$ = additive genetic relationship within a group
\[ a_{ij} = \text{additive genetic relationship between individual(s) in groups } i \text{ and } j \]
\[ a_{hj} = \text{additive genetic relationship between the individual in the breeding goal (} h \text{) and individuals in group } j \]

**P-matrix**

**diagonal:**
- Variance of \( m \) records of a given type
  \[
  \frac{1+(m-1)t}{m} \sigma_p^2 \quad (= \sigma_p^2 \text{ for } m=1)
  \]  
  \[ (4.49) \]
  with \( t = \text{repeatability for repeated records} \)
  \[ t = ah^2 + c^2 \text{ for multiple individuals} \]

**off-diagonal:**
- Covariance between mean of \( m \) records on different traits (\( k \) and \( l \)) for the same group:
  \[
  \frac{r_{pj} \sigma_{pj} \sigma_{pj} + (m-1)ar_{sl} \sigma_{sk} \sigma_{sl}}{m} \quad (= r_{pj} \sigma_{pj} \sigma_{pj} \text{ for } m=1)
  \]  
  \[ (4.50) \]
- Covariance between (mean of) record(s) on same trait \( k \) for different groups (\( i \) and \( j \)):
  \[
  (a_{ij} h_k^2 + c_k^2) \sigma_{pk}^2
  \]  
  \[ (4.51) \]
- Between records on different traits (\( k \) and \( l \)) in different groups (\( i \) and \( j \)):
  \[
  a_{ij} r_{sl} \sigma_{sk} \sigma_{sl}
  \]  
  \[ (4.52) \]

**G-matrix**
- Covariance of the genetic value for trait \( k \) on the breeding goal animal (\( h \)) with records on trait \( l \) for group \( j \)
  \[
  a_{hj} r_{sl} \sigma_{sk} \sigma_{sl} \quad (= a_{hj} \sigma_{sk}^2 \text{ if } k=l)
  \]  
  \[ (4.53) \]

**C-matrix**

**Diagonal:**
- Variance of genetic value for trait \( k \)
  \[
  \sigma_{sk}^2
  \]  
  \[ (4.54) \]

**Off-diagonal:**
- Covariance between genetic values for traits \( k \) and \( l \) on breeding goal animal
  \[
  r_{sl} \sigma_{sk} \sigma_{sl}
  \]  
  \[ (4.55) \]

**4.4.2.2.1 Example Index of individual record and full-sib mean performance**

Imagine a situation where we have an observation on the individual’s performance plus the mean performance of that individual’s \( m \) full sibs, and we wish to predict the individual’s breeding value. The index will then take the form,
\[ I = g = b_1 x_1 + b_2 x_2 \]

where \( x_1 \) is the individual’s phenotype and \( x_2 \) is the full-sib mean phenotype, both expressed as deviations from the population mean. Then \( P \) and \( G \) will take the form,

\[
P = \begin{bmatrix}
\sigma^2_{x_1} & \sigma_{x_1 x_2} \\
\sigma_{x_1 x_2} & \sigma^2_{x_2}
\end{bmatrix}, \quad G = \begin{bmatrix}
\sigma^2_{x_1 g} \\
\sigma_{x_1 g}
\end{bmatrix}
\] (4.56)

Elements of \( P \) and \( G \) can be derived using the equations developed in the previous section. As an example, consider a selection index based on individual phenotype and the mean performance of 5 full sibs for animals in a population recorded for growth rate with a heritability of 0.5. We will assume there is no common environmental component.

Then:

\[
P = \begin{bmatrix}
1 & \frac{1}{2} h^2 \\
\frac{1}{2} h^2 & \frac{1 + (m-1)\frac{1}{2} h^2}{m}
\end{bmatrix} \sigma^2_p = \begin{bmatrix}
1 & .25 \\
.25 & .4
\end{bmatrix} \sigma^2_p
\] (4.57)

and:

\[
G = \begin{bmatrix}
\frac{1}{2} h^2 \\
\frac{1}{2} h^2
\end{bmatrix} \sigma^2_p = \begin{bmatrix}
.5 \\
.25
\end{bmatrix} \sigma^2_p
\] (4.58)

Selection index coefficients are given by \( b = P^{-1} G \) which, since \( \sigma^2_p \) cancels out, gives

\[
b = \begin{bmatrix}
1 & .25 \\
.25 & .4
\end{bmatrix}^{-1} \begin{bmatrix}
.5 \\
.25
\end{bmatrix} = \begin{bmatrix}
.4074 \\
.3704
\end{bmatrix}
\]

Hence, the selection index would be

\[ I = g = 0.4074 x_1 + 0.3704 x_2 \]

The accuracy of this index or EBV is given by

\[
r_{HI} = r_{gs} = \sqrt{\frac{b' G}{\sigma^2_g}} = \sqrt{\frac{.4074 \cdot .5}{0.5 \sigma^2_p}} = 0.77
\] (4.59)

We can compare this accuracy with the accuracy of 0.707 for phenotypic selection on the same trait as shown in Section 2.8.1. By adding information on the mean performance of 5 full sibs, the accuracy of evaluation is increased from 0.71 to 0.77, i.e. by 8.9%. And, since \( S = i r_{gs} \sigma_g \), and \( i \) and \( \sigma_g \) are not affected by the addition of extra information to the index, expected response will also increase by 8.9%.
4.5 Selection Index and Animal Model BLUP

An assumption in the use of selection indexes to estimate breeding values is either that there are no fixed effects in the data used, or that fixed effects are known without error. This may be true in some situations. An example are some forms of selection in egg-laying poultry where all birds are hatched in one or two very large groups and reared and recorded together in single locations. But in most cases, fixed effects are important and not known without error. For example, with pigs, different litters are born at different times of the year, often in several different locations. In progeny testing schemes in dairy cattle, cows are born continuously, begin milking at different times of year and in a very large number of different herds.

For this reason (and others) genetic evaluation in practice is often based on methods of Best Linear Unbiased Prediction, BLUP, which is a linear mixed model methodology which simultaneously estimates random genetic effects while accounting for fixed effects in the data in an optimum way. Relationships among animals can be included in the model. A sire model would account for relationships through the sire, i.e. half-sibships. A sire and dam model accounts for relationships through both the sire and the dam, i.e. full and half-sibships. An animal model accounts for all relationships among all animals in the data set. A description of the theory and application of BLUP, and animal model BLUP in particular, can be found in Schmidt (1988), Mrode (1996), and Lynch and Walsh (1998).

When relationships are included in a BLUP procedure, the method is equivalent to a selection index with the additional ability to efficiently estimate and correct the data for fixed effects. In the absence of fixed effects, BLUP with relationships is identical to a selection index. For example, a BLUP sire and dam model without records on the sire and dam would be the same as a selection index based on individual, full sib and half-sib records. An animal model BLUP would be equivalent to a selection index based on all related individuals, including ancestors, with records.

These equivalences are important for the design of breeding programs, because it means that in many situations, many aspects of selection programs with BLUP evaluation can be effectively studied with simulations based on equivalent selection indexes. There are two approaches to modeling Animal model BLUP EBV using selection index:

1) Develop a selection index based only on those relatives providing the greatest amount of information, rather than all possible relatives as in the animal model. For example, when records on parents, full and half sibs, and progeny are accounted for, information on more distant relatives may only provide a trivial increase in accuracy of selection.

2) Develop a selection index that includes parental EBV as sources of information, along with records on the individual itself, collateral relatives, and progeny, if available. In such an index, the parental EBV account for all ancestral information.

Development of the first type of index follows from the previous sections. We will describe the development of the second type of index in more detail in the following.
Consider the following information sources to estimate the BV of individual \( i \) for a hierarchical breeding design in which each sire is mated to \( m \) dams and each dam has \( n \) progeny (Figure 4.1):

- \( x_i \) = the animal’s own record,
- \( x_{fs} \) = the average of single records on the individual’s \( n-1 \) full sibs
- \( x_{hs} \) = the average of single records on the individual’s \((m-1)n\) half sibs
- \( \hat{g}_s \) = the EBV of the individual’s sire, excluding \( x_i \), \( x_{fs} \), and \( x_{hs} \)
- \( \hat{g}_d \) = the EBV of the individual’s dam, excluding \( x_i \), \( x_{fs} \), and \( x_{hs} \)
- \( \bar{g}_m \) = the mean EBV of the \((m-1)\) mates of the sire that produced the individual’s half sibs

Based on this information, the selection index to estimate the individual’s BV can be formulated as:

\[
I_i = \hat{g}_i = b_1 x_i + b_2 x_{fs} + b_3 x_{hs} + b_4 \hat{g}_s + b_5 \hat{g}_d + b_6 \bar{g}_m
\]  

(4.60)

\[
P = \begin{bmatrix}
\sigma^2_{x_i} & \sigma_{x_i x_{fs}} & \sigma_{x_i x_{hs}} & \sigma_{x_i \hat{g}_s} & \sigma_{x_i \hat{g}_d} & \sigma_{x_i \bar{g}_m} \\
\sigma_{x_{fs}} & \sigma^2_{x_{fs}} & \sigma_{x_{fs} x_{hs}} & \sigma_{x_{fs} \hat{g}_s} & \sigma_{x_{fs} \hat{g}_d} & \sigma_{x_{fs} \bar{g}_m} \\
\sigma_{x_{hs}} & \sigma_{x_{hs} x_{fs}} & \sigma^2_{x_{hs}} & \sigma_{x_{hs} \hat{g}_s} & \sigma_{x_{hs} \hat{g}_d} & \sigma_{x_{hs} \bar{g}_m} \\
\sigma_{\hat{g}_s} & \sigma_{\hat{g}_s x_{fs}} & \sigma_{\hat{g}_s x_{hs}} & \sigma^2_{\hat{g}_s} & \sigma_{\hat{g}_s \hat{g}_d} & \sigma_{\hat{g}_s \bar{g}_m} \\
\sigma_{\hat{g}_d} & \sigma_{\hat{g}_d x_{fs}} & \sigma_{\hat{g}_d x_{hs}} & \sigma_{\hat{g}_d \hat{g}_s} & \sigma^2_{\hat{g}_d} & \sigma_{\hat{g}_d \bar{g}_m} \\
\sigma_{\bar{g}_m} & \sigma_{\bar{g}_m x_{fs}} & \sigma_{\bar{g}_m x_{hs}} & \sigma_{\bar{g}_m \hat{g}_s} & \sigma_{\bar{g}_m \hat{g}_d} & \sigma^2_{\bar{g}_m} \\
\end{bmatrix}
\]  

(4.61)

\[
G = \begin{bmatrix}
\sigma_{g_i x_i} & \sigma_{g_i x_{fs}} & \sigma_{g_i x_{hs}} & \sigma_{g_i \hat{g}_s} & \sigma_{g_i \hat{g}_d} & \sigma_{g_i \bar{g}_m} \\
\end{bmatrix}
\]  

(4.62)
\[ \mathbf{P} = \begin{bmatrix}
1 & \frac{1}{2} h^2 + c^2 & \frac{1}{4} h^2 & \frac{1}{2} r_s^2 h^2 & \frac{1}{2} r_d^2 h^2 & 0 \\
\frac{1+ (n-2)(\frac{1}{2} h^2 + c^2)}{n-1} & \frac{1}{4} h^2 & \frac{1}{2} r_s^2 h^2 & \frac{1}{2} r_d^2 h^2 & 0 \\
\frac{1}{4} h^2 + \frac{1}{4} h^2 + c^2 + \frac{1}{m-1} & \frac{1}{2} \frac{1}{2} r_s^2 h^2 & \frac{1}{2} r_s^2 h^2 & 0 \\
\frac{1}{m-1} & \frac{1}{2} r_s^2 h^2 & 0 & 0 \\
\frac{1}{n(m-1)} & \frac{1}{2} r_d^2 h^2 & 0 & 0 \\
\frac{1}{m-1} & \frac{1}{2} r_d^2 h^2 & 0 & \frac{r_s^2 h^2}{m-1} \\
\end{bmatrix} \sigma_p^2 \text{ (4.63)} \]

\[ \mathbf{G} = \begin{bmatrix}
h^2 & \frac{1}{2} h^2 & \frac{1}{4} h^2 & \frac{1}{2} r_s^2 h^2 & \frac{1}{2} r_d^2 h^2 & 0 \\
0 & \sigma_p^2 \\
0 & \sigma_p^2 \\
0 & \sigma_p^2 \\
0 & \sigma_p^2 \\
0 & \sigma_p^2 \\
\end{bmatrix} \text{ (4.64)} \]

\[ \mathbf{x}_{hs} = \sum_{k=1}^{m-1} \sum_{l=1}^{n} \frac{x_{kl}}{n} / (m-1) \text{ (4.65)} \]

Where

\[ x_{kl} = \frac{1}{2} g_s + \frac{1}{2} g_d + g_{mskl} + c_{kl} + e_{kl} \text{ (4.66)} \]

Thus

\[ \mathbf{x}_{hs} = \frac{1}{2} g_s + \frac{1}{2} g_d + \frac{\sum_{k=1}^{m-1} (g_{dk} + c_k)}{n} + \frac{\sum_{k=1}^{n} (g_{mskl} + e_{kl})}{n(m-1)} \text{ (4.67)} \]

And

\[ \sigma_{x_{hs}}^2 = \frac{1}{4} \sigma_g^2 + \frac{1}{4} \sigma_g^2 + \frac{1}{2} \sigma_g^2 + \frac{1}{2} \sigma_g^2 + \frac{1}{2} \sigma_g^2 \text{ (4.68)} \]

Also,

\[ \sigma_g^2 = r_g^2 \sigma_g^2 \text{ (4.69)} \]

And

\[ \sigma_{x_{hs}}^2 = \sigma_{(1/2 g_s + 1/2 g_d + g_{ms} + e_s)}^2 = \sigma_{(1/2 g_s + e_s)}^2 = \frac{1}{2} r_s^2 \sigma_g^2 \text{ (4.70)} \]

As before, index weights can be derived as:

\[ \mathbf{b} = \mathbf{P}^{-1} \mathbf{G} \]

And accuracy as:

\[ r_{g,h} = \sqrt{\mathbf{b}^T \mathbf{Pb} / \sigma_g^2} \]

Because elements of the \( \mathbf{P} \) and \( \mathbf{G} \) matrices depend on accuracy of EBV of the sire and dam, which in turn depend on the EBV of their parents, iteration must be used to derive the final index and its accuracy. This can be done by using some starting value for accuracy of parental EBV, e.g. \( r_s = r_d = h \), deriving the index and its accuracy, and then using the resulting accuracy as the new accuracy for \( r_s \) and \( r_d \), resolving the index, etc.. This process of iteration is akin to building pedigree information; in each iteration, an additional ancestral generation with data is added, which increases accuracy but at a diminishing rate, until accuracy asymptotes (see example).

In the previous selection indexes were used to provide genetic evaluations for a single trait based on records of that trait on the individual and/or other relatives. This is known as single-trait...
evaluation. It should be clear from selection index theory, that information on other traits could also be included in the index, to give a multi-trait evaluation (see Villanueva et al. 1993).

$$\text{EBV}_i = b' \left[ x_i, x_{fs} , x_{hs} , g_{\text{sire}} , g_{\text{dam}} , g_{\text{mates}} \right]$$

$$b = P^{-1} G \gamma_i$$

**Iterative Procedure to Compute Accuracy of EBV**

1) Set accuracy of $$g_{\text{sire}}, g_{\text{dam}}, g_{\text{mates}}$$ equal to $$h$$ (own record)
2) Set up index $$(P, G)$$ and derive accuracy $r_{\gamma_i} = \sqrt{b'Pb}$
3) Set accuracy of $$g_{\text{sire}}, g_{\text{dam}}, g_{\text{mates}}$$ equal to $$r_{\gamma_i}$$
4) Repeat steps 2 and 3 until accuracy converges

Needs adaptation if selection after progeny

---

**Iteration builds pedigree info**

<table>
<thead>
<tr>
<th>Iteration 1</th>
<th>Iteration 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example</strong></td>
<td><strong>Example</strong></td>
</tr>
<tr>
<td>Iteration 1</td>
<td>Iteration 2</td>
</tr>
<tr>
<td><strong>Iteration 1</strong></td>
<td><strong>Iteration 2</strong></td>
</tr>
<tr>
<td>n = 100</td>
<td>75</td>
</tr>
<tr>
<td>$h^2 = 0.25$</td>
<td>$h^2 = 0.75$</td>
</tr>
<tr>
<td>$c^2 = 0$</td>
<td>$c^2 = 0.50$</td>
</tr>
<tr>
<td><strong>Starting Value</strong></td>
<td><strong>Starting Value</strong></td>
</tr>
<tr>
<td><strong>For Accuracy of Parental EBV</strong></td>
<td>0.5000</td>
</tr>
<tr>
<td>$r_{\theta_{sire}}r_{\theta_{dam}}r_{\theta_{mates}} = 0.5000$</td>
<td>(start with n=1)</td>
</tr>
<tr>
<td>$P = x_j$</td>
<td>$P = x_j$</td>
</tr>
<tr>
<td>$100.00, 12.50, 6.25, 3.13, 3.13, 0.00$</td>
<td>$100.00, 12.50, 6.25, 5.93, 5.93, 0.00$</td>
</tr>
<tr>
<td>$x_{fs}$</td>
<td>$x_{fs}$</td>
</tr>
<tr>
<td>$12.50, 22.22, 6.25, 3.13, 3.13, 0.00$</td>
<td>$12.50, 22.22, 6.25, 5.93, 5.93, 0.00$</td>
</tr>
<tr>
<td>$x_{hs}$</td>
<td>$x_{hs}$</td>
</tr>
<tr>
<td>$6.25, 6.25, 7.04, 3.13, 0.00, 0.16$</td>
<td>$6.25, 6.25, 7.04, 5.93, 0.00, 0.09$</td>
</tr>
<tr>
<td>$g_{\text{sire}}$</td>
<td>$g_{\text{sire}}$</td>
</tr>
<tr>
<td>$3.13, 3.13, 3.13, 6.25, 0.00, 0.03$</td>
<td>$3.13, 3.13, 3.13, 11.86, 0.00, 0.03$</td>
</tr>
<tr>
<td>$g_{\text{dam}}$</td>
<td>$g_{\text{dam}}$</td>
</tr>
<tr>
<td>$3.13, 3.13, 3.13, 0.00, 0.00, 0.00$</td>
<td>$3.13, 3.13, 3.13, 0.00, 0.00, 0.00$</td>
</tr>
<tr>
<td>$g_{\text{mates}}$</td>
<td>$g_{\text{mates}}$</td>
</tr>
<tr>
<td>$0.00, 0.00, 0.16, 0.00, 0.00, 0.33$</td>
<td>$0.00, 0.00, 0.31, 0.00, 0.00, 0.62$</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td><strong>G</strong></td>
</tr>
<tr>
<td>$25.00, 12.50, 6.25, 3.13, 3.13, 0.00$</td>
<td>$25.00, 12.50, 6.25, 5.93, 5.93, 0.00$</td>
</tr>
<tr>
<td><strong>Acc</strong></td>
<td><strong>Acc</strong></td>
</tr>
<tr>
<td>0.6888</td>
<td>0.7024</td>
</tr>
</tbody>
</table>

**Example**

$$b = P^{-1} G$$

$$0.170, 0.391, 0.464, 0.030, 0.267, 0.223$$

$$0.169, 0.234, 0.500, 0.048, 0.298, 0.250$$

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.688</td>
</tr>
<tr>
<td>1</td>
<td>0.690</td>
</tr>
<tr>
<td>2</td>
<td>0.692</td>
</tr>
<tr>
<td>3</td>
<td>0.694</td>
</tr>
<tr>
<td>4</td>
<td>0.696</td>
</tr>
<tr>
<td>5</td>
<td>0.698</td>
</tr>
<tr>
<td>6</td>
<td>0.700</td>
</tr>
<tr>
<td>7</td>
<td>0.702</td>
</tr>
<tr>
<td>8</td>
<td>0.704</td>
</tr>
</tbody>
</table>

---

60