DESIGN AND OPTIMISATION OF ANIMAL BREEDING PROGRAMMES

Lecture notes prepared by:

Jack C.M. Dekkers
Iowa State University
Ames, USA

John P Gibson
Institute for Genetics and Bioinformatics
Armidale, Australia.

Piter Bijma and Johan A.M. van Arendonk
Animal Breeding and Genetics group
Wageningen University
Wageningen, The Netherlands

for AnS 652 A and B, S05, Iowa State University
by Jack Dekkers

Preface

The first version of the notes were written by John Gibson for a Nordic graduate course in Denmark in 1992. When Johan was asked to give a similar course in Denmark four years later, it was decided to combine the material into a joint set of lecture notes. From that point onwards, the (parts of the) notes have been used in several other courses. In recent years, Jack Dekkers has made a considerable contribution to the notes. Over time, the notes have changed which reflects the developments in the fields as well as feed back we have had in teaching animal breeding and genetics.

For the Chapter on Inbreeding, we have made use of material prepared by John Woolliams and Theo Meuwissen for an international course in 2000 in The Netherlands. We greatfully acknowledge their contribution.

We would like to thank the students attending courses in which these or an earlier version of the notes were used for making teaching an enjoyable experience and for all their comments and suggestions towards improving these notes.

Wageningen, July 2004

Jack Dekkers
Department of Animal Science
Iowa State University
239D Kildee Hall
Ames IA 50011-3150 USA
e-mail: jdekkers@iastate.edu

John Gibson
Institute for Genetics and Bioinformatics
University of New England
Armidale NSW 2351, Australia.
e-mail: john.gibson@pobox.une.edu.au

Johan van Arendonk and Piter Bijma
Animal Breeding and Genetics Group
Wageningen University
PO Box 338
6700 AH Wageningen, The Netherlands
e-mail: johan.vanarendonk@wur.nl and Piter.Bijma@wur.nl

Chapter 1

Introduction

(Based on Bijma and van Arendonk, 2004)

There are two fundamental questions faced by animal breeders. The first asks: **"What is the best animal?"** Is the best Labrador the one with show-winning conformation or the one with exceptional retrieving instinct? Is the best dairy cow the one that gives the most milk; the one with the best feet, legs and udder support; or the one that combines performance in these traits in some optimal way? These are matters of intense debate among breeders, and, in truth, no one has all the answers. The question is an important one, however, because the answers determine the desired direction of genetic change for breeding organisations and people keeping farm or companion animals. The second question asks, "**How do you breed animals so that their descendants will be, if not "best", at least better than today's animals?"**. In other words, how can we genetically improve animal populations? This question involves genetic principles and animal breeding technology, and is the subject of this course.

1. What is the best animal

"Best" is a relative term. There is no best animal for all situations. The kind of animal that works best in one environment may be quite different from the best animal under another set of circumstances.

When we describe animals, we usually characterise them either in terms of appearance or performance or some combination of both. In any case, we talk about **traits**. A trait is any observable or measurable characteristic of an animal.

Some examples of *observable* traits –traits we would normally mention in describing the appearance of an animal- are coat colour, size, muscling, leg set, udder conformation, and so on. Some examples of *measurable* traits –traits we would likely refer to in describing how an animal has performed- are body weight, daily milk production, time to run a mile, etc. There are hundreds of traits of interest in domesticated animals. Note that in none of the examples of traits mentioned above is the appearance or performance of a particular animal described. An animal may be red and weigh 343 kilograms at 1 years of age, but *red* coat colour and *343 kg* yearling weight are not the traits- the traits are simply coat colour and weaning weight. *Red* and *343 kg* are the observed categories or measured levels of performance for the traits of coat colour and yearling weight. They are the **phenotypes** for these traits.

In animal breeding, we are mainly concerned with changing animal populations genetically. From a genetic point of view, therefore, we want to know not only the most desirable phenotypes, but the most desirable **genotypes** as well. That is because an animal's genotype provides the genetic background for its phenotypes and it is the genetic material that is passed on from parents to its offspring. Summarised in an equation:

$$\mathbf{P = G + E}$$

where P represents an individual's phenotype, G represents its genotype, and E represents the **environmental effects**- the effects that external (nongenetic) factors have on an animal's performance[1]. In other words, its genotype and the environment it experiences determine an animal's phenotype.

The word *genotype* is used in several ways. We can speak of an animal's genotype in general, referring to all the genes and gene combinations that affect the array of traits of interest to us. An example used later on in this section involves a "tropically adapted" genotype. In this case, the genotype includes all the genes and gene combinations affecting heat resistance, parasite resistance, and other traits that make up tropical adaptation. This sense of the word *genotype* is generally implied in this chapter. We can also speak of an animal's genotype for a particular trait, referring to just those genes and gene combinations that affect that trait (e.g., heat resistance). Or, as we will see later in this course, we can limit the definition of genotype even further in which case it refers to a particular gene only (e.g., an animal has genotype AA for the kappa-casein gene). In any case, the genotypes of our animals' descendants are what we can change with breeding methods. Favourable changes in genotypes result in improved phenotypes.

To answer the question "What is the best animal?" we need to determine what traits are of primary importance and what genotypes are most desirable for those traits. Most breeders, if they have some experience, have an opinion about the key traits and better genotypes. A Thoroughbred breeder, for example, might describe the perfect animal as "…. fast, but with enough endurance and heart for the longer distances, and easily rated". A pig breeder version might be "…. a healthy pig with a good growth and good carcass quality." There are probably as many opinions of this sort as there are breeders and for the most part they are quite subjective. In order to develop a sense of the important traits and best genotypes in a more objective way it is important to understand the role of the genotype in the system of the farm. This means that the importance of traits will depend on the physical environment under which animals are kept, the management system as well as economic factors. If you think about it, it will become clear that a number of the components of the system will interact with each other. For example, the best preventive health program (management) depends on the kinds of pathogens in the area (physical environment) and the costs of vaccines, dewormers, etc. (economics). To determine which health program is the most cost-effective, you must have knowledge of alternative programs, local pathogens, and treatment costs and understand how treatment programs interact with these other factors to affect profitability. Similarly, the best genotype depends on the local environment, the management practises in use, and the costs of inputs and prices of animal products. To determine the best genotype, you must have knowledge of environmental, management, and economic components and understand how they interact with the genotype to affect profitability.

The genotype of domestic animals determines the degree to which the animals are suited for their function in society. The key to determining the traits of importance and optimal genotypes for those traits is a thorough analysis of the function of the animal in the entire system and an understanding of the many interactions among components of the system.

Knowledge of the function of the animal and the interactions between the genotype and other components of the system is necessary if we want to develop sensible goals for breeding programs, in other words, if we want to develop appropriate **breeding objectives**. Knowing, for

[1] This mathematical expression is oversimplified but it will do fine for the purposes of this discussion. Later on we will see that there might also be an interaction between the G and E.

example, that parasite resistance is critically important in tropical climates, breeding objectives in the Tropics emphasise traits such as tick count (a measure of tick resistance). In temperate regions, on the other hand, less emphasis is placed on parasite resistance and more emphasis is placed on other traits.

2. Population structure and breeding objective

In the process of determining the best animal, you might ask, "Best for whom?". The answer to this question depends on the function of the animal, the structure of the population and the role of the "breeder"[2] within that structure. Most populations can be thought of as having a pyramidal structure: a relatively small number of breeders at the top selling breeding stock to a larger number of multipliers who in turn sell animals to a great number of end users.

The pyramid suggests a flow of **germ plasm** – genetic material in the form of live animals, semen, or embryos – from the top down, the elite breeders producing the most advanced animals, breeders at the multiplier level replicating those animals, and end users benefiting from the genetic improvement occurring at the higher levels. Ideally, breeders at each level try to produce animals that will be in the greatest demand by their customers at the next level down, with the ultimate result that the best animal is the animal that is the most useful or profitable for the end user. *End users* can thus be defined as the individuals whose particular needs should form the basis for determining breeding objectives.

In food and fibre producing species (sheep, cattle, swine, and poultry), the end users are commercial producers. These are the persons whose primary products are commodities for public consumption. Commercial dairy farmers produce milk; commercial swine producers produce pork; commercial poultry farmers produce eggs, chicken and turkey. Commercial producers are in most cases not the end of the production chain; beyond them are the processors (dairy plant, slaughterhouses), the retailers and consumers. But the commercial producers are end users because their particular needs reflect the requirements of the entire production chain. They need animals that are physically and reproductively sound, healthy and perform efficiently in their environment. They also need animals that possess the product and performance characteristics required by the retailers and consumers. The importance of these latter characteristics should be reflected – when the market systems functions well - in the prices paid to the commercial producers for their products. In the Western world, the interest of consumers in the system of production has increased over time. This increased awareness of consumers has resulted in an increased emphasis on health and welfare traits in the breeding objective of farm animals and reduced emphasis on primary production traits (e.g. amount of milk, growth rate and litter size).

The breeding industries for recreational and companion animal species (horses, dogs, cats, etc.) differ somewhat in structure from the livestock industries. The pyramid arrangement is still present, and markets for specialised types of animals exist, but seedstock/commercial divisions are usually less clear and the end users may not be breeders at all. Consider, for example, Labrador retrievers. The end users of Labs are hunters and pet owners. These persons may or may not choose to breed their animals, and the qualities that are important to them are those that contribute to retrieving ability, companionship, health, aesthetics, or some combination of these

[2] Person answering the question

traits. Among Labrador breeders there are elite breeders and multipliers, but the term commercial producers does not really fit here because no consumable commodity like meat, eggs or milk is being produced. The various horse industries provide similar examples. End users of horses range from owners of the most valuable racing animals to causal riders to those that keep miniature horses as pets.

3. How are animal populations improved?

The purpose of animal breeding is not to genetically improve individual animals- once an individual is conceived, it is too late to change the genotype of that animal- but to improve animal **populations**, to improve future generations of animals. To this task breeders bring two basic tools: selection and mating. Both involve decision-making. In selection, it is decided which individuals become parents, how many offspring they may produce, and how long they remain in the breeding population. In mating, it is decided which of the males we have selected will be bred to which of the females we have selected.

Selection

Selection is used to make long-term genetic change in animals. It is the process that determines which individuals become parents, how many offspring they may produce, and how long they remain in the breeding population. Most of us are familiar with the term **natural selection**. Natural selection is the great evolutionary force that fuels genetic change in all living organisms. We commonly think of natural selection as affecting wild animals and plants, but in fact it affects both the wild and domestic species. All animals with lethal genetic defects, for example, are naturally selected against- they never live to become parents. Natural selection cannot be ignored but the kind of selection of primary interest in animal breeding is **artificial selection**. The idea behind selection is simply this: to let individuals with the best sets of genes reproduce so that the next generation has, on average, more desirable genes than the current generation of animals. The animals with the best sets of genes are said to have the best **breeding values**. They are –from a genetic point of view- the individuals with the greatest value as parents. In selection, we try to choose those animals with the best breeding values: the animals that will contribute the best genes to the next generation. The result of successful selection is then to genetically improve future generations of a population by increasing over time the proportion of desirable genes.

To see how selection works, consider the simplest form of selection: **phenotypic selection** or **mass selection**. In this type of selection, the performance of the individual is the only information used in making selection decisions. No attention is paid to the pedigree of the animal or the performance of its sibs (brothers and sisters) or of any progeny it may have produced. For example, if you were using phenotypic selection for weaning weight to determine whether a particular ewe lamb was to be kept for breeding, you would base your decision strictly on her own weaning weight. In practise (meaning outside of scientific laboratories), phenotypic selection in its pure form is increasingly rare, but it makes a good example, as we will also see later on during this course.

Figure 1.1 depicts phenotypic selection for increased body size in mice. The largest mice in each generation are chosen to become parents of the next generation, and the result over time is an increase over time in average body size. The idea of using the phenotype for body size as the

selection criteria is based on the expectation that phenotype for size is a reasonable indicator of the genes affecting body size. It is the genes, after all, which are transmitted from parent to offspring. In other words, it is assumed that phenotype for body size in mice is somehow related to breeding value for body size. If that were not the case, phenotypic selection for this trait would be a waste of time. The relationship between phenotype and breeding value is therefore a very important one, and this relationship is reflected by the **heritability**. When heritability of a trait is high, phenotypes are generally good indicators of underlying breeding values, and phenotypic selection will be effective in changing the level of the trait. When heritability is low, phenotypes reveal little about breeding values, and phenotypic selection will be ineffective. Judging by the rapid increase in body size of the mice in Figure 1.1, body size must be quite heritable. Not all traits are as heritable. The heritability of fertility in mammals, for example, is generally quite low. Estimating the heritability of a trait involves statistical techniques to estimate the extent to which relatives resemble each other for the trait of interest, compared with unrelated animals. The actual methodology involved and a description of the methods is beyond the scope of this course.

Most animal breeders are unlikely to limit themselves to individual performance information alone in making selection decisions. They will use information on relatives as well. For example, when a dog breeder purchases an eight-week old puppy from another breeder, she probably does not base here choice on just the conformation and personality characteristics evident in such a young puppy. She wants to evaluate those same traits in the littermates, the dam and the sire. She might want to see a copy of the puppy's extended pedigree to learn more about its ancestors. Similarly, when beef cattle breeders evaluate a sire to use via artificial insemination (A.I.) they look further than the sire's own performance for growth rate. They want to know something about the growth performance of his progeny.

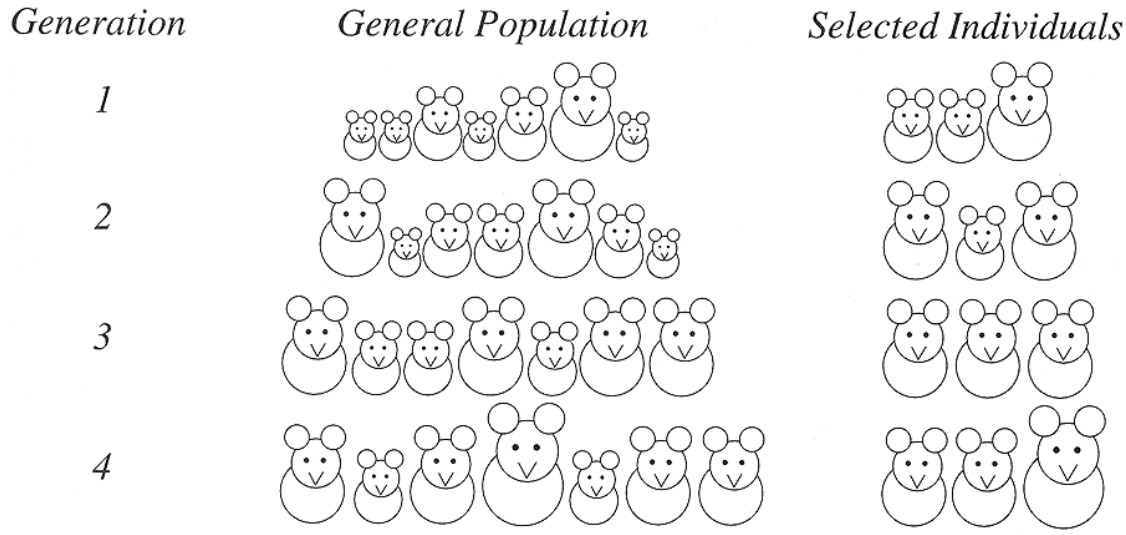


Figure 1.1. Illustration of phenotypic selection for increased body size in mice

The above examples illustrate that selection decisions are based on a combination of information. In this course we will outline how the different sources of information can be combined into a single prediction of the breeding value of the animal. The strength of the relationship between the true breeding value and its prediction is measured by the **accuracy**. When accuracy is high, predictions of breeding values will normally be good ones – they will closely reflect the

differences in true breeding values of the animals being evaluated. And because the predictions of breeding values are accurate, we can do a good job in selection.

The traits mentioned so far in this chapter – such as weaning weight in sheep, body size in mice, fertility, conformation and personality in dogs, milk production in dairy cattle- have all been **polygenic traits**. Many genes affect polygenic traits, and no single gene is thought to have an overriding influence. The genetic variation in these traits is due to segregation at many loci. Until recently, we knew little about the specific genes affecting these traits – we just know there were lots of them. As long as we cannot identify specific genes, we have to rely on phenotypic performances, predictions of breeding value to characterise the genotypes of animals. There are good grounds for believing that there is a range in the size of effects of genes for any trait, from a few with large effect, down to a large number having very small effects. We will see in this course that the developments in molecular biology now make it feasible to identify individual genes that affect quantitative trait. Information on genetic markers linked to individual genes can be used in selection programmes to improve the accuracy of selection (so-called marker-assisted selection). Once an individual gene has been identified, its biochemical and physiological roles can be studied. The results of these studies will greatly increase our understanding of the nature of genetic variation in traits.

Most traits in animals are polygenic in nature. Some traits, however, are **simply-inherited** – they are affected by a single or only a few genes. A good example is the horned/polled character in cattle of European origin (Polled means naturally without horns). A single gene determines whether a cow is horned or polled. There are also a large number of single-gene disorders that are considered to be serious problems but do not prevent affected individuals from reproducing. Well known examples include the inherited eye disorder is dogs, the malignant hyperthermia syndrome ("halothane gene") in pigs. Because only a few genes influence simply-inherited traits, selection for simply-inherited traits is different from selection for typical polygenic traits. With simply-inherited traits, we do not deal with breeding values and their predictions, or even with the concept of heritability. Rather, we are interested only in knowing whether an individual possesses the specific allele or alleles of interest, and we select animals based on that knowledge. If the disorder can be detected either by clinical examination or by DNA-testing prior to reproductive age, it is possible to select against the disorder effectively. The detection of the gene for malignant hyperthermia syndrome in pigs and the subsequent development of a DNA-test have greatly increased the opportunity for pig breeders to eliminate the disorder from the population. Malignant hyperthermia in pigs is an autosomal recessive disorder which means that it is not possible to discriminate between a phenotype of animals with two normal alleles (homozygous animals) and animals carrying one defect allele (heterozygous animals, so-called carriers). The power of the DNA-test lies in the fact that it facilitates the detection of carriers- animals that are heterozygous at the gene causing the genetic disorder- prior to reproductive age.

When we think of selection, we normally envision selection of individual animals within a breed. It is also possible to select between breeds. In setting-up a farm or breeding program, we need to choose a breed to work with. **Between-breed selection** provides a way of using breed differences to make very rapid genetic change. For many traits, breed differences can be very large. By taking advantage of such large differences, between-breed selection can produce genetic change much faster than the gradual change possible from selection within a breed. For example, the milk production of Black and White cattle in The Netherlands has increased enormously in the

1970's – not through selection within the Dutch Friesian population, but through importation of semen from the more productive Holstein-Friesian in the United States and Canada.

Mating

Selection is the first of the two basic tools used by animal breeders to make genetic change. The second tool is **mating**. Mating is the process that determines which (selected) males are bred to which (selected) females. It is distinctly different from selection. In selection, you choose the group of animals you want to be parents; in mating, you match males and females from the selected group.

There are many different methods for mating animals, and each method can be defined by a set of mating rules: a **mating system**. There are three reasons for using mating systems: (1) to produce offspring with extreme breeding value, (2) to make use of complementarity, and (3) to obtain hybrid vigour. Extreme phenotypes can be obtained by mating parents with extreme breeding values (high*high and low*low). If an animal of intermediate size is desired, mating large animals to small animals is one way to produce it. The parental genotypes are quite different, and neither one is optimal, but the mating is complementary because the offspring is optimal. Mating a Charolais to an Angus is an example of **crossbreeding**; the mating of sires of one breed to dams of another. In crossbreeding often used to produce breed complementarity, and in fact, the Charolais x Angus mating is a complementary one. Charolais are large French cattle known for their fast growth and heavy muscling, Angus are smaller British cattle known for their maternal ability, and the crossbred offspring benefit for having both kinds of parents. Another reason for crossing these two breeds is to produce **hybrid vigour** or **heterosis**. Hybrid vigour is an increase in performance of crossbred or hybrid animals over that of the pure-breds. Hybrid vigour occurs to a greater or lesser degree in many traits, but it is most noticeable in traits like fertility and survivability.

4. Multiple trait selection

In this course, a lot of the discussion of selection and the examples used for illustration will be limited to **single-trait selection**, selection for just one trait. That is because single-trait selection provides a simple framework within which to learn the principles of animal breeding. But in the real world of animal breeding, selection for a single trait is rare. Breeders are typically interested in improving a number of traits. They practise **multiple-trait selection.** Dairy farmers select for traits related to milk production, health, reproduction, type and longevity.

Selection for one trait rarely affects just that one trait. Usually other traits are affected as well. Genetic change in a trait resulting from selection on another trait is termed **correlated response to selection**. Correlated response to selection is probably caused by a number of genetic mechanisms and results in so-called genetic correlation between traits.

Genetic correlations between traits and the correlated response to selection brought about by them can be beneficial. However, if we are unaware of or choose to ignore unfavourable genetic correlations, selection for one trait can lead to undesirable response in others. In cattle, for example, blind selection for growth rate leads to larger birth weights and more dystocia. If we

want faster growth, but cannot tolerate increased dystocia, we must avoid simply selecting for growth or against dystocia. We need a way to select for growth rate and against dystocia at the same time. We need a method for multiple-trait selection as introduced in this course.

5. Inbreeding

Inbreeding is the mating of related individuals. That is the simplest definition anyway. Because all animals within a population are related to some degree, a more technically correct definition of inbreeding is the mating of individuals more closely related than average for the population. Inbreeding has a number of effects, but the chief one and the one from which all the others stem is an increase in homozygosity- an increase in the number of homozygous loci in inbred animals and an increase in the frequency of homozygote genotypes in an inbred population. Because inbred individuals have fewer heterozygous loci than non-inbreds, they cannot produce as many different kinds of gametes. The result is fewer different kinds of zygotes and therefore less variation in the offspring. This illustrates, as we will see in more detail furtheron in this course, that inbreeding (more precisely the level of inbreeding in the population) is related to the amount of genetic variation. A second consequence of inbreeding is the expression of deleterious recessive alleles with major effects, and it is this aspect of inbreeding, more than any other, that gives inbreeding a bad reputation. People associate inbreeding with genetic defects. It is true that defects caused by recessive alleles often surface in inbred populations. But inbreeding does not create deleterious recessive alleles; they must already have been present in a population. Inbreeding by itself simply increases homozygosity, and it does so without regard to whether the newly formed homozygous combinations contain dominant or recessive alleles. It therefore increases the chance of deleterious alleles becoming homozygous and expressing themselves. Expression of deleterious recessive alleles with major effects, particularly lethal genes, is a very visible consequence of inbreeding. It is an example of the effect of inbreeding can have on certain simply-inherited traits. Less obvious is the expression of unfavourable recessive alleles influencing polygenic traits. The individual effects of these genes are small but, taken together, can significantly decrease performance- a phenomenon known as **inbreeding depression**.

6. Biodiversity

An important issue arises in situations where a breed that is native to a particular area appears to have lost its function in that area or elsewhere, and consequently is in danger of becoming extinct. The question to be raised in this situation is whether such a breed should be preserved. The arguments in favour of preservation are that we do not know what type of animals will be required in the future, and that we should therefore preserve the available genetic variation between breeds (bio-diversity) as an insurance against the unknown future. On the other hand, it is argued that people who aim to earn a living from animals cannot afford to look too far into the future; they appreciate the arguments in favour of preservation, but are unable to meet the relatively high cost of preserving populations that they are unlikely ever to utilise during their own lifetimes. At both the national and international level, e.g. FAO and Rare Breeds International, concerted efforts are being made to gather relevant data on breeds that seem threatened by extinction, and to act, where possible, to save them. Interestingly, the two areas that are probably of greatest concern are at the either end of the spectrum of animal improvement. At one end we have a large variety of locally adapted native populations (often in developing countries) that are under threat from the influx of “improved” breeds and strains from developed

countries. And at the other end we have an increasing number of poultry selection lines that are discarded when yet another independent poultry breeding company is taken over by a larger and often multinational breeding company.

7. Technology and animal breeding

The face of animal breeding has changed significantly over the past decades. Animal breeding used to be in the hands of a few distinguished "breeders", individuals who seem to have specific arts and skills to "breed good animals". Nowadays, breeding in particular in livestock species is dominated by science and technology. In some livestock species, animal breeding is in the hands of a few large companies, and the role of the individual breeders seems to have decreased. There are several reasons for this change. Firstly, the breeding industry has adopted scientific principles. Looking was replaced by measuring, and an intuition was partly replaced by calculations and scientific prediction. Other major developments grew from the introduction of biotechnology.

Biotechnology can be broadly defined as the application of biological knowledge to practical needs. These technologies fall generally into two categories, reproductive and molecular. Not all of this is new. Artificial insemination was introduced in cattle in the fifties. There is no doubt that technology had a major impact on rates of genetic improvement in dairy cattle and is just as important to the structure of animal breeding programs. Nowadays, technologies like ovum pick up, in vitro fertilisation, embryo transfer, cloning of individuals, and selection with the use of DNA-information is all on the ground. Some of the technologies are already applied, others are further developed, or waiting application. Finally, rapid development of computer and information technology has greatly influenced data collection and genetic evaluation procedures in animal populations, now allowing comparison of predicted breeding values across farms, breeds or countries.

It is important to recognise that the introduction and exploitation of new technologies have large social impacts. The introduction of breeding methods typically needs to find the right balance between what is possible from a technological point of view and what is accepted by the decision makers and users within the socio-economic context of the production system. Ultimately it is the consumer who decides which technology is desirable and which is not. In most western societies, consumers are increasingly aware of health, environmental and animal welfare issues. Food safety and methods of food production are part of their buying behaviour. However, price and production efficiency are still major factors determining the sustainability of a livestock sector. Successful animal breeding programs need to find and apply the accepted technologies that help them remain competitive. This course is mostly concerned with the technical issues involved in the application of new technologies in animal breeding.

8. Components of breeding programs

Very generally, the aim of animal breeding is to genetically improve populations of livestock so that they produce more efficiently under the expected future production circumstances. Genetic improvement is achieved by selecting the best individuals of the current generation and by using them as parents of the next generation. A breeding program is the organized structure that is put into place to genetically improve livestock populations. This chapter deals with the set-up and evaluation of animal breeding programs.

Message

A breeding program is the organized structure that is set up in order to realize the desired genetic improvement of the population.

Successful genetic improvement requires breeding programs to have (at least) the following components: *i*) A system to record data on selection candidates. Without data on selection candidates it is impossible to identify the best individuals. *ii*) Methods and tools to estimate the genetic merit (breeding value) of selection candidates. This step is referred to as "breeding value estimation" or "genetic evaluation system". *iii*) A system to select the animals that become parents of the next generation, and mate them to produce the next generation. *iv*) A structure to disseminate the genetic improvement of the breeding program into the production population. In most cases, the breeding population and the production population are (partly) separated. Since the aim is to improve livestock production, genetic improvement created in the breeding population should be disseminated into the production population.

Data recording and collection*:* Estimation of breeding values requires phenotypic data on selection candidates. Thus a system has to be set up to routinely record data on selection candidates. The way data is collected depends on the species and the traits in the breeding goal. For example, the product of a dairy cattle breeding company is a straw of semen from a bull. However, milk yield cannot be recorded on bulls. Thus to identify bulls of high genetic merit for milk yield, one has to collect data on daughters of bulls. Dairy cattle breeding schemes therefore have a system to record data on daughters of test bulls. Milk yield of those daughters is recorded on common dairy herds, meaning that farmers are involved in the data recording. In beef cattle breeding, growth performance of bulls can be recorded on the selection candidates themselves, meaning that progeny testing is not necessary. In beef cattle breeding, data collection therefore takes place at testing stations where the performance of selection candidates is recorded. *The quality of the data is fundamental to the success of breeding programs.* Without high quality data, it is impossible to accurately estimate genetic parameters and breeding values.

Breeding value estimation: After data are recorded, breeding values have to be estimated. The common procedure to estimate breeding values in applied livestock breeding is called "BLUP". BLUP and selection index theory have the same theoretical basis; both are based on regression of breeding values on phenotypes. Compared to selection index theory however BLUP has the following advantages; *i*) It accounts for systematic environmental effects. *ii*) BLUP is more flexible than selection index theory and therefore more suitable as an operational tool. *iii*) BLUP takes account of selection.

Selection and mating: Selection and mating takes place after breeding values are estimated. Selection refers to the process of choosing parents to produce the next generation, whereas mating refers to the pairing of selected individuals. Thus selection precedes mating. The selection process determines the genetic improvement of the population over time, whereas the mating process determines how maternal and paternally derived alleles are combined within individuals. This chapter will introduce a number of selection and mating procedures and present theory to understand the effects of the different procedures.

Dissemination of genetic progress: In most species, the breeding and production populations are distinct. Genetic progress is created in the breeding population, but the final aim is to improve livestock production in the entire population. Thus genetic improvement created in the breeding population has to be disseminated into the production population.
In dairy cattle, the breeding and production populations are not strictly separated. Superior cows from the production population can enter the breeding population, meaning that they are selected as bull dams. Genetic progress created in the breeding program is transferred to the dairy farms by the sale of semen of progeny tested bulls to the farmers. The sale of semen is the primary source of income for dairy cattle breeding companies. In addition, a limited number of embryos from the breeding population are sold to the dairy farmers.

The situation is different in pig and poultry breeding. Pig and poultry production are based on crossbreeding systems. The breeding populations consist of purebred lines, which are mated together to produce crossbred offspring. Crossbred offspring are sold to fattening farms or egg producers. The breeding and production populations are therefore completely separated; crossbred production animals cannot enter the purebred breeding populations. Dissemination of genetic superiority of the purebred breeding populations takes place by the sale of crossbred offspring.

Message
A breeding program has the following components: *i*) a data recording system, *ii*) methods and tools for breeding value estimation, *iii*) a selection and mating system and *iv*) a structure to disseminate the genetic improvement into the production population.

9. Design and evaluation of breeding programs

Design of breeding programs: The structure of breeding programs depends on both the species and the breeding goal. The optimum design of a breeding program will differ between species with large reproductive capacity and species with small reproductive capacity, between breeding programs that aim to improve production or reproduction traits, and low heritable traits versus high heritable traits.

Judging the quality of breeding programs: Choosing the best breeding scheme among a number of alternatives requires yardsticks to measure the quality of breeding schemes. Such yardsticks can be developed only when there is a well-defined breeding goal. Given that the breeding goal is clearly defined, there are three criteria that summarize the quality of a breeding program. These are:

1. Selection response for the breeding goal traits.
2. Maintenance of genetic diversity as measured by the rate of inbreeding.
3. Costs of the breeding program.

Selection response for the breeding goal traits is the revenue of a breeding program, whereas loss of genetic diversity and financial costs are the expenses of a breeding program. Selection response, loss of genetic diversity and financial costs are expressed in different units. The problem therefore is to combine them into a single criterion for the quality of a breeding program.

A comparison of breeding schemes based on selection response and the rate of inbreeding can be done as follows. To avoid long-term loss of genetic diversity an upper limit can be set to the rate of inbreeding. Next, alternative breeding schemes can be judged by comparing their selection response at the same rate of inbreeding. The scheme with the highest selection response at the same rate of inbreeding (e.g. 1%/generation) is the best scheme.
It is more difficult to combine selection response and cost into a single criterion. The question is whether the revenues from an increase in selection response, for example in the form of increased market share, makes up for the cost of increased selection response. Hence, this is not a genetic issue but primarily a commercial and business issue.

Evaluation of breeding programs: Once a breeding program is operational it is essential to routinely evaluate the results. Evaluation may consist of comparing realized genetic improvement and rates of inbreeding with values expected when designing the breeding program. When there are clear differences between expected and realized selection response and inbreeding, then one needs to find the causes of those discrepancies and if possible improve the breeding program. Reasons that breeding programs do not yield the expected genetic improvement are: *i*) the use of inappropriate models for breeding value estimation, for example when the models do not include systematic environmental effects that are present in the data; *ii*) overestimation of the genetic parameters (e.g. h^2) resulting in biased EBVs and overprediction of the expected response; iii) preferential treatment among selection candidates resulting in selection of individuals that received "good treatment" instead of genetically superior individuals, and *iv*) unexpected correlated response in other traits.

Message
The quality of alternative breeding schemes can be judged by comparing selection response, rate of inbreeding and costs of the alternatives.

Methods to design and evaluate breeding programs: To compare alternative breeding programs we need methods to quantify expected rates genetic improvement and inbreeding of the alternatives. In other words, we need methods to predict rates of gain and inbreeding of breeding programs. From a methodological point of view, quantifying the expected rates of gain and inbreeding can be done in two manners, either stochastically or deterministically. Stochastic simulation is often the easiest way, but in most cases deterministic simulation gives more insight.
With stochastic simulation, the breeding program is simulated in detail on a computer. Stochastic simulation consists of the following cycle. 1. Breeding values and phenotypes of individuals in the base generation are simulated. 2. Breeding values are estimated for the base generation animals by performing BLUP analyses on their simulated phenotypes. 3. Based on the estimated breeding values coming from the BLUP analyses, a number of animals is selected to become parents of the next generation. 4. The selected animals are mated and offspring from the matings are simulated. Next, steps 2, 3 and 4 are repeated until the desired number of generations is simulated. Because in stochastic simulation we simulate an entire population of "real" animals, the rates of gain and inbreeding can simply be estimated from the simulated data. Hence, after simulating the breeding scheme, the next step is to analyze the simulated data to quantify the rate of gain and inbreeding of the breeding scheme. Multiple replicates of the population are simulated, and the rates of gain and inbreeding are averaged over replicates.

The advantage of stochastic simulation is that one can mimic the true breeding program in detail, because the individual animal is simulated. Hence, stochastic simulation can be very precise. However, there are two disadvantages related to the use of stochastic simulation to evaluate breeding schemes. First, stochastic simulation is time consuming, particularly when large populations are simulated. Even with modern computers, simulation of a sufficient number of replicates of a large breeding scheme may take several hours or even days. Hence, stochastic simulation is less suited as an operational tool to quickly evaluate a number of alternatives. Second, with stochastic simulation the user often does not gain as much insight into the breeding scheme as with deterministic simulation. For example, with stochastic simulation the user will observe that shorter generation intervals generally go together with higher gain, but the deterministic equation $\Delta G = ir_{IH}\sigma_H/L$ directly shows that gain is inversely proportional to the generation interval. Hence, because stochastic simulation does not explicitly model mechanisms like accuracy, generation interval, etc, it can be difficult to extend the result to other breeding schemes that were not simulated themselves.

Instead of using stochastic simulation, one can use deterministic methods to quantify expected gain and inbreeding from alternative breeding schemes. Deterministic methods do not mimic the breeding program on the individual animal level, but use (deterministic) equations to predict gain and inbreeding. For example, prediction of the rate of gain by using the expression that $\Delta G = ir_{IH}\sigma_H/L$ is a deterministic methods. Hence, modeling the mechanisms that determine gain and inbreeding as mathematical equations allows us to quantify the expected outcome of a breeding program. To use deterministic methods one needs to know/derive the mechanisms determining gain and inbreeding; it requires more insight into quantitative population genetics than stochastic simulation.

Advantages of deterministic methods are 1). It takes limited computation time, so that many alternatives can be compared within limited time, and 2). Because the mechanisms are modeled explicitly, it gives a lot of insight into gain and inbreeding in breeding programs. In some cases, however, it may be difficult to derive accurate deterministic methods. Hence, there is a risk that deterministic methods are not precise if they do not properly model the mechanisms determining gain and inbreeding in populations. In complicated cases, stochastic simulation may be used to check the accuracy and validity of the deterministic equations. Hence, stochastic simulation may be helpful to validate and improve deterministic models, and in this way we improve our understanding of the mechanisms determining genetic improvement and rates of inbreeding in populations.

In this course we will mainly deal with deterministic models. The reason is that for many important situations deterministic methods are available and they provide more insight than stochastic models.

Message

The expected selection response and inbreeding of breeding schemes can be determined by using either stochastic simulation or deterministic methods. Deterministic methods provide more insight and are computationally fast. Stochastic simulation is precise and useful to validate deterministic methods.

10. This course

The course "Animal breeding strategies" introduces the quantitative genetic principles underlying the design and implementation of genetic improvement programs in livestock species. Those principals will also apply to companion animals, populations of endangered breeds and zoo populations. The basic quantitative genetic principles used in this course are handled in Falconer and Mackay (1996)[3].

The lectures start with a general overview of the field. This course focuses on the definition of breeding objectives and the genetic evaluation of breeding strategies. To achieve this, much of the course is devoted to the general principles involved in deriving economic weights of the various traits that might be genetically improved, making selection decisions between animals, designing breeding strategies and determining which strategies will make optimum progress. What is presented is a selection of some of the more common tools used in defining breeding objectives and designing and evaluating breeding strategies. These tools should be adequate to tackle many basic practical problems in animal breeding and provide background to using more complex methods.

Estimation of breeding values using best linear unbiased prediction (BLUP) is an important element in animal breeding but this lies outside the scope of this course (see AnS562). Attention will be paid to selection index theory but the emphasis lies on prediction of genetic gain and not on genetic evaluation of animals.

Lecture notes provide students with detailed knowledge on issues related to the design of breeding programmes for farm animals. Lectures will 'guide' the student through these notes. In addition problems will be supplied.

[3] Falconer D.S. and T.F.C. Mackay, 1996. Introduction to quantitative genetics. Longman fourth edition.

Genetic andEconomic Aspects of Animal Breeding Strategies

Chapter I Introduction

- Building blocks and structure of animal breeding programs
- Methods to model/evaluate breeding programs/strategies
- Design of breeding programs/strategies
- Marker-assisted Selection

What Do We Mean By Breeding Strategies?

- Tactics designed to integrate new technologies and to improve old ones, for the purpose of maximizing performance of existing stock (Charles Smith)
- Integration of the components of a breeding program into a structured system for genetic improvement, with the aim to maximize an overall objective

General aim for animal breeding strategies:

Obtain future generations of animals that will produce more efficiently under future production circumstances

Basic Principle of making genetic progress in a population

Mate the "best" to the "best" and do that as quickly as possible.

$$\text{Genetic Gain / Yr} = \frac{\text{genetic superiority selected parents}}{\text{generation interval}} = \frac{\text{intensity} \times \text{accuracy} \times \text{genetic st. dev.}}{\text{generation interval}}$$

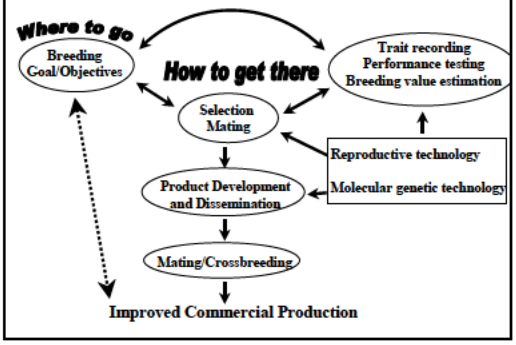
Mate the "Best" to the "Best" and Do That As Quickly As Possible

Some Questions

- How to find/identify the "best"?
- "Best" for what?
- What are the limits to use of only the "best"?
- How can we shorten the generation interval?
- What are the limits?
- How can a breeding company make a profit from this?
  - "Breeding is a business" Lush, 1945
- How do technologies enter into this?

BREEDING STRATEGIES

Basic Components of a Successful Breeding Program/Strategy



Basic Components of a Successful Breeding Strategy

- Breeding Goal or Objectives - where should we go?
  - Which traits must be improved? - Economic traits
  - How important is each trait? - Economic values
    - Focus on improvement of Economic efficiency/profit
    - Consider (future) consumer demands
- Trait recording, Performance testing, Breeding value estimation

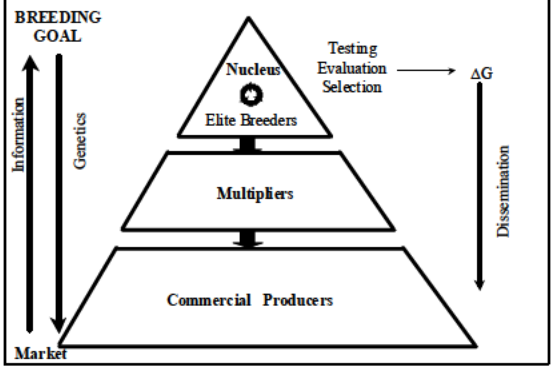
  Identify animals with "best" genetics - relative to breeding goal
  - trait performance recording and testing programs
    - which traits should be recorded and on which animals?
      - field recording
      - performance test stations / nucleus herds
      - progeny testing
  - pedigree registration

Genetic Evaluation ↔ Selection Index (Total merit index)

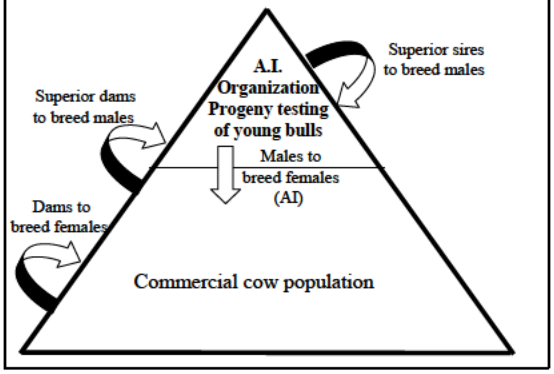
Components of a Successful Breeding Strategy (cont'd)

- Selection and mating
  - use best animals to breed next generation → genetic improvement
    - How many and which animals should we select?
    - How should we mate them?
    - Should reproductive technology be used to increase # progeny/parent?
    - balancing rate of genetic gain and inbreeding (and cost)
- Product Development and Dissemination
  - program for marketing and distribution of superior genes into the commercial sector
    - progeny testing, AI
    - multipliers
- Mating/Crossbreeding
  - optimize combinations of genetic material in commercial animals

Typical Structure of Livestock Breeding Programs



Dairy Cattle Progeny Testing Program



Basic steps in the design of breeding programs (Harris '84)

1) Describe the production system(s)
2) Formulate the objective -simplified and comprehensive- of the system
3) Choose a breeding system and breeds
4) Estimate selection parameters and (discounted) economic values
5) Design an animal evaluation system
6) Develop selection criteria
7) Design matings for selected animals
8) Design a system for expansion - dissemination - of genetic superiority
9) Compare alternative programs

Development of Breeding Strategies Summary

- Integration of the components of a breeding program into a structured system for genetic improvement, with the aim to maximize an overall objective (genetic gain, market share).
- Evaluate opportunities for improving upon current strategies.
- Evaluate the potential of new technologies.
  - How can they best be incorporated into current strategies?
  - Can their benefits best be capitalized on in a redesigned breeding structure?

Breeding Strategies - Summary

What tools are necessary to develop optimum strategies?

- Quantitative genetics theory
  - Predicting response to selection, selection index, inbreeding, etc
- Systems analysis
  - Predicting and optimizing response in overall objective
- Common sense
- An open mind

Chapter 2

Stochastic Methods to Model Breeding Programs

2.1 Introduction

The objective of genetic improvement of livestock is to enhance the genetic level for traits of interest in a population through genetic selection such that some overall goal is achieved or enhanced. The overall goal can usually be described in economic terms (e.g. maximize profit per animal per year) and will be discussed further in chapter 7.

There are many factors that determine the success of a breeding program. These include design and implementation issues. In this course, we will primarily focus on factors related to the design of genetic improvement programs, which include factors such as population size, numbers of animals to select, criteria for selection, etc.. Because of the number of factors involved, the number of alternative programs is numerous. However, ultimately only one program can be implemented; animal breeders don't have the luxury of trying out different options and then deciding which one to go with. Thus, we need some other means of deciding *a priori* which breeding program will maximize our overall objective. This requires the ability to model breeding programs and to predict outcomes from alternative breeding programs. Furthermore, if a good understanding can be developed of the impact alternative design factors have on program outcomes, this will lead to the development and choice of better breeding programs. The development of this knowledge and associated methods and tools are the focus of this course.

2.2 Quantitative Genetic Model

Because most traits of interest in livestock are multifactorial in nature, i.e. affected by a potentially large number of individual genes along with environmental factors, quantitative genetic theory has become the primary basis for the development of methods to develop, model, and evaluate alternative breeding programs. The basis of this theory is the infinitesimal genetic model (Falconer and Mackay, 1996). The purpose of this section is to briefly review this theory as a basis for developing methods to model breeding programs.

The quantitative genetic model for the phenotype of animal i is: $y_i = \mu + g_i + e_i$ (2.1)

where μ is an overall mean (or sum of fixed effects), g_i is the animal's genetic value, and e_i it's random environmental effect. For the purposes of the majority of this course, we will assume we are dealing with additive traits such that g_i refers to the additive genetic or breeding value.

Variables g_i and e_i are assumed normally distributed with means zero and standard deviations σ_g and σ_e. Strictly, these assumptions hold for g_i only for an unselected (base) population and both the mean and variance will change as a result of selection, as will be described later on in the course.

With the exception of the sex chromosomes, which we will ignore for the moment, all animals carry two copies of every gene. One copy is inherited by random sampling from the two copies carried by the male parent (sire) and the other copy is inherited by random sampling from the two copies carried by the female parent (dam). It follows that the additive genetic value of an offspring, g_o, can be partitioned into three sources, and modeled as follows:

$$g_o = ½\, g_s + ½\, g_d + g_m \tag{2.2}$$

where g_s and g_d are the additive genetic values of the sire and dam and g_m is the Mendelian sampling contribution. The Mendelian sampling contribution reflects the random selection of copies of parental genes. Since genes are inherited at random from the parents, the average values of g_m over a large number of progeny is expected to be zero.

Mathematically, it is said that the expectation of g_m, $E(g_m)$, is zero. But for any particular individual, g_m has a real value which varies between individuals. The range of values of g_m is determined by its variance, which in the absence of inbreeding, is expected to be

$$E(\boldsymbol{\sigma}^2_{g_m}) = ½\, \boldsymbol{\sigma}^2_{g_0} \tag{2.3}$$

where $\boldsymbol{\sigma}^2_{g_0}$ is the initial genetic variance in the population prior to any selection. The reason for noting the requirement that there be no prior selection in the population will become clear later in the course.

With inbreeding, the expected variance of Mendelian sampling terms is reduced by a factor $[1- ½(F_s+F_d)]$, where F_s and F_d are the inbreeding coefficients of the sire and dam. Thus:

$$E(\boldsymbol{\sigma}^2_{g_m}) = ½\, [1- ½(F_s+F_d)]\, \boldsymbol{\sigma}^2_{g_0} \tag{2.4}$$

2.3 Stochastic Models for Evaluation of Breeding Programs

The simple quantitative genetic models described in the previous paragraph can be used to simulate a breeding program and evaluate its outcomes. Simulations in animal breeding can be divided into three types:

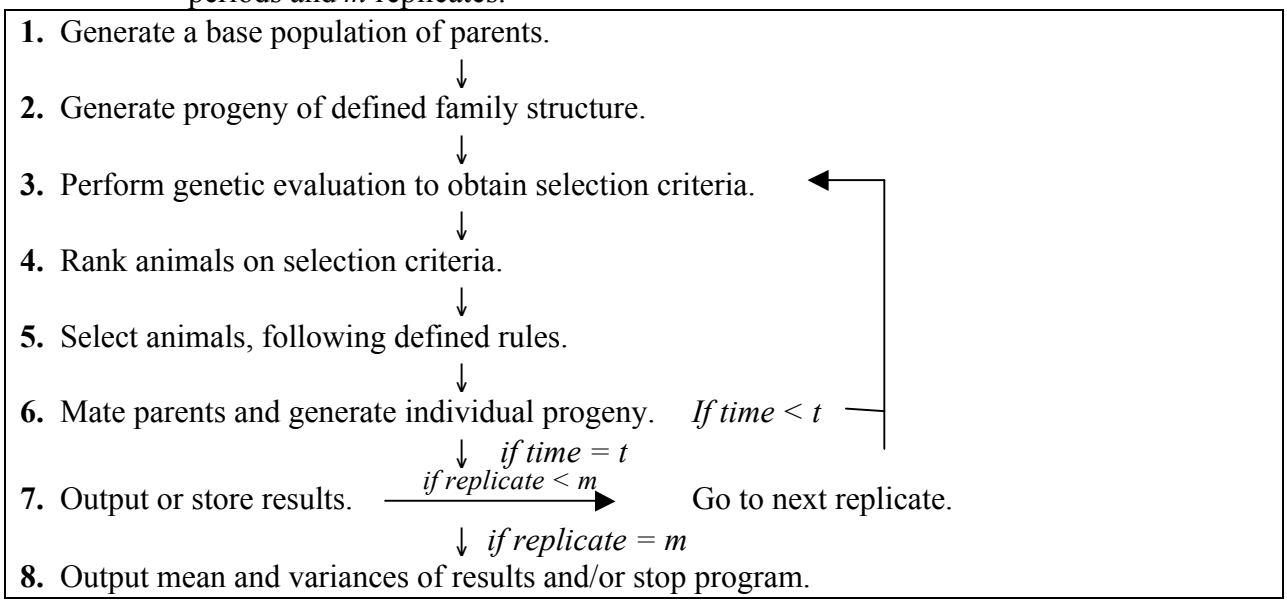
1) stochastic simulation (or sometimes called Monte Carlo simulation)
2) deterministic simulation
3) combination of stochastic and deterministic simulation.

Stochastic simulations use random number generators to simulate variability. The two most common types of random generators needed are those for the uniform and the normal distribution. Most statistical software programs have functions that can generate these. Excel has a uniform random number generator: RAND(), which returns a uniform number between 0 an 1.

Using the Inverse Transform method (http://www.mathwave.com/articles/random-numbers-excel-worksheets.html) this function can be used in combination with inverse cumulative distribution functions to generate numbers from other distributions in Excel. For example, to generate a random number from a standard normal distribution, use: NORMINV(RAND(),0,1). The function NORMINV(p, mean, st.dev.) returns the truncation point for a normal distribution that has a fraction p below it. So by drawing p from a random uniform distribution (0,1), a random truncation point is generated based on the cumulative distribution function.

With stochastic simulations in animal breeding, which will be described here, a population of animals is simulated by generating records for each animal in the population by random sampling from pre-defined distributions which are determined by the rules of inheritance and origins of environmental effects imposed on the model. A model for stochastic simulation of a breeding program is schematically represented in Figure 2.1. The steps involved are described in further detail in what follows.

Figure 2.1 General schematic of a stochastic simulation of a breeding program with *t* time periods and *m* replicates.



2.3.1 Generating Base Population Parents

A base population is generated according to the rules of inheritance and structure of the population defined by the program control variables. For example, if the phenotype of a single trait, explained by the simple additive inheritance model plus a random environment effect, is

$$y_i = \mu + g_i + e_i$$

and there are n_m males and n_f females in the base assumed to be randomly selected, unrelated, and non-inbred, then the effects for an animal in the base population could be defined by the following programming steps:

Do for each animal i:

1. *r = random number from normal distribution with mean 0 and variance 1*
2. $g_i = r * \sigma_{g_o}$ *where* σ_{g_o} *is the additive genetic st. deviation in the base population.*
3. *r = new random number from normal distribution with mean 0 and variance 1*
4. $e_i = r * \sigma_e$ *where* σ_e *is the standard deviation of environmental effects.*
5. $p_i = \mu + g_i + e_i$, where μ is the pre-defined population mean, i.e. a constant.
6. *Store* p_i, g_i; *and* e_i

This can be repeated for all animals in the base population. In order to enable the construction of a pedigree file, animals should be given a unique identification number. The simulation can be extended to include other genetic effects, such as dominance or systematic environmental effects such as age, herd or year. Virtually all programming languages have a random number generator or an associated library of subroutines containing a routine for random number generation.

2.3.2 Generating Progeny

Once parents are generated, mating pairs are allocated and progeny generated. Recalling from equation 2.1 that the phenotype of progeny *k* of male parent *i* and female parent *j* is

$$y_{ijk} = \mu + \tfrac{1}{2} g_{s_i} + \tfrac{1}{2} g_{d_j} + g_{m_{ijk}} + e_{ijk} \qquad (2.5)$$

where g_{s_i} and g_{d_j} are the known additive genetic values of the sire and dam, $g_{m_{ijk}}$ is the Mendelian sampling contribution for individual *k* and e_{ijk} is the environmental effect. The contributions of $g_{m_{ijk}}$ and e_{ijk} are obtained for each progeny in turn by sampling from a random normal distribution with mean *0* and variance 1 and multiplying the random number by σ_{g_m} or σ_e, where $\sigma^2_{g_m} = \tfrac{1}{2}\sigma^2_{g_o}$ in the absence of inbreeding, or $\sigma^2_{g_m} = \tfrac{1}{2}(1 - \tfrac{1}{2}F_{s_i} - \tfrac{1}{2}F_{d_j})\sigma^2_{g_o}$ in the presence of inbreeding, where F_{s_i} and F_{d_j} are the inbreeding coefficients of the two parents. Fixed effects can then be added to p_{ijk} according to the structure specified by the design.

2.3.3 Deriving the Selection Criterion

The selection criterion, such as the phenotypic record, a selection index, or BLUP evaluation, would be estimated for each simulated animal as if in real life. A subroutine of the program would be written to perform the evaluations. The nature of the selection criterion will determine the amount of data to be stored. For example, a selection index involving only collateral relatives would not require the parental records to have been stored, whereas animal model BLUP evaluation would require all animals and relationships back to the base population to be stored. In contrast to selection indexes, BLUP evaluation will be expensive for computing time because of the iterative nature.

Selection index or BLUP requires defined variances of traits for single trait evaluation and variance/covariance matrixes for multiple traits. Usually these would be set to the base population values, though false values may be given deliberately if estimation of sensitivity to parameter for BLUP is under investigation. If relationships back to the base generation are included, BLUP automatically allows for change in genetic variance due to selection (see Chapter 5).

With selection indexes, the appropriate variance/covariance among traits and relatives at each generation are required. A decision will therefore have to be taken as to whether to use constant parameters over time or to allow them to change. When the same set of parameters is used over time it seems logical to use the parameters from the base population, which were also used in simulating the data. In real life, the base population parameters can only be estimated and it might therefore be interesting to investigate the consequences of using other than the true parameters. Population parameters will change over time as a result of selection. These changes can be allowed for in constructing the selection index. In that case a method is needed to obtain the parameters at each point in time. The parameters could be estimated from the phenotypic and the true additive genetic values (g_{ijk}, g_{s_i}, g_{d_j}). This, however, would not be possible in real life and hence would not give realistic results. Alternatively, parameters could be estimated using phenotypic records or changes in parameters could be predicted from the selection strategy. Interpretation of the results will obviously depend on the assumptions made.

2.3.4 Selecting and Mating Animals for Breeding

In order to produce the next generation of offspring, one needs to define the method of selecting the animals to be used as parents and the procedure used in mating the selected parents. In the previous step, the selection criterion has been estimated for all candidates for selection. Truncation selection is commonly used for selection, in which the animals with the highest value for the selection criteria are selected. This requires that males and females are separately ranked in order of merit for the selection criteria. Efficient ranking routines are available in most language libraries. Apart from the method of selection, the user has to specify the number of animals to be selected and the category of animals, which are eligible for selection. One might, for example, restrict the selection to animals of one particular age class only or have no restriction other than that animals need to be old enough to be able to reproduce. In the latter case, selection will be across age groups and it is important to specify up to what age animals are eligible for selection.

In the absence of restrictions on selection, selection is simply a process of designating the required number of top ranking animals as parents. With complete assortative mating, the top ranked male is allocated to the n top ranked females, the second ranked male to the next n females and so on; where n is the number of females per male. With random mating, each selected female is allocated a random deviate, and the females are then ranked on the random deviate and mating proceeds as above.

An advantage of stochastic simulation is that restrictions can be imposed on selection and mating. Common examples would be restrictions defining the maximum number of full and half

sibs that can be selected as parents, and restrictions that full and half sibs may not be mated together. The imposition of restrictions may make some animals ineligible for mating so that more animals must be available for mating than indicated by the defined proportions to be selected.

2.3.5 Inbreeding Coefficients

Traditional methods of estimating inbreeding coefficients of individual animals by tracing path coefficients, or directly from a complete relationship matrix rapidly become time consuming and expensive of storage space as population sizes and number of generation's increase. With this method it was often impractical to estimate inbreeding coefficients in stochastic simulations. Several algorithms have been developed, however, for efficiently deriving inbreeding coefficients from a pedigree file (e.g. Tier, 1990). Use of these algorithms reduces computer time 10-100 fold compared to traditional methods. An additional trick is to recognize that all full sibs have the same inbreeding coefficient so that only one member of the family needs to have the coefficient estimated. Even so, calculation of inbreeding coefficients can still be expensive of computing time when simulating several thousand animals in each of several generations.

2.3.6 Completing the Cycle

Once mating pairs are allocated, progeny can be produced and the cycles repeated until the desired number of time periods has been achieved. At this point, summary statistics can be printed or stored, and the next replicate started. The number of replicates required will depend principally on the required accuracy of estimates of response and variance of response, which are largely dependent on the size of the population and the number of generations simulated. Large populations have low variance of response and therefore require fewer replicates for a given level of accuracy.

Stochastic simulations are often used to validate deterministic simulations. In this case it is desirable to have very accurate estimates of output parameters to estimate biases in the deterministic program. Typically, with smaller populations, several hundreds to 1000 replicates are run. But when using stochastic simulations to evaluate alternative breeding programs, very small differences between alternatives are rarely of practical interest so that often fewer, say 100, replicates can suffice. In practice the number of replicates required can be determined once a few initial runs have indicated the variance to be expected between runs for a particular size and type of population.

2.3.7 Multiple Trait Simulations

Multiple trait simulations are a little more difficult because they require simulation of correlated random variables. For Excel, a user-defined function is available from http://homepage2.nifty.com/hashimoto-t/misc/mnormrand-e.html#download that allows you to generate correlated random variables based on a defined vector of means and a variance-covariance matrix. See Excel file mnormrand.xls .

Alternatively, simulation of correlated random variables can be achieved by deriving the *n* uncorrelated principal components of the genetic and environmental variance covariance matrix among the *n* traits, generating random deviates for each principal component in turn and then back-transforming these to obtain random deviates for the original traits. Alternatively, an approach using Cholesky decomposition of the original variance covariance matrixes can be used which has advantages in terms of computing ease and time. The Cholesky decomposition approach is explained in Appendix C and some examples of simulating correlated traits and records for related individuals are given by Van Vleck (1993). These same methods can deal with simulations involving other covariances among random variables, such as *g* x *e* covariance and additive x dominance genetic covariances.

2.3.8 Genome-level models

In the previous, the genetic component was modeled as a normally distributed variable, using the infinitesimal genetic model. This model assumes that the trait is affected by a large number of unlinked loci, each of small effect. Stochastic models also allow the modeling of a more realistic genetic architecture of the trait by simulating individual loci and their placement on chromosomes within the genome, along with genetic markers. These models require specification of the number of loci, the number and length of chromosomes that these loci are located on, and their position (in centi-Morgans, cM) on these chromosomes. Then, the following parameters must be specified for each locus:

1) Locus position - loci could be positioned on chromosomes at random by sampling from a uniform distribution, or evenly distributed across the genome.
2) Number of alleles.
3) Allele frequencies in the base population – these could be set to be equal or sampled from some distribution
4) Genotypic effects associated with each genotype - these can, for example, for a locus with two alleles B, b, be based on the standard single locus genetic model with genotypic values of $+a_l$, d_l, and $-a_l$ for genotypes BB, Bb, and bb at locus *l* (Falconer and MacKay, 1996). Genotypic values assigned to each locus could be sampled from an assumed distribution of gene effects, such as from a gamma distribution (e.g. Hayes and Goddard, 2003), in an attempt to reflect reality. In addition, epistatic effects could be allowed for by assigning genotypic effects to combinations of genotypes at multiple loci.

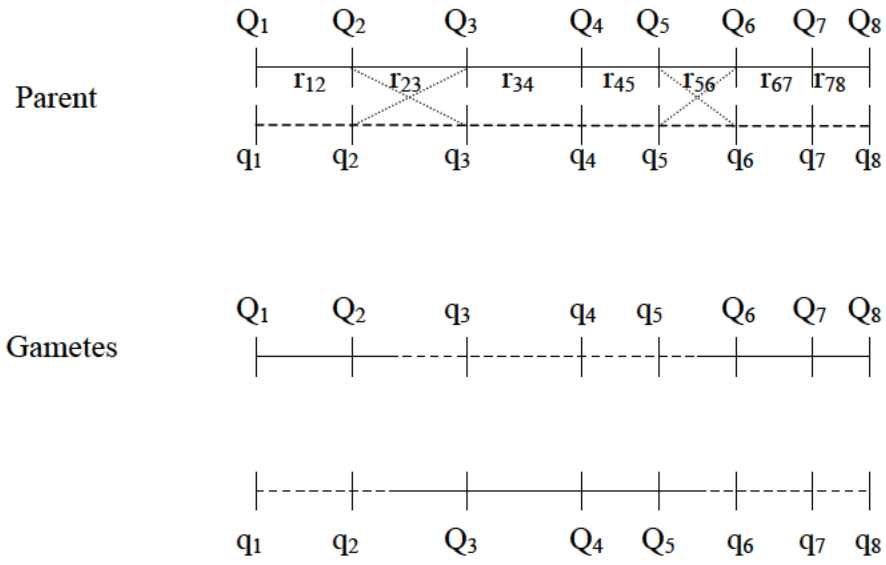
For the base population, alleles at locus *l* for individual *i* can then be assigned by drawing two random numbers *u* from a uniform (0,1) distribution. For example, for a locus *l* with allele frequency f_j^l for alleles B_j (j=1, . . , m_l), allele *j* is assigned if $\sum_{k=1}^{j-1} f_k^l < u < \sum_{k=1}^{j} f_k^l$. This random sampling of alleles assumes the base population is in Hardy-Weinberg and gametic phase equilibrium (Falconer and MacKay, 1996).

The genetic value of individual *i* then is the sum of genetic effects at each of the *q* loci:

$g_i = \sum_{l=1}^{q} g_i^l$, where g_i^l is the genotypic value at locus *l* for individual *i*, which is based on the simulated genotype of for locus *i* and the genotypic value that is associated with this genotype.

If all loci are unlinked, progeny genotypes at each locus can be simulated by randomly drawing one of the two alleles of the sire and one of the two alleles of the dam. If loci are linked, recombination must be allowed for. Consider the two haplotypes for a parent in Figure 2.1.

Figure 2.1. Simulation of Mendelian inheritance with linked loci with recombination in intervals 23 and 56.



To create a progeny from this parent, the first step is to simulate the production of two gametic chromosomes through meiosis. This can be simulated as follows

1) Starting with the first interval, 12, the probability of recombination (r_{12}) or not ($1-r_{12}$) is drawn from a uniform normal distribution. If $u[0,1] < r_{12}$, then a recombination takes place and we end up with the following two recombinant haplotypes: $Q_1,q_2,q_3,q_4,q_5,q_6,q_7,q_8$ and $q_1,Q_2,Q_3,Q_4,Q_5,Q_6,Q_7,Q_8$, since all alleles downstream from the cross-over are switched. If $u[0,1] > r_{12}$ then the parental chromosomes stay intact.

2) Proceed to the next interval and draw presence or absence of a recombination event in that interval: if $u[0,1] < r_{23}$ then there is a recombination event and we end up with the following two recombinant haplotypes (assuming there also was recombination in interval 12): $Q_1,q_2,Q_3,Q_4,Q_5,Q_6,Q_7,Q_8$ and $q_1,Q_2,q_3,q_4,q_5,q_6,q_7,q_8$. If there is no recombination event, then the haplotypes generated in step 1 remain intact.

3) Proceed through all intervals consecutively as described above.

Once a pair of recombinant gametes has been created, a random one of the two is sampled to generate the progeny. A similar procedure is used to generate the other parental chromosome.

Note that this method assumes that recombination events in adjacent intervals are independent (no interference – Haldane mapping function). If there is interference, probabilities of recombination in interval *i* must be adapted, depending on presence or absence of a recombination event in interval *i*-1.

Simulation of genomic selection programs or data for genome-wide association analysis also requires simulation of historical generations of the population, in order to generate linkage disequilibrium between loci. A useful freely available software program for this purpose is QMSim (http://www.aps.uoguelph.ca/~msargol/qmsim/QMSim_documentation.pdf Sarargolzaei and Schenkel, 2009, University of Guelph). After download, you can run this program from command line, using ./QMSim [parameterfile] -o The download provides several example input parameter files.

2.4 Advantages and Disadvantages of Stochastic Models

Stochastic simulation depends on relatively simple rules determining inheritance from one generation to the next, along with description of the criteria on which all animals will be selected for breeding. Thus, for a given degree of complexity of the breeding program, stochastic simulations are often relatively easy to write compared to the deterministic models that will be described later. In addition, stochastic models allow alternative genetic models to be evaluated, while deterministic models are primarily restricted to the infinitesimal genetic model. However, see Chapter 12 for deterministic models with individual genes along with an infinitesimal polygenic component.

With stochastic simulation, the result of any one run reflects random sampling events so that to obtain the mean expected response, many replicate runs must be made; but this also allows the variance of the response to be estimated. Because each animal in the population is individually identified, stochastic programs can take up a large amount of storage space and involve a very large number of mathematical operations for every run. This, combined with the need to replicate, means that stochastic programs take much longer, often very much longer, to run than deterministic programs.

Stochastic simulation also does not provide much insight into the impact of various factors on response to selection and does not lend itself easily to optimization of breeding programs. Hence, in the remainder of this course, the main focus will be on deterministic models, to facilitate an understanding of the factors that affect the outcomes of breeding programs. With the tremendous increases in computing power, however, stochastic models have become more and more attractive and used for the evaluation and analysis of breeding programs in both research and practice.

Chapter 3

Basic Principles of Response to Selection

3.1 Introduction

When comparing different breeding programs the first question usually asked is "what are the expected responses to selection of the various plans". A considerable part of this course will focus on methods of designing breeding programs, which maximize response to selection. Although breeding plans are often quite complex, most can usually be understood in terms of a few simple principles of response to selection. In this chapter we briefly review these principles as a foundation for what follows in the rest of the course.

As in many fields of science, there are often many different ways of deriving a particular result. If you are familiar with the basic principles of quantitative genetics (e.g. as in Falconer and Mackay, 1996), the results given here should be familiar to you. However, the approach used here is slightly different to that given in other texts. You should be familiar with the derivations given in texts such as Falconer and Mackay (1996), as those derivations are generally more rigorous and go back to first principles. However, the derivations given in this course will often be more useful when it comes to designing breeding strategies and deriving statistics necessary for such designs.

3.2 Predicting Genetic Merit of Progeny

The basic guiding principle behind genetic improvement and predicting response to selection is that parents with high additive genetic values (breeding values) tend to have progeny with high additive genetic values (and therefore high phenotypes). This follows from the quantitative genetic model for the additive genetic value of progeny:

$$g_o = \tfrac{1}{2}g_s + \tfrac{1}{2}g_d + g_m \qquad (3.1)$$

where g_s and g_d are the additive genetic values of the sire and dam and g_m is the Mendelian sampling contribution, as described in the previous chapter.

Since $E(g_m) = 0$, the expectation of the progeny additive genetic value, $E(g_i)$, from a given pair of parents is given by

$$E(g_o) = \tfrac{1}{2}g_s + \tfrac{1}{2}g_d$$

i.e., the expected additive genetic value of the progeny is equal to the mean additive genetic value of the two parents.

For determining response to selection, we are interested in the mean of the genetic value of the progeny generation, $E(\bar{g}_o)$. This can be obtained from the average genetic value of the selected parents $\bar{g}^*_s$ and $\bar{g}^*_d$, where * indicates that the variable refers to selected individuals:

$$E(\bar{g}_o) = \tfrac{1}{2}\bar{g}^*_s + \tfrac{1}{2}\bar{g}^*_d \tag{3.2}$$

For the purpose of understanding and predicting response to selection, it is useful to express the mean genetic value of selected parents in terms of a deviation from the mean genetic value of all individuals from which they were selected ($\bar{g}_s$ and $\bar{g}_d$):

Thus:

$$\begin{aligned} E(\bar{g}_o) &= \tfrac{1}{2}(\bar{g}^*_s - \bar{g}_s + \bar{g}_s) + \tfrac{1}{2}(\bar{g}^*_d - \bar{g}_d + \bar{g}_d) \\ &= \tfrac{1}{2}(\bar{g}_s + \bar{g}^*_s - \bar{g}_s) + \tfrac{1}{2}(\bar{g}_d + \bar{g}^*_d - \bar{g}_d) \\ &= \tfrac{1}{2}(\bar{g}_s + S_s) + \tfrac{1}{2}(\bar{g}_d + S_d) \\ &= \tfrac{1}{2}(\bar{g}_s + \bar{g}_d) + \tfrac{1}{2}(S_s + S_d) \end{aligned} \tag{3.3}$$

Here, S is the genetic superiority of the selected parents, which is defined as the difference between the mean genetic value of the selected individuals from the mean of the group they were selected from, e.g.:

$$S_s = \bar{g}^*_s - \bar{g}_s \tag{3.4}$$

Response to selection is defined as the difference of the mean genetic value of progeny of selected parents from the mean genetic value of progeny of all possible parents. Response is often denoted as R or Δg. Using the R notation, the expectation of R is given by:

$$E(R) = \bar{g}_o - \bar{g}_p \tag{3.5}$$

Where

$$\bar{g}_p = \tfrac{1}{2}(\bar{g}_s + \bar{g}_d)$$

Using this and the expression of $\bar{g}_o$ in terms of means of the parental generation and genetic superiorities of the selected parents (equation 3.3), expected response from the current to the next generation simplifies to:

$$\begin{aligned} E(R) &= \tfrac{1}{2}(\bar{g}_s + \bar{g}_d) + \tfrac{1}{2}(S_s + S_d) - \tfrac{1}{2}(\bar{g}_s + \bar{g}_d) \\ &= \tfrac{1}{2}(S_s + S_d) \end{aligned} \tag{3.6}$$

Thus, expected response from the current to the next generation is determined entirely by genetic superiority of the selected parents.

Note that for the simple case of equal selection in males and females, $S_s = S_d = S$ and $E(R) = S$.

In general we do not know the genetic value of parents. But we may have a prediction of their genetic value through an estimated breeding value (EBV), $\hat{g}$. Usually this prediction is based on a recognized method of genetic evaluation using different sources of phenotypic information. Examples are simple phenotypic selection, family index selection, pedigree index selection, BLUP, and so on. Whatever the method used, provided the estimate is unbiased, i.e. that

$$E(g \mid \hat{g}) \quad \hat{g}$$

then the expectation of the genetic value of an individual progeny is equal to the mean of the parental predictions, i.e.

$$E(g_o) = \tfrac{1}{2}\hat{g}_s + \tfrac{1}{2}\hat{g}_d = \hat{g}_p$$

where $\hat{g}_p$ is the mean estimated genetic value of the two parents.

Then, the expected mean genetic value of the progeny generation can be written in terms of the mean EBV of the selected and all parents by replacing $\bar{g}$ in (3.2) and (3.3) by $\bar{\hat{g}}$ as:

$$\begin{aligned} E(\bar{g}_o) &= \tfrac{1}{2}\bar{\hat{g}}_s^* + \tfrac{1}{2}\bar{\hat{g}}_d^* \\ &= \tfrac{1}{2}(\bar{\hat{g}}_s + \hat{S}_s) + \tfrac{1}{2}(\bar{\hat{g}}_d + \hat{S}_d) \end{aligned} \qquad (3.7)$$

Where $\hat{S}$ is the estimated genetic superiority of the selected parents, which can be obtained from (3.4) as:

$$\hat{S} = \bar{\hat{g}}^* - \bar{\hat{g}} \qquad (3.8)$$

Similarly, knowing the EBV of the parents, response from the current to the next generation can be predicted based on (3.5) and (3.6) as:

$$E(R) \quad \hat{g}_o \quad \bar{g}_p = \tfrac{1}{2}(\hat{S}_s + \hat{S}_d) \qquad (3.9)$$

It should be noted that equation (3.1) can be extended back so that the sire and dam terms are replaced by their respective sire and dam terms (i.e. grandsires and grandams of individual *i*) and so on back through the ancestor pathways, e.g.

$$g_o = \tfrac{1}{2}(\tfrac{1}{2}g_{ss} + \tfrac{1}{2}g_{ds} + g_{ms}) + \tfrac{1}{2}(\tfrac{1}{2}g_{sd} + \tfrac{1}{2}g_{dd} + g_{md}) + g_m \qquad (3.10)$$

where *ss* is sire of the sire, *ds* is dam of the sire, etc., and g_{ms} and g_{md} are the sire and dam

Mendelian sampling terms.

However, the expectation of g_o in terms of $\hat{g}_s$ and $\hat{g}_d$ in cannot easily be pushed back to include grandparental ($\hat{g}$) terms since the expectation of these terms depends on the degree of selection of the parents. However, solutions to most problems of design of breeding programs can be found using the parent-offspring relationships.

3.3 Predicting Response per Generation

The previous section allows us to predict response to selection if we have a particular group of chosen parents. This can be useful where we have an existing population of real animals and we want to predict the effects of choosing different combinations of animals as parents from that population. For example, in dairy cattle we might have several hundred bulls available for use, each with an estimated breeding value for milk yield. Assuming that the genetic evaluation procedure is unbiased, we could ask the consequences of using different numbers of bulls. Should we use the best 10 available or the best 20? Semen price is often (but not always!) related to quality, so that the top 10 bulls will often be more expensive than the next best 10 bulls. We could then ask how much genetic improvement would we expect when using the cheaper second set of 10 bulls rather than using the more expensive 10 best bulls. We will return to this problem later.

In many cases we are not interested in a particular group of existing animals but in predicting response to selection in future generations or in the consequences of different designs of animal breeding programs. We might ask, if we had a population of 100 bulls (which do not yet exist), what would be the expected response to selection if we use only the best 10 in comparison to using the best 20 every generation? The problem is then to predict the genetic superiority (S) of different types of possible parents in a hypothetical population as a result of a particular selection program.

A selection program typically is described by the fraction or number of males and females that are selected and by the criterion on which they are selected. Our objective here is to develop theory that can be used to predict the genetic superiority of selected parents based on this information.

We can assume that in this hypothetical population we have an estimate of each animal's genetic value, which we will call an index value that is used as the selection criterion. We do not need to know at this stage how this index is derived. But we will assume that there is a linear relationship between the index value and the true genetic value. We can then derive predictions of genetic superiorities of selected parents based on standard regression theory.

A standard equation for the regression of a dependent variable, y, on an independent variable, x, takes the form

$$y_i = a + b_{yx} x_i + e_i \qquad (3.11)$$

and a prediction of y given x is

$$\hat{y}_i = \bar{y} + b_{yx}(x_i - \bar{x}) \qquad (3.12)$$

where $\bar{y}$ is the mean value of y over all values of x, $\bar{x}$ is the mean value of x in the population of all possible values, and x_i is the observed value of x for the i^{th} individual for whom we wish to predict a value of y. From standard regression theory, the regression coefficient, b_{yx}, of y on x is given by

$$b_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = r_{xy} \frac{\sigma_y}{\sigma_x} \tag{3.13}$$

where σ_{xy} is the covariance of x and y, σ_x^2 is the variance of x, and r_{xy} is the correlation between y and x, which is given by

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_y^2 \sigma_x^2}} \tag{3.14}$$

In our breeding problem, we want to predict the genetic value of an individual (that will become a parent) given a recorded or estimated index value, I_i. Hence from (3.12),

$$\hat{g}_i = \bar{g} + b_{gI} (I_i - \bar{I}) \tag{3.15}$$

where I_i is the index value of individual i, $\bar{g}$ is the mean genetic value of individuals in the population, $\bar{I}$ is the mean index value of individuals in the population, and b_{gI} is the regression of genetic values on index values.

If we are predicting the average genetic value of a group of selected (chosen) animals, we get:

$$\bar{\hat{g}}^* = \bar{g} + b_{gI} (\bar{I}^* - \bar{I}) \tag{3.16}$$

To obtain a prediction of the genetic superiority of the selected parents, we can substitute (3.16) into (3.8), recalling that it is the genetic value of parents we are predicting, to get:

$$\hat{S} = \bar{\hat{g}}^* - \bar{g} = b_{gI} (\bar{I}^* - \bar{I}) \tag{3.17}$$

The right-hand side of equation (3.17) in parentheses, $(\bar{I}^* - \bar{I})$, is the deviation of index values of selected animals from the mean index value of all animals in the population. We can define the intensity of selection, i, as the deviation of selected from average animals in standard deviation units, i.e.

$$i = (\bar{I}^* - \bar{I}) / \sigma_I \tag{3.18}$$

where σ_I is the standard deviation of index values. It then follows from (3.18) that

$$(\bar{I}^* - \bar{I}) = i \sigma_I \tag{3.19}$$

and substituting (3.18) into (3.17) we get

$$\hat{S} = b_{g.I} \; i \; \sigma_I \tag{3.20}$$

From standard regression theory (equation 3.13), we recall that

$$b_{g.I} = r_{gI} \frac{\sigma_g}{\sigma_I} \tag{3.21}$$

hence, $$\hat{S} = r_{gI} \frac{\sigma_g}{\sigma_I} (i\ \sigma_I) = i\ r_{gI}\ \sigma_g \qquad (3.22)$$

Equation (3.22) gives a general formula to predict genetic superiorities of selected parents, which are needed to predict the response to selection. This formula applies whenever the value on which animals are selected, I, is linearly related to their additive genetic value. Predicted superiorities can be used to model the genetic level of future generations in a recursive manner using equation (3.7):

$$\begin{aligned} E(\bar{g}_o) &= \tfrac{1}{2}(\bar{g}_s + \hat{S}_s) + \tfrac{1}{2}(\bar{g}_d + \hat{S}_d) = \\ &= \tfrac{1}{2}(\bar{g}_s + i_s\, r_{g,I_s}\, \sigma_g) + \tfrac{1}{2}(\bar{g}_d + i_d\, r_{g,I_d}\, \sigma_g) \end{aligned} \qquad (3.23)$$

or model response per generation using equation (3.9):

$$R = \tfrac{1}{2}(S_s + S_d) = \tfrac{1}{2}(i_s\, r_{g,I_s}\, \sigma_g + i_d\, r_{g,I_d}\, \sigma_g) \qquad (3.24)$$

Methods to derive the accuracy of selection, r_{gI} , based on various sources of information will be reviewed and developed in Chapter 4. To illustrate, its derivation for the simplest case, phenotypic selection based on own phenotype, will be given in section 3.4. The intensity of selection, i, can be obtained from Normal distribution theory and will be further discussed in section 3.6. For the moment, we will assume that the genetic standard deviation, σ_g, is known and remains constant over generations. The latter assumption will be relaxed in Chapter 5.

In the remainder of this chapter, we will first illustrate equation (3.22) for phenotypic selection, then present how equation (3.23) fits in a general diagram for a deterministic simulation model, followed by a discussion of approximations for intensity of selection, and finally develop extensions of this equation to prediction of response with selection across multiple age groups, response per unit of time, and correlated response to selection.

3.4 Example of Phenotypic Selection

The generality of equation (3.22) can be seen by considering the specific and familiar case of phenotypic selection. In this case, the index value, I, is simply the phenotype of the animal. Assuming only additive genetic and random environmental effects, and assuming phenotype is adjusted for fixed effects (e.g. the mean), we can write the phenotypic value of an animal, y_i as

$$y_i = g_i + e_i$$

where e_i is the environmental effect, assumed uncorrelated with the additive genetic effect, g_i. Then, $$\sigma_{gI} = \sigma_{gy} = \sigma_{g,g+e} = \sigma_g^2$$

Thus $$r_{gI} = r_{gy} = \frac{\sigma_g^2}{\sqrt{\sigma_g^2 \sigma_p^2}} = \frac{\sigma_g}{\sigma_p} = h \qquad (3.25)$$

Where h is the square root of heritability.

Thus, from (3.22), $\hat{S} = i\, h\, \sigma_g$ (3.26)

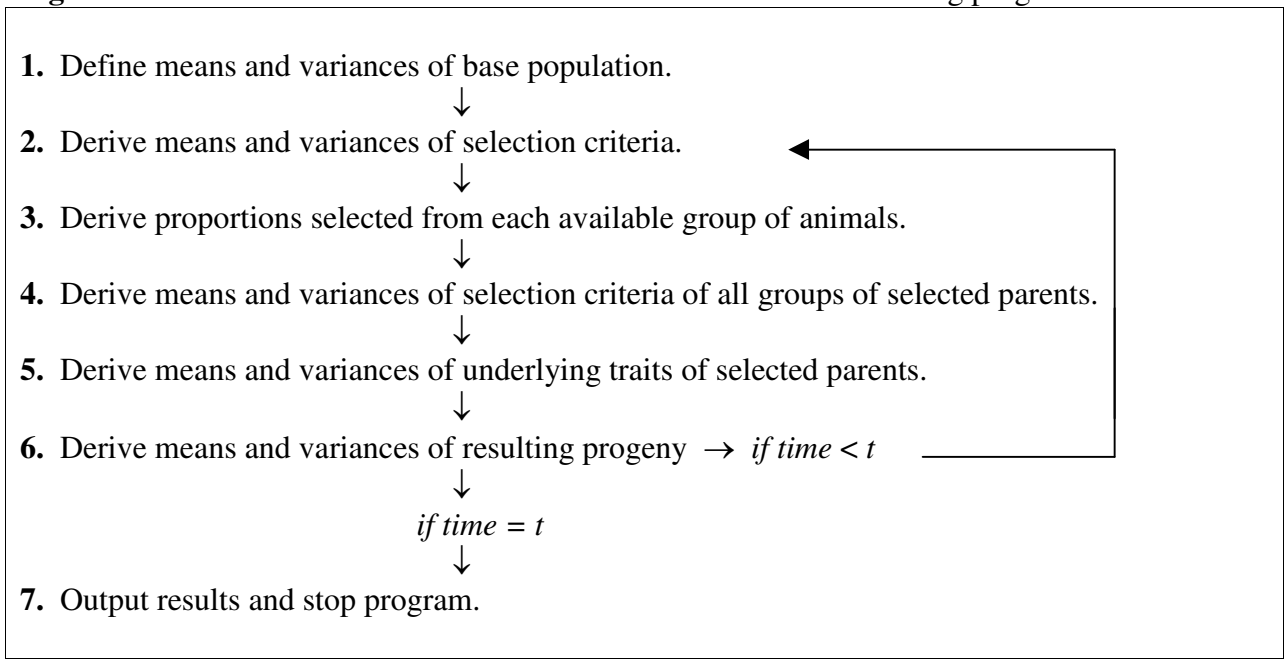
Recalling that heritability is $h^2 = \frac{\sigma_g^2}{\sigma_p^2}$, we get $\hat{S} = i\, h^2 \sigma_p$ (3.27)

Equation (3.27) should be familiar as the standard form for prediction of response to phenotypic selection. What we have shown here is that this standard response to phenotypic selection is just a special case of the general form of response to selection given by equation (3.22).

3.5 Simple Deterministic Model for Predicting Response to Selection with Multiple Age Groups

A general schematic for a simple deterministic simulation of a breeding program is given in Figure 3.1. Comparing to Figure 2.1 for a stochastic simulation, it should be clear that while the general flow of deterministic and stochastic simulations are similar, their fundamental nature is quite different. Whereas stochastic simulations model individual animals and their genetic and phenotypic characteristics, deterministic simulations model means and variances of genetic and phenotypic characteristics of groups of individuals. Recurrence equations such as equation (3.23) for computing the mean genetic value of progeny are used to compute characteristics of progeny. Other recursive equations, such as those for variances, will be presented in later Chapters. Another important component of deterministic simulations is the derivation of the means and variances of the selection criterion that is used. Variance of the selection criterion depends on the accuracy of selection. Methods to derive accuracy of selection are presented in Chapter 4.

Figure 3.1 General schematic of a deterministic simulation of a breeding program.



It is clear that, by modeling means and variances, deterministic simulations are computationally less demanding than stochastic models, besides the fact that deterministic models give expected responses and are not subject to stochastic variation in response. However, to accurately model all aspects of a breeding program deterministically does require more complicated models. Some of these will be described in the remainder of this chapter, while others follow in later chapters.

3.6 Selection Intensity with Truncation Selection

The prediction of response to selection given by (3.24) does not require that we know how animals are selected, merely that we know the mean index value of selected animals and hence are able to derive the intensity of selection, *i*.

Generally in animal breeding we consider the special case of truncation selection. In this form of selection, all animals above a certain index value, *x*, are chose for breeding and all animals below this value are discarded. Usually the truncation point is determined by the proportion, *p*, of animals to be used for breeding. In many cases, index values will be normally distributed. If so, and under the assumption of large population size, the relationships between *p*, *x* (measured in s.d. units), and *i* can be derived from the properties of the normal distribution to be equal to:

$$i = z/p \tag{3.28}$$

where *z* is the height of the normal distribution at the truncation point *x* and is given by

$z = \dfrac{e^{-1/2x^2}}{\sqrt{2\pi}}$ and π, to 9 decimal places, is 3.141592654.

For individual cases it is often convenient to look up the intensity of selection corresponding to a particular proportion selected from tables, such as those supplied by Falconer and MacKay (1996). When simulating breeding programs on the computer, many computer languages supply a routine that returns the truncation point, *x*, corresponding to a particular proportion selected, *p*.

Realized selection intensity in small populations will be less than predicted by $i=z/p$ as a result of order statistics (Hill 1976). Special tables are provided in Falconer and MacKay (1996) for specific population sizes. Analytically, intensities for finite population size can be approximated by adjusting *p* to p^* as follows:

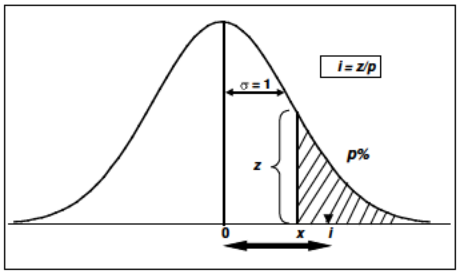
$$p^* = \frac{(s + {}^1/_2)}{n + \dfrac{s}{2n}} \tag{3.29}$$

where *s* is the number selected and *n* is the population size (i.e. uncorrected $p = s/n$), and then estimating the adjusted *i*, i^* as

$$i^* = \frac{z^*}{p^*} \tag{3.30}$$

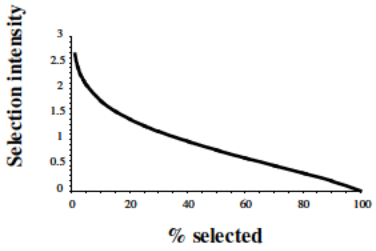
where z^* is the height of the normal distribution at the truncation point x^* corresponding to p^*.

Derivation of selection intensity from the standard Normal distribution



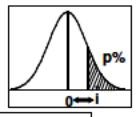
Effect of % Selected on Selection Intensity

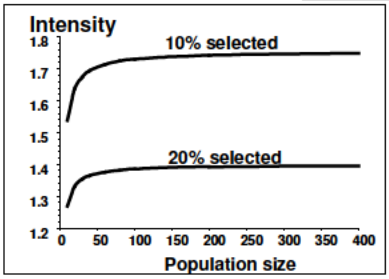
Large population size



Effect of Population Size

(Falconer and MacKay, 1996)





The second assumption that is made in the standard equation for selection intensity (3.28) is that there is no correlation between the selection criterion (EBV) of the different candidates of selection. Correlations between the selection criterion of different candidates are generally due to: 1) genetic relationships between candidates of selection; and 2) the use of the same information in calculating the EBV for different animals.

The most extreme example of such a correlation occurs when the population consist of n_{fs} full sib families with n_w individuals per family and selection based on pedigree information ($\hat{g}_o = \frac{1}{2}\hat{g}_s + \frac{1}{2}\hat{g}_d$). Note that the same pedigree information is used for all member of the family and, because this is the only information used, the correlation between their EBV is equal to 1.

The impact of a correlation between the selection criterion of candidates on intensity is related to the impact of population size on intensity. This is easy to see from the above example by noting that the number of alternative values the selection criterion has among all candidates is not $n = n_{fs}n_w$ but only n_{fs}. Thus, if n_c individuals are to be selected, selection is of n_c/n_w families out of n_{fs}, rather than of n_c individuals out of $n_{fs}n_w$.

Rawlings (1976) proposed a method of adjusting intensity for correlations between EBV, as well as finite population size based on:

$$i^* = \sqrt{1 \quad t_{av}} \; i \tag{3.31}$$

where t_{av} is the average correlation between the selection criterion across all possible pairs of selection candidates. For a population with unrelated full sib families, t_{av} can be derived based on the correlation of the EBV of full sibs, t_{fs}, and the correlation of the EBV of unrelated individuals (=0), each weighted by the number of full-sib pairs and unrelated pairs that exist in the population (Rawlings, 1976). The result is:

$$t_{av} = t_{fs} \frac{n_w \quad 1}{n_w n_{fs} \quad 1} \tag{3.32}$$

The correlation between the selection criterion of full sibs (t_{fs}) that is required for these computations can be derived based on the information that contributes to the selection criterion of each full sib. Computation of these correlations for more complex selection criteria will be covered in section 6.1, once selection index methods to derive EBV have been developed.

Meuwissen (1991) extended the method of Rawlings (1976) for populations where full sib families are nested within half sib families. This situation is more common in livestock populations and originates from mating each of n_{hs} sires to n_{fs} dams and where each dam produces n_w offspring. The resulting population consists of n_{hs} half-sib families with n_{fs} full sib families of n_w progeny per half-sib family. The selection intensity adjusted for finite population size and correlated EBV can then be approximated as a weighted average of the correlation between EBV of full-sibs (t_{fs}), the correlation between EBV of half-sibs (t_{hs}), and the correlation between EBV of unrelated individuals (0). Weighting each correlation by the number of pairs

that have that specific relationship results in the following equation for the average correlation between all possible pairs of individuals:

$$t_{av} = \frac{t_{fs}(n_w \quad 1) + t_{hs} n_w (n_{fs} \quad 1)}{n_w n_{fs} n_{hs} \quad 1} \tag{3.33}$$

Meuwissen (1991) compared this approximation with Monte Carlo simulation for a range of correlations and population sizes and found that the approximation worked well when low correlations between EBV were present or when the number of half-sib families was greater than 10. The approximation, however, overestimated the Monte Carlo results by up to 32% for a scheme with high correlations. A modified approximation for situations with high correlations between EBV was suggested by Meuwissen (1991).

Modern sire and dam evaluation methods use all available information for the prediction of breeding values. The use of more family information increases correlations between EBV of family members. In some breeding schemes, selection focuses on young animals because older animals tend to lag behind genetically. However, young animals have little information on individual or on progeny performance. In that case, family information dominates the prediction of EBV and correlations between EBV of relatives are expected to be high. For a correct comparison of schemes, it is therefore important to consider the effect of correlations between EBV, especially when the number of families is limited. In some animal selection experiments or in the nucleus herd of an animal breeding program, the population is often reproduced by rather few families, perhaps as few as 10, of at least half sibs. Even when the total size is larger, breeding may be carried out through the year with selection only among contemporaries at any time, and these may represent few families. In calculating the selection intensity in those cases, the correlation between family members should not be ignored (Hill, 1976).

3.7 Modeling Selection Across Multiple Age Groups

In many breeding populations, candidates for selection may come from several distinct groups, each with a different genetic mean and a different variance for the selection criterion. Examples might be: 1) dairy sires of various ages, where older sires have lower average genetic merit but will be more accurately evaluated and hence have higher variance for the selection criterion when their second crop of daughters become available; 2) selection of boars of different ages, where older boars will have lower average genetic merit; 3) selection of cows, where older cows have more lactations and therefore more accurate evaluations.

Genetic means of progeny generations and responses to selection can in these cases be derived by extending the principle obtained before. Considering sires and dams separately, assume that sires can be selected from three age groups, with the relative number of selection candidates in each age group equal to w_{s1}, w_{s2}, and w_{s3} ($\Sigma w_i = 1$). Fractions selected from each age group are p_{s1}, p_{s2}, and p_{s3}, for a total proportion selected of

$$P_s = p_{s1}\, w_{s1} + p_{s2}\, w_{s2} + p_{s3}\, w_{s3} \tag{3.34}$$

Let the genetic mean in age group i be denoted by $\bar{g}_{si}$ and the accuracy of the selection criterion by $r_{si.}$ For the moment we will assume the genetic standard deviation is the same in each age group and equal to σ_g. This assumption we be relaxed in later chapters.

Then, the genetic mean of selected sires in age group i is equal to:

$$\bar{g}^*_{si} = \bar{g}_{si} + S_{si} \tag{3.35}$$

where S_{si} is the genetic superiority of the selected sires from age group i over the mean of all males in that age group, and can be predicted as before based on

$$\hat{S}_{si} = i_{si}\, r_{si}\, \sigma_g \tag{3.36}$$

where i_{si} is the intensity that corresponds to a fraction selected p_{si}.

Using a weighted average based on the relative number of sires from each age group, the mean genetic value of selected sires can be computed as:

$$\begin{aligned}\bar{g}^*_s &= \frac{1}{P_s}\{p_{s1}\, w_{s1}\, \bar{g}^*_{s1} + p_{s2}\, w_{s2}\, \bar{g}^*_{s2} + p_{s3}\, w_{s3}\, \bar{g}^*_{s3}\} \\ &= \frac{1}{P_s}\sum p_{si}\, w_{si}\,(\bar{g}_{si}+S_{si})\end{aligned} \tag{3.37}$$

Similarly, the mean genetic value of dams can be derived as:

$$\bar{g}^*_d = \frac{1}{P_d}\sum p_{di}\, w_{di}\,(\bar{g}_{di}+S_{di}) \tag{3.38}$$

and the average genetic value of the progeny as

$$\begin{aligned}E(\bar{g}_o) &= \tfrac{1}{2}\bar{g}^*_s + \tfrac{1}{2}\bar{g}^*_d \\ &= \tfrac{1}{2}\frac{1}{P_s}\sum p_{si}\, w_{si}\,(\bar{g}_{si}+S_{si}) + \tfrac{1}{2}\frac{1}{P_d}\sum p_{di}\, w_{di}\,(\bar{g}_{di}+S_{di})\end{aligned} \tag{3.39}$$

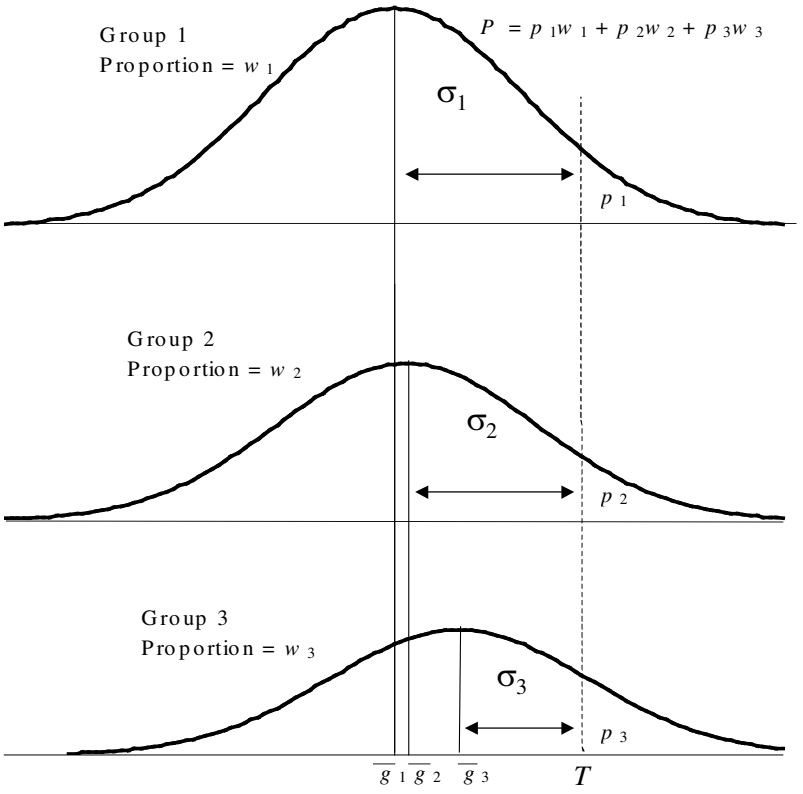
These equations allow for recursive prediction of the genetic mean of the population in successive time periods. In Chapter 8, we will formalize these recursive equations in the form of gene flow.

In the previous, the proportions selected from each age group were pre-determined. These proportions may, however, not maximize the average genetic value of the selected parents and, thereby, the genetic value of progeny. Thus, referring to sires, the problem is to determine the proportions to select from each age group such that the average genetic value of the selected group is maximized, but subject to the constraint that the total proportion selected is equal to P_s.

To address this problem, we'll assume that the selection criterion I_i for each age group i is unbiased. This implies that $E(g_i|I_i) = I_i$ and also that the selection criterion can be compared across age groups. Thus, individuals with the same value v of the selection criterion in different age groups are expected to have the same genetic value v.

The general problem is illustrated in Figure 3.2. Given the assumptions for the selection criterion, individuals should be selected by truncating across the distributions of the selection criterion; replacing an individual in age group 1 that falls just above the truncation point with an individual from age group 2 that falls just below the truncation point will reduce the expected genetic value of selected parents. Thus, the same truncation point should be used for all distributions. In practice, this would be equivalent to ranking all individuals based on their EBV regardless of the age group they belong to, and selecting the top ones.

Figure 3.2 Schematic representation of truncation selection of a total proportion P across multiple overlapping distributions



Thus, to maximize the genetic value of selected parents, the objective is to find the truncation point T where selection of sires across all available distributions yields a total proportion selected of P_s. There is no algebraic solution to this problem and the answer must be found iteratively. Bisection is a general, simple, and effective optimization method that can be used for this problem. A schematic of a simple computer subroutine to do this is illustrated below.

1. Find for all i the (unstandardized) truncation point, T_i, of the i^{th} distribution that corresponds to a proportion P selected from that distribution ($T_i = \overline{g}_i + x_i\sigma_i$, where x_i is the standardized truncation point and σ_i the standard deviation of the i^{th} distribution ($\sigma_i = r_{si}\sigma_g$ for our case))

2. Choose the lowest T_i as a lower bound for $T \rightarrow T_l$
   Choose the highest T_i as a upper bound for $T \rightarrow T_u$. (T must lie between T_l and T_u.)

3. Compute the mean of the upper and lower bound $\rightarrow$ $T_m = ½ (T_u + T_l)$

4. For each distribution i, find the proportion selected, p_i, that corresponds to truncation at T_m.

5. Find the total proportion selected for truncation at T_m: $P_m = \Sigma p_i w_i$

6. If $|P_m - P| < \varepsilon$, where ε is a pre-set convergence criterion, exit the routine and return T_m as the optimized truncation point.

7. If $P_m < P$ then T_m becomes the new upper bound $\rightarrow$ set $T_u = T_m$
   If $P_m > P$ then T_m becomes the new lower bound $\rightarrow$ set $T_l = T_m$

8. *Return to step* 3.

Even with a large number of distributions, this program will iterate to a solution with high accuracy fairly rapidly. For most applications no more than 5 or 6 rounds of iteration should be required.

The proportion of animals in each distribution, w_i, might reflect structural differences in numbers (different numbers produced in different groups as designed in the breeding program) and losses from groups over time due to death, disease, sales, etc. Differences between groups in reproductive capacity (fertility) could be incorporated directly into w_i, or treated as a separate factor affecting the effective numbers (in terms of contributions to progeny) in each group after selection.

3.8 Asymptotic Response per Unit Time

Response defined by equations (3.22) and (3.24) is the response from one generation to the next. If conditions remain constant over generations, it is also the response per generation. ***Generation interval*** is generally defined as the average age of the parents when their progeny are born or as the average time between birth of parents and birth of progeny.

Generation intervals vary widely across species. For example, a generation interval for poultry and swine can be as short as 1 year, whereas for progeny testing schemes in cattle, generation intervals for sires are often 7 years or more. Generation intervals can also be altered within species by changing the age at which animals are selected and bred.

In general, it is more useful to estimate response per unit time, usually response per year. Response per year is often given the same notation as response per generation, R.

When selection is equal in males and females and, therefore, response per generation is equal to $R = S = ir_{gI}\sigma_g$, response per year is obtained by dividing equation (3.22) by the generation interval, L, to get

$$R = \frac{ir_{g,I}\sigma_g}{L} \tag{3.40}$$

(Note, in general, as here, we must be careful to know whether response, R, is expressed per generation, per year, or in some other unit of time).

Equation (3.40) holds the key to designing breeding programs. Response per unit of time is proportional to the intensity of selection, the accuracy of genetic evaluation, and the square root of the genetic variance, and is inversely proportional to the generation interval.

3.8.1 Multiple Pathways of Selection

The derivations leading to equation (3.40) assumed that males and females are treated alike. In practice this is often not the case. For example, in most species, males have a higher reproductive rate than females, thus we need fewer males for breeding and consequently can have a higher intensity of selection in males than females. In some species, traits of interest are recorded only in one sex, obvious examples being milk yield in dairy cattle, litter size in swine, and rate of egg production in poultry. This can lead to different accuracies of evaluation in the two sexes, since one sex has it's own performance contributing to it's evaluation while in the other sex genetic evaluation must be based entirely on information from relatives. Similarly, different sexes can have different generation intervals for a variety of reasons, e.g. the sex with the highest reproductive rate (usually males) may take less time to produce replacement offspring and hence potentially have the shortest generation interval.

In these cases, response per unit of time can be derived by deriving the sum of genetic superiorities in males and females (S_s and S_d) by the sum of their generation intervals (L_s and L_d):

$$R = \frac{S_s + S_d}{L_s + L_d} \tag{3.41}$$

This is referred to as the '***steady state***' or '***asymptotic***' response to selection, which is the expected response per unit of time after the breeding program has been in operation for several years. The reason for this assumption will be made clear in the derivation of the equation, which follows.

In practice it may take several generations to approach this steady state, and in some cases a true steady state may never be reached. It is therefore generally safer to think of R predicted by equation (3.41) as the prediction of the average rate of response per year, recognizing that predicted response may well vary from one year to the next. Even where a steady state response rate is eventually achieved, genetic response will usually be variable from one year to the next in the early generations of the breeding program.

Note that responses from year to year can always be predicted from the recursive equation (3.23). A comparison of this approach with the asymptotic response is given in Figure 3.3 Note that, starting from an unselected population, expected responses fluctuate during the initial years but stabilize to the asymptotic response after several years of selection.

Figure 3.3 Example of predicted annual versus asymptotic responses Starting from an unselected population

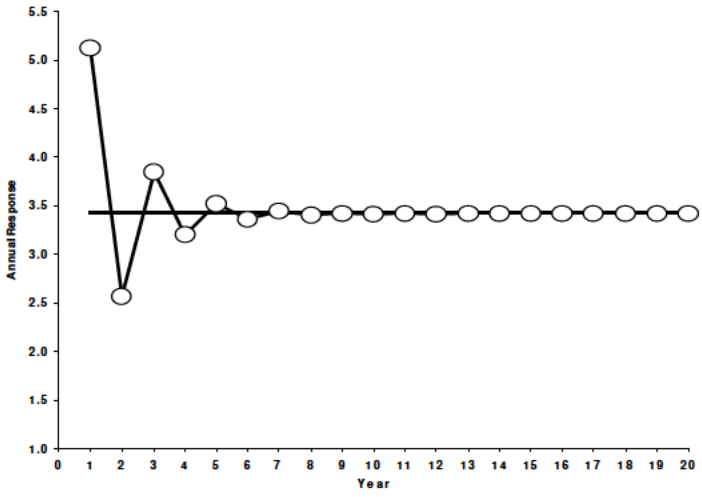
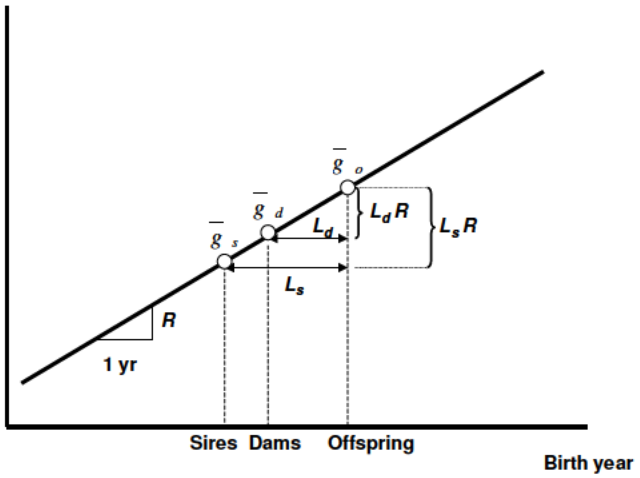


Figure 3.4 Asymptotic response to selection for breeding programs with overlapping generations



To derive equation (3.41), we start by describing the genetic mean of progeny in terms of the average of the genetic mean of the selected parents, from equation (3.23):

$$\bar{g}_o = \tfrac{1}{2}\bar{g}_s^* + \tfrac{1}{2}\bar{g}_d^* = \tfrac{1}{2}(\bar{g}_s + S_s) + \tfrac{1}{2}(\bar{g}_d + S_d) \qquad (3.42)$$

Now, referring to Figure 3.4, note that if the asymptotic response of R per year has been achieved, the genetic mean of male selection candidates is expected to be L_sR lower than the genetic mean of the progeny generation. This is because males are on average L_s years older than their progeny and the gain per year is equal to R. Thus, the genetic mean of male candidates can be expressed as:

$$\bar{g}_s = \bar{g}_o - L_sR$$

and similarly,

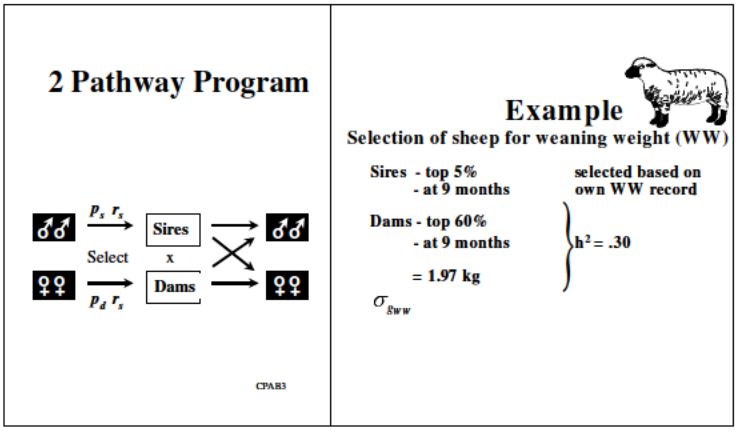
$$\bar{g}_d = \bar{g}_o - L_dR$$

Substituting into equation (3.42) we get:

$$\begin{aligned}\bar{g}_o &= \tfrac{1}{2}(\bar{g}_o - L_sR + S_s) + \tfrac{1}{2}(\bar{g}_o - L_dR + S_d)\\ &= \bar{g}_o - \tfrac{1}{2}R(L_s + L_d) + \tfrac{1}{2}(S_s + S_d)\end{aligned}$$

Rearranging and solving for R results in equation (3.41).

Equation (3.41) applies to a so-called two-path selection program, in which selection differs between males and females.



Predicting Response in WW

Path	%	i	r = $\sqrt{h^2}$	Genetic Superiority	Gen. Interval
Sire	5	2.06	.55	2.23	1.17 yr
Dam	60	.64	.55	.69	1.17 yr
				2.92	2.34 yr

ΔG_{WW} = 2.92/2.34 = 1.25 kg/yr

Rendel and Robertson (1950) and Robertson and Rendel (1950) pointed out that in any breeding program there are actually four basic pathways of genetic improvement, corresponding to the four sources of parental genes of male and female progeny. These four pathways are:

- male parents of male progeny (sires of males, *sm*)
- female parents of male progeny (dams of males, *dm*)
- male parents of female progeny (sires of females, *sf*)
- female parents of female progeny (dams of females, *df*).

Robertson and Rendel showed that where each of the four pathways of genetic improvement were separately recognized, response per generation as predicted by equation (3.41) can be rewritten as:

$$R = \frac{S_{sm} + S_{dm} + S_{sf} + S_{df}}{L_{sm} + L_{dm} + L_{sf} + L_{df}} = \frac{\sum_i S_i}{\sum_i L_i} \tag{3.42}$$

For each path, genetic superiorities can be derived as shown before as: $S_i = i_i r_i \sigma_g$

When for a particular path selection is across multiple age groups, genetic superiority for that path can be computed as a weighted average of genetic superiorities achieved within each age group. To illustrate, referring to the example of selection across three age groups of section 3.6, the superiority of that path would be computed as:

$$S_s = \frac{1}{P_s}\{p_{s1}\, w_{s1}\, S_{s1} + p_{s2}\, w_{s2}\, S_{s2} + p_{s3}\, w_{s3}\, S_{s3}\} \tag{3.43}$$

Similarly, the generation interval for this path would be computed as:

$$L_s = \frac{1}{P_s}\{p_{s1}\, w_{s1}\, L_{s1} + p_{s2}\, w_{s2}\, L_{s2} + p_{s3}\, w_{s3}\, L_{s3}\} \tag{3.44}$$

Selection Across Age Groups

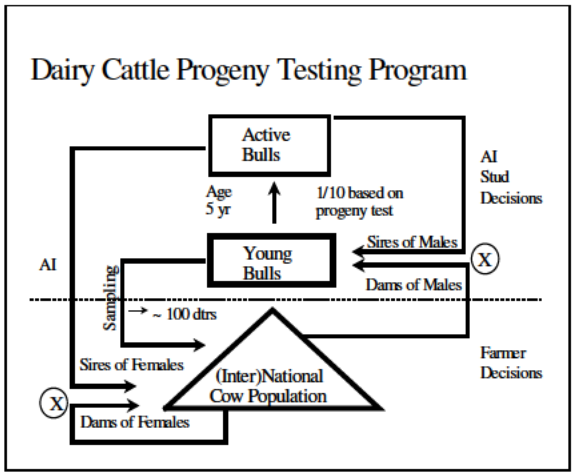
e.g. Selection of Bull Dams σ_g = 550 kg

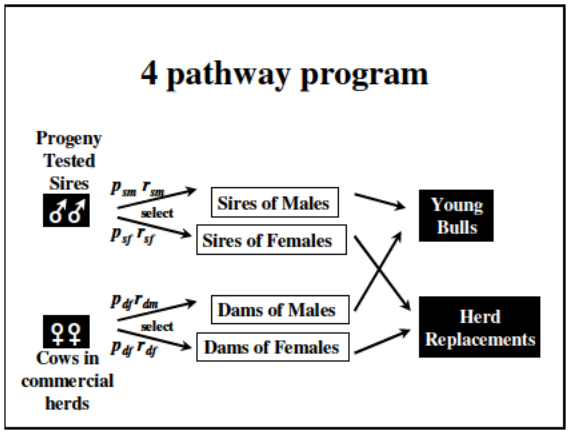
Age Group	Age at Birth of Progeny	% of Bull Dams	% Selected	i	r	Genetic Superiority $ir\sigma_g$
Heifers	2 yr	50%	2.5%	2.34	.55	707.9
1st Lact.	3 yr	30%	1.5%	2.53	.68	946.2
2nd Lact.	4 yr	20%	1.5%	2.53	.72	1001.9

Selection Across Age Groups

- **Pooled Generation Interval**
  L_{dm} = 50% * 2
  + 30% * 3
  + 20% * 4
  = 2.7 yr
- **Pooled Genetic Superiority**
  S_{dm} = 50% * 707.9
  + 30% * 946.2
  + 20% * 1001.9
  = 838.2 kg

To illustrate a breeding program in which all four pathways of improvement are recognized, we can consider a conventional progeny testing program for improvement of milk production in dairy cattle with the use of artificial insemination. For simplicity we assume all cows reproduce naturally without the aid of embryo transfer. In such a scheme, young bulls are tested by mating to a (hopefully) random sample of cows, the resulting heifers are reared, and their first lactation performance is recorded. This daughter lactation information is then used to produce a genetic evaluation on each young bull, often referred to as the "first proof" of a bull. At this stage the best bulls can be selected for breeding and the remainder discarded. In contrast, heifers and cows are evaluated largely based on their own lactation performance. In a population of several hundred thousand recorded dairy cows, several hundred young bulls, perhaps up to a thousand, would be tested each generation.





We can now consider each of the four pathways of genetic improvement in a highly efficient hypothetical progeny-testing program.

Sires of males: Since we only test a few hundred young bulls, and every sire can produce tens of thousands of doses of semen, we need only a few sires to produce these young bulls each generation. Thus we need to select only the top 1 or 2% of tested bulls as sires of sons. These sires have high accuracy of genetic evaluation, since progeny tests generally give high accuracy. The generation interval will, however, be at least 6 years because of the time from birth of the young bull, through the birth of his first crop of test daughters, through their first lactation to the birth of his sons.

Sires of females: Since there are several hundred thousand cows to be bred, many more bulls are required to produce the necessary amount of semen each generation. In an efficient scheme, the top 10-15% of young bulls can be selected, giving a lower selection intensity than for sires of sons. Accuracy of selection is the same as for sires of sons because they are chosen on the basis of the same information. The generation interval is, however, about a year longer because it takes time to breed a large population of cows and the better bulls will be used by farmers for a little longer than the not so good bulls.

Dams of males: Since there are several hundred thousand cows and only a few hundred sons are tested, dams of sons can be selected very intensely, perhaps only the best 0.1 to 0.5% being required. But evaluation is based on their own performance, which has lower accuracy than a progeny test. These cows could be bred in their second lactation based on their first lactation performance and part of their second lactation performance, so that they would be around 4½ to 5 years old at the birth of their sons.

Dams of females: Dairy cows have a very low reproductive rate, producing less than one live calf per year, after allowing for average calving intervals and mortality of fetuses and calves. Allowing for disease and other losses of growing heifers and for the fact that only half the calves are females, only about 1 in 3 calvings result in a potential replacement heifer for the dairy herd. Since average life in the herd in many western countries is often not much over three lactations, the average cow barely has sufficient time to produce a

replacement before she leaves the herd. There is thus very little room for selection of dams of cows, with perhaps 90% of all cows required for breeding. Accuracy of selection would be very similar to that for dams of sires. However, generation interval is generally increased by a year or two, since the average cow takes close to three calving to produce a replacement.

The parameters applying to each pathway are summarized in Table 3.1.

Table 3.1. Intensity and accuracy of selection and generation interval in a highly efficient hypothetical progeny-testing program for improving milk yield in dairy cattle.

Pathway	Proportion Selected (p_i)	Intensity (i_i)	Accuracy (r_i)	Genetic Superiority ($S_i = i\ r_i\sigma_g$)	Generation Interval (yr) (L_i)
Sires of males	2 %	2.42	0.90	2.178 σ_g	6
Sires of females	10 %	1.75	0.90	1.575 σ_g	7
Dams of males	0.5 %	2.89	0.60	1.743 σ_g	5
Dams of females	90 %	0.19	0.60	0.114 σ_g	6
TOTAL				$\Sigma S = 5.601\sigma_g$	$\Sigma L = 24$

If we assume that genetic variance is the same for all pathways (a common assumption but not always strictly true; see Chapter 5), then we can use the parameter values in Table 3.1 to obtain an estimated annual rate of response for this particular breeding program, of

$$R = \frac{5.601}{24}\sigma_g = 0.233\ \sigma_g \text{ per yr}$$

Response could of course be expressed in many units, but the three most common and probably most useful are in genetic standard deviations, σ_g, per year (as above), absolute units per year (e.g. kg milk per year), or as a percentage of the mean per year.

Imagine that the dairy cattle population above has a mean yield of 6000 kg, that the heritability (h^2) of milk yield is 0.25, and that coefficient of variation (CV) is 0.18, all fairly typical values for intensive dairy production. Since

$$\sigma_g^2 = h^2\ \sigma_p^2$$

and $\sigma_p^2 = (cv \times \bar{x})^2$,

then $\sigma_p^2 = (0.18 \times 6000)^2 = (1080)^2$.

Hence $\sigma_g^2 = 0.25(1080)^2$

And $\sigma_g = \sqrt{\sigma_g^2} = 0.5 \times 1080 = 540$ kg.

Hence $R = 0.233 \times 540 = 125.82$ kg per year

or, alternatively, $R = 125.82/6000 = 2.1\%$ per year.

The choice of units will depend on how the results are to be used. Use of genetic standard deviation units may be useful to geneticists who think in such terms and allow results to be readily converted from one population to the next if it is believed that the major variation between populations is in the absolute amount of genetic variance. For example, this would be true if h^2 and *cv* were the same for different populations but the mean level of performance differed.

Absolute units, such as kg milk per year, are often the most intelligible to people familiar with the species and trait(s) in question. For example, there would probably be little point in presenting results in σ_g per year if the audience is made up of non-geneticists, such as dairy farmers, industry, or government officials.

Expressing results in terms of percentage change per year is likely to be understood by a wide audience. It also has the advantage of allowing relatively meaningful comparisons of response for different traits across species. A good example is given by Smith (1984), who compared the theoretical response rate for typical breeding programs for sex-limited traits in poultry, swine, sheep, and cattle. The traits were egg production in poultry, litter size in swine, litter size in sheep, and milk production in cattle. His estimates of absolute response rates were 5.46 eggs per year, 0.3 piglets per year, 0.04 lambs per year, and 75 kg milk per year. Expressed in absolute units, it is clearly very difficult to interpret these results or make any comparison across species. However, expressed as percentage change per year, the same results were 2.1, 3.0, 2.1, and 1.5% per year for poultry, swine, sheep, and dairy cattle. Although not perfect, this does allow us to draw such general conclusions, as that selection for sex-limited traits should give roughly similar relative rates of response in different species. It may come as a surprise to those working with dairy cattle, that the relative rates of response are lowest for milk production in cattle.

Accounting for use of young bulls

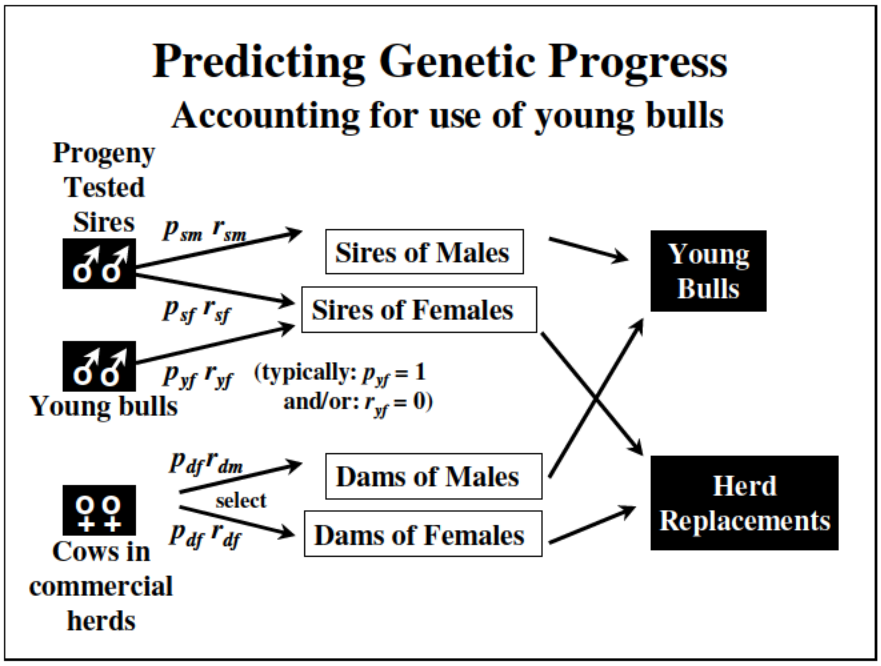
In the previous, the sire to female path only accounted for the use of progeny-tested sires to breed cows to produce herd replacements. However, young bulls also contribute to the next generation of females; in a practical breeding program, semen from young bulls can represent as much as 20% of all inseminations. To account for this, the genetic superiority and generation interval for sires of females must be computed as a weighted average. Assuming y is the proportion of females produced from young bulls, genetic superiority of the sire to female path is computed as:

$$S_{sf} = y\, S_{yb,f} + (1\text{-}y)\, S_{pb,f}$$

where $S_{yb,f}$ and $S_{pb,f}$ are genetic superiorities of young and progeny-tested bulls that are used to breed female replacements. In most cases, $S_{yb,f} = 0$ because $p_{yb,f} = 1$ and thus $i_{yb,f} = 0$, unless there is additional selection of young bulls that are entered into the progeny tests, above and beyond selection of their parents (which is already covered through the *sm* and *dm* pathways). An example where $S_{yb,f} > 0$ is preselection of young bulls based on genetic markers (see Chapter 12).

Similarly, the generation interval for the *sf* pathway is computed as a weighted average of the generation intervals for the *yb,f* and *pb,f* pathways:

$$L_{sf} = y\, L_{yb,f} + (1\text{-}y)\, L_{pb,f}$$



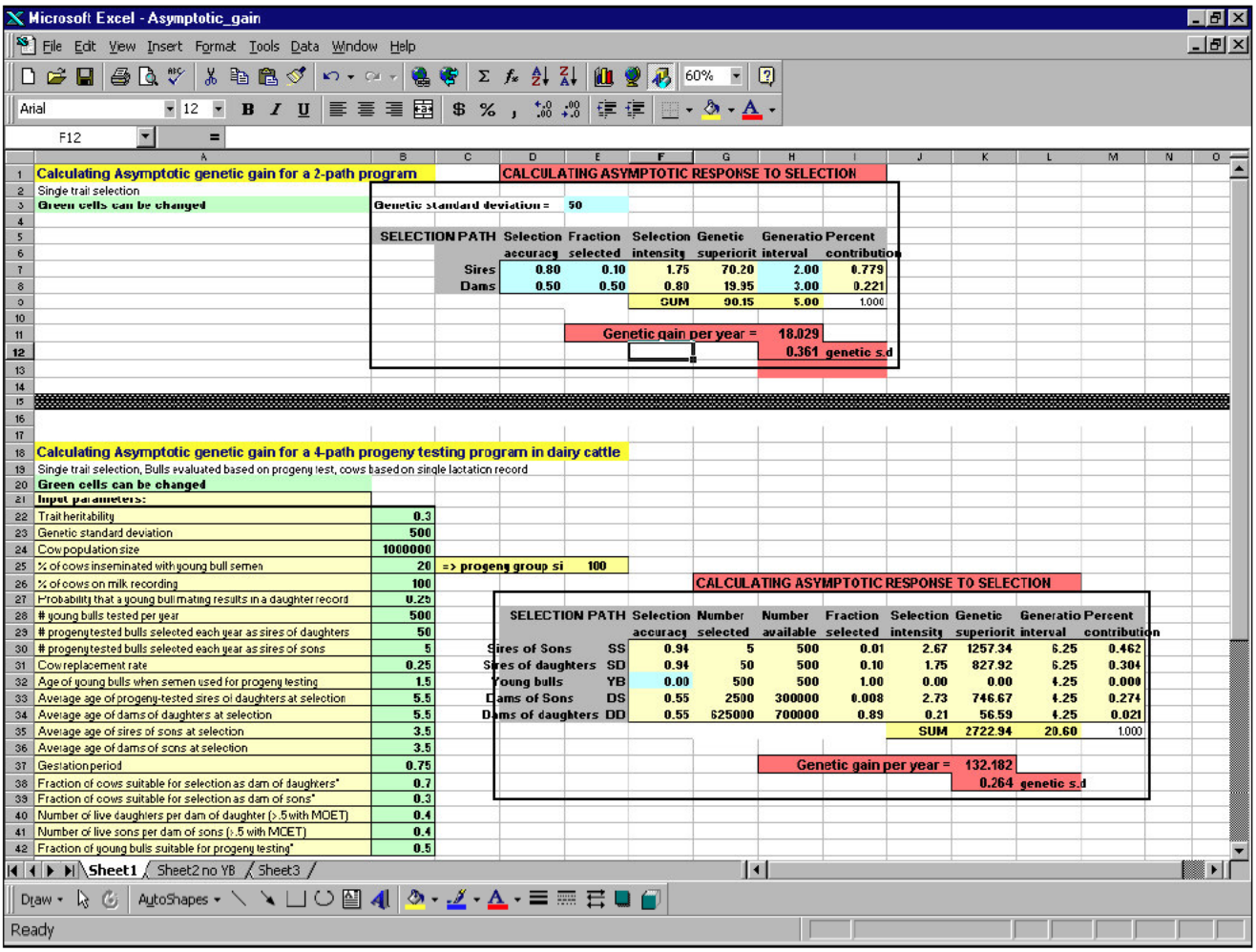
An example is given in Table 3.2, which assumes $y = 0.2$

Table 3.2. Intensity and accuracy of selection and generation interval in a highly efficient hypothetical progeny-testing program for improving milk yield in dairy cattle with accounting for 20% use of young bulls to breed female replacements.

Pathway	Proportion Selected (p_i)	Intensity (i_i)	Accuracy (r_i)	Genetic Superiority ($S_i = i\, r_i \sigma_g$)	Generation Interval (yr) (L_i)
Sires of males	2 %	2.42	0.90	$2.178\sigma_g$	6
Sires of females - Young	100 %	0	0.50	0	2
Sires of females - Proven	10 %	1.75	0.90	1.575	7
				$1.260\sigma_g$	6
Dams of males	0.5 %	2.89	0.60	$1.734\sigma_g$	5
Dams of females	90 %	0.19	0.60	$0.114\sigma_g$	6
TOTAL				$\Sigma S = 5.268\sigma_g$	$\Sigma L = 23$

Now response per year becomes: $R = \frac{5.268}{23}\sigma_g = 0.230\ \sigma_g$ per yr

Note that, compared to Table 3.1, response is slightly lower. By changing y, this approach can be used to optimize the proportion of the population to inseminate with young bulls. Note, however, that increasing y also increases the number of young bulls that can be tested or, alternatively, the number of progeny per young bulls. This has consequences for other parameters of the breeding program. Nevertheless, this method provides a means to look at the impact of various factors on genetic gain. A spreadsheet to evaluate alternative program parameters is provided.



3.9 Correlated Response to Selection

Selection for trait i will not only result in genetic change in trait i (R_i) but also in traits that are genetically correlated to the selected traits. Genetic change in trait j to selection on trait i is referred to as correlated response to selection and will be denoted R_{ji}, in contrast to direct response, which is denoted by R_i. Similarly, genetic superiorities of parents selected on trait i will be denoted by S_i and superiorities for trait j by S_{ji}.

Following equation (3.22), genetic superiority of parents for trait 2 as a result of selection on an index for trait 1, I_1, can be obtained based on the general equation:

$$S_{2.1} = i\, r_{g_2 I_1} \sigma_{g_2} \tag{3.45}$$

Here $r_{g_2 I_1}$ is the correlation of the genetic value for trait 2 with the criterion that selection is based on, i.e. I_1. When the selection criterion I_1 is only based on records for trait 1 (single trait

evaluation), this correlation can be expressed in terms of the accuracy of selection for trait 1 and the genetic correlation as: $r_{g_2 I_1} \quad r_{g_2 g_1} r_{g_1 I_1}$

Then:
$$S_{2.1} = i\, r_{g_2 g_1} r_{g_1 I_1}\, \sigma_{g_2} = r_{g_1 g_2} \frac{\sigma_{g_2}}{\sigma_{g_1}} i\, r_{g_1 I_1}\, \sigma_{g_1} = r_{g_1 g_2} \frac{\sigma_{g_2}}{\sigma_{g_1}} S_1 = b_{g_2 g_1} S_1 \qquad (3.46)$$

where $b_{g_2 g_1}$ is the regression of genetic values for trait 2 on genetic values for trait 1. This regression coefficient quantifies the expected genetic change in trait 2 for every unit genetic change in trait 1. When the selection criterion is not exclusively based on records for trait 1, e.g. the index is a multiple-trait index, the same principle holds but derivation of the regression coefficient becomes more complex. This will be dealt with in Chapter 4.

Correlated response to selection can now be predicted from direct response by simple regression techniques:
$$R_{2.1} = b_{g_2 g_1} R_1 = = r_{g_1 g_2} \frac{\sigma_{g_2}}{\sigma_{g_1}} R_1 \qquad (3.47)$$

Where R_1 can be predicted using equation (3.41).

Predicting Correlated Response to Selection

- Select on trait 1
  - Response in trait 1
    - ⇨ Direct response to selection
  - Response in trait 2 IFF $r_{g_{1,2}} \neq 0$
    - ⇨ Correlated response to selection
    - ⇨ Change in trait due to selection on other trait

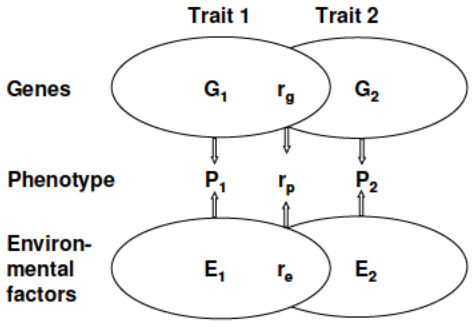
Correlations between Traits

Phenotypic correlation (r_p) = correlation between phenotypes for traits 1 and 2 observed on the same individuals

Causes for existence of a phenotypic correlation:

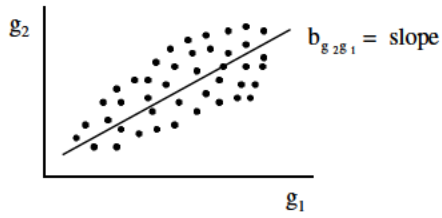
1) Some genes can have effects on both traits = Pleiotropy
   ⇨ genetic correlation (r_g)
2) Some environmental factors can affect both traits
   ⇨ environmental correlation (r_e)

Correlations between Traits



Correlated Response to Selection

Regression of BV for trait 2 on BV for trait 1



Example

Selection of sheep on weaning weight

- Direct response = ΔG_{WW} = 1.25 kg/yr
- Correlated response in birth weight?

$\sigma_{g_{BW}} = .5$ kg
$r_{g_{WW,BW}} = +.3$
$\Delta G_{BW.WW} = b_{A_{BW}A_{WW}} \Delta G_{WW}$

Predicting Response in W W

Path	%	i	$r = \sqrt{h^2}$	Genetic Superiority	Gen. Interval
Sire	5	2.06	.55	2.23	1.17 yr
Dam	60	.64	.55	.69	1.17 yr
				2.92	2.34 yr

ΔG_{WW} = 2.92/2.34 = 1.25 kg/yr

Prediction of Correlated Response

$$b_{A_{BW}.A_{WW}} = r_g \frac{\sigma_{g_{BW}}}{\sigma_{g_{WW}}} = (.3)\ \frac{.5}{1.97} = .076 \text{ kg/kg}$$

$$\Delta G_{BW.WW} = (.076)(1.25) = .095 \text{ kg/yr}$$

Indirect Selection

- Genetic improvement of a trait of economic importance through selection on EBV for a correlated trait, e.g.
  - Select on Somatic Cell Count to improve mastitis resistance.
  - Select on conformation traits to improve herd life.
  - Select on scrotal circumference to improve fertility (sheep).

Indirect Selection (cont'd)

- Advocated over direct selection if:
  - Correlated trait is recorded and direct trait not.
  - Correlated trait is less expensive to measure.
  - Correlated trait is measured earlier in life ∴L↓
  - Correlated trait has higher h^2.

Efficiency of Indirect vs Direct Selection

1 = correlated trait
2 = economic trait

- Direct selection:

$$\Delta G_2 = \frac{\Sigma i r_{\hat{A}_2} \sigma_{g_2}}{\Sigma L}$$

Efficiency of Indirect vs Direct Selection (cont'd)

- Indirect selection: (correlated response in trait 2 to selection on trait 1)

$$\Delta G_{2.1} = r_{g_{1\,2}} \frac{\sigma_{g_2}}{\sigma_{g_1}} (\Delta G_1)$$

$$\uparrow \quad \Delta G_1 = \frac{i r_{\hat{A}_1} \sigma_{g_1}}{\Sigma L}$$

$$\text{Efficiency} = \frac{\Delta G_{2.1}}{\Delta G_2}$$

3.10 Design of Breeding Programs

The prediction of rate of response to selection given by equation (3.40) and in its more complete form by equation (3.42) holds the key to understanding many of the basic principles of design of breeding programs. In general, response is positively related to intensity and accuracy of selection and to amount of genetic variation, and is negatively related to generation interval. Altering a breeding program will often affect several parameters simultaneously and it is the net effect of all these changes that determines the predicted response to selection.

Consider the dairy cattle progeny testing scheme outlined in section 3.8.1. We could, for example, ask the consequence of waiting until potential dams of sires were older and thus had more lactation records than in the scheme originally outlined. This would increase accuracy of evaluation in this pathway somewhat, because of the increase in information available, but would also increase the generation interval. Later in this course you will have the tools to predict the expected change in accuracy, but at this stage we will simply state that by waiting for an extra year, the accuracy of evaluation in the dams of sires pathway would increase from 0.6 to 0.64 while the generation interval increases from 5 to 6 years. Thus the predicted rate of response is now $R = \frac{(2.42 \text{ x } 0.9 + 1.75 \text{ x } 0.9 + 2.89 \text{ x } 0.64 + 0.19 \text{ x } 0.6}{6 + 7 + 6 + 6} \sigma_g = 0.229 \ \sigma_g$ per year

which is less than the predicted response of 0.233 σ_g per year when selecting younger dams of sires. Assuming our parameters are appropriate, we would conclude that we should not wait for extra lactation records on our potential dams of sires.

As another example, we could go on to ask what would happen if we tested more young bulls in our progeny test program each generation. If testing resources were limited by having more young bulls to test, we would have to produce fewer daughters per bull. Thus accuracy of selection would decrease (due to having fewer daughters) and intensity of selection would increase (due to having more young bulls to choose among) in both sire pathways. But also, if we had more young bulls tested, we would need more dams to produce these bulls, which would increase the proportion selected and reduce intensity of selection in the dams of sons pathway. In such a situation we could vary the number of young bulls tested per generation, calculating the appropriate selection intensities and accuracies in each pathway and hence derive the expected rate of response to selection for each number tested. The number of bulls tested that maximized response rate could then be identified.

As we will see later in this course, the above approach is only an approximation to the real world. But in many cases this approximation can be quite reliable in its own right. Adapting this approximation to more complex (realistic?) situations is not necessarily particularly difficult.

Another consideration is that the design that maximizes genetic response is not necessarily the design that maximizes economic progress. To evaluate the optimum design from an economic perspective requires that the economic costs be weighed against the economic benefits of the designs considered. In some cases a wide range of designs can give similar rates of genetic progress, but often at widely differing costs. In such cases the economically optimum design may give slightly less than maximum genetic response.

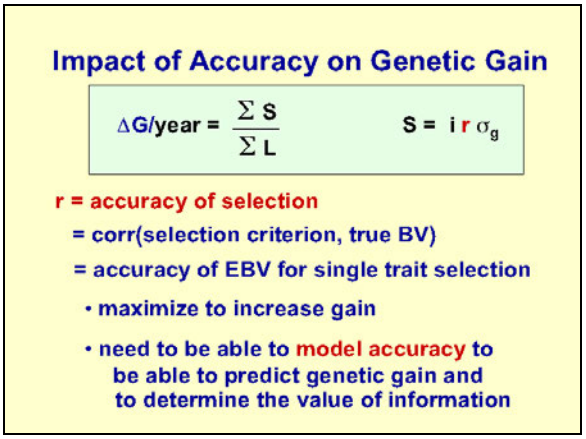
Chapter 4

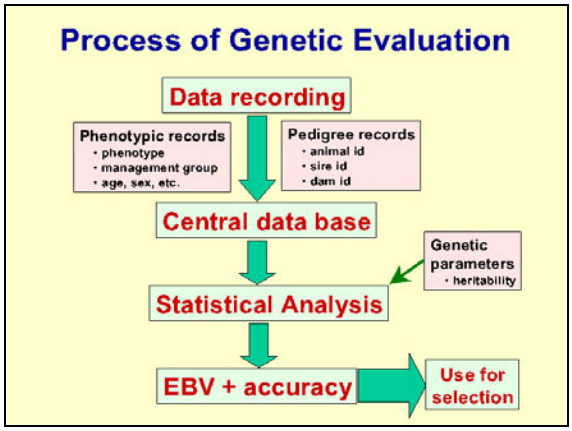
Deterministic Models for Estimated Breeding Values

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 (σ_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 and records on 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.





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:

$$r = r_{g,\hat{g}} = b_{g,\hat{g}} \frac{\sigma_{\hat{g}}}{\sigma_g} = \frac{\sigma_{\hat{g}}}{\sigma_g} \quad (4.1)$$

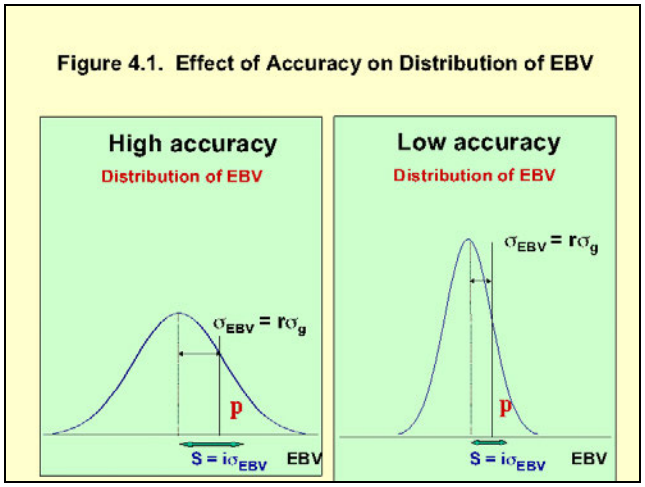
and the covariance between true and estimated BV as:

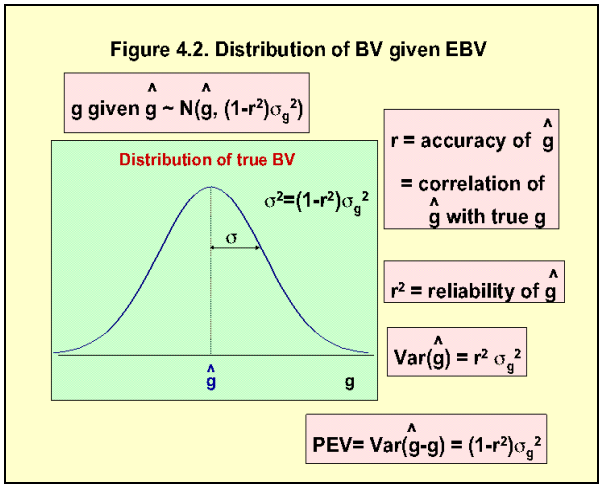
$$\sigma_{g,\hat{g}} = r_{g,\hat{g}} \sigma_g \sigma_{\hat{g}} = \sigma_g^2 \quad (4.2)$$

The variance of EBV is then equal to:

$$\sigma_{\hat{g}}^2 = r^2 \sigma_g^2 \quad (4.3)$$

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:

$$\sigma_\varepsilon^2 = \text{var}(g_i - \hat{g}_i) = \sigma_g^2 + \sigma_{\hat{g}}^2 - 2\sigma_{g,\hat{g}} = \sigma_g^2 + \sigma_g^2 - 2\sigma_g^2$$

$$= \sigma_g^2 - \sigma_g^2 = \sigma_g^2 - r^2\sigma_g^2$$

$$= (1-r^2)\sigma_g^2 \qquad (4.4)$$

Note that $\sigma_g^2 = \sigma_{\hat{g}}^2 + \sigma_\varepsilon^2$

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: $\sigma_{\hat{g},\varepsilon} = \sigma_{\hat{g},\hat{g}-g} = \sigma_{\hat{g}}^2 - \sigma_{g,\hat{g}} = \sigma_{\hat{g}}^2 - \sigma_{\hat{g}}^2 = 0$

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.

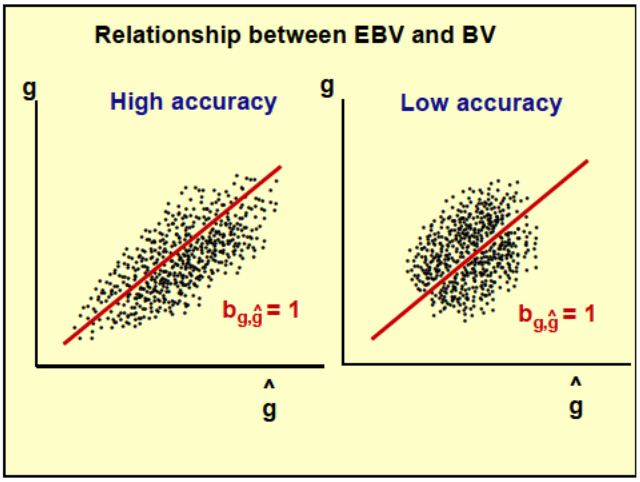
Given an animal's EBV and assuming normality, the animal's *true* BV is expected to follow a Normal distribution with mean equal to the EBV and variance equal to $(1-r^2)\sigma_g^2$:

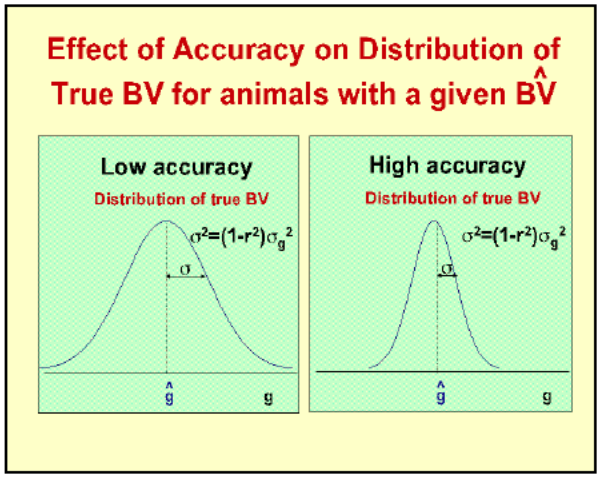
$$g_i | \hat{g}_i \sim N(\hat{g}_i, (1-r^2)\sigma_g^2) \qquad (4.5)$$

This distribution is illustrated in Figure 4.2.

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

$$\varepsilon_i \sim N(0,(1\text{-}r^2)\sigma_g^2) \qquad (4.6)$$





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:

$$\hat{g}_i = b_{g,x}\, x_i = b_{g,x}\,(\text{phenotype of individual}) \qquad (4.7)$$

The regression coefficient can be derived as:

$$b_{g,x} = \sigma_{x_i g_i}/\sigma_p^2 = \sigma_{g_i+e_i g_i}/\sigma_p^2 = \sigma_g^2/\sigma_p^2 = h^2 \qquad (4.8)$$

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

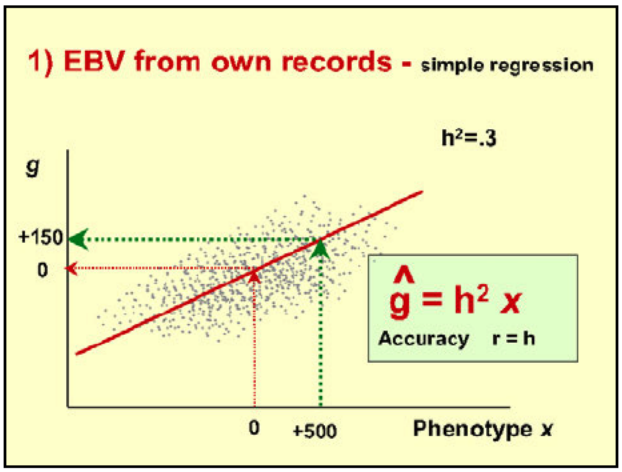
$$\hat{g}_i = h^2 x_i \qquad (4.9)$$

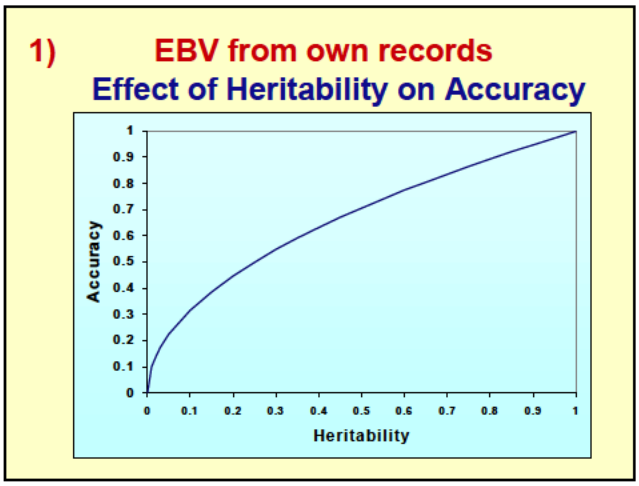
where x_i is the phenotype of individual i expressed as a deviation from the population mean.

The accuracy of selection is: $r = r_{g,\hat{g}} = \sigma_{\hat{g}_i h^2 x_i}/\sigma_g \sigma_{h^2 x} = h^2\sigma_g^2/h\sigma_g^2 = h$ (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:

$$x_{ij} = g_i + pe_i + te_{ij} \qquad (4.11)$$

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,

$$t = \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_p^2} \quad \text{or} \quad \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_g^2 + \sigma_{pe}^2 + \sigma_{te}^2} \tag{4.12}$$

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 $r_{x_1x_2} = \dfrac{\sigma_{x_1x_2}}{\sqrt{\sigma_{x_1}^2 \sigma_{x_2}^2}}$

where
$$\sigma_{x_1x_2} = \sigma_{(g_i + pe_i + te_{i1},\, g_i + pe_i + te_{i2})}$$
$$= \sigma_g^2 + \sigma_{pe}^2$$

Hence, $$r_{x_1x_2} = \frac{\sigma_g^2 + \sigma_{pe}^2}{\sigma_p^2} = t$$

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,

$$\hat{g}_i = b_{g\bar{x}}\, \bar{x}_i \tag{4.13}$$

where
$$\bar{x}_i = \sum_{j=1}^{m} x_{ij} / m \tag{4.14}$$

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

$$\bar{x}_i = \sum_{j=1}^{m} (g_i + pe_i + te_{ij}) / m \tag{4.15}$$

Then, $$b_{g\bar{x}} = \sigma_{g\bar{x}} / \sigma_{\bar{x}}^2$$

The variance of $\bar{x}_i$ is: $$\sigma_{\bar{x}}^2 = \sigma_g^2 + \sigma_{pe}^2 + \frac{\sigma_{te}^2}{m} = t\sigma_p^2 + \frac{(1-t)\sigma_p^2}{m} = \frac{(mt + 1 - t)\sigma_p^2}{m}$$

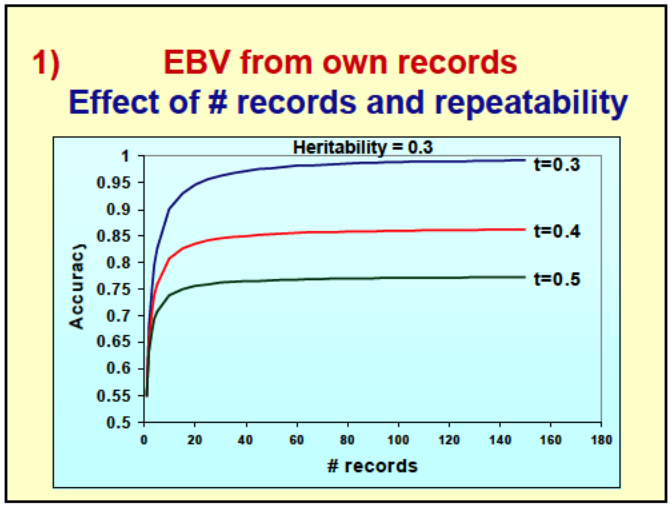
$$= \frac{((m-1)t + 1)\sigma_p^2}{m} \tag{4.16}$$

The covariance is: $$\sigma_{g\bar{x}} = \sigma_g^2 \tag{4.17}$$

Thus,
$$b_{g\bar{x}} = \frac{m\sigma_g^2}{\sigma_p^2\left((m-1)t+1\right)} = \frac{mh^2}{(m-1)t+1} \qquad (4.18)$$

And accuracy of selection is given by: $r = corr_{g\bar{x}} = \sqrt{\frac{mh^2}{(m-1)t+1}\frac{\sigma_g^2}{\sigma_g^2}} = \sqrt{\frac{mh^2}{(m-1)t+1}}$ (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

$$b_{g\bar{x}} = 0.25,\ 0.333,\ 0.42 \text{ or } 0.45$$

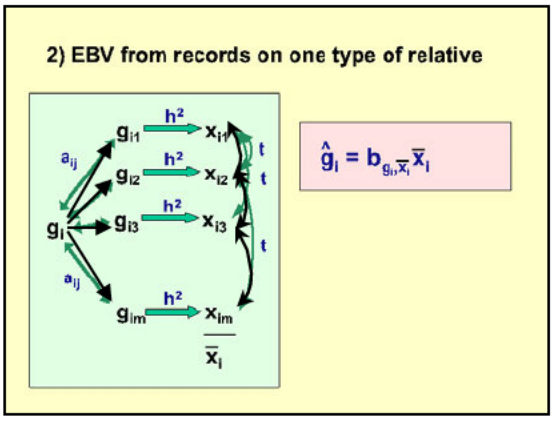
and accuracies of

$$r = 0.5\ ,\ 0.58\ ,\ 0.65 \text{ or } 0.67.$$

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

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

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

Then,

$$b_{g\bar{x}} = \sigma_{g\bar{x}} / \sigma^2_{\bar{x}}$$

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

$$\begin{aligned} t = r_{x_{ij}x_{ij'}} &= \sigma_{x_{ij}x_{ij'}} / \sigma^2_p = \sigma_{(g_{ij}+e_{ij},\, g_{ij'}+e_{ij'})} / \sigma^2_p \\ &= (a_{jj'}\sigma^2_g + c^2\sigma^2_p)/\sigma^2_p \\ &= a_{jj'}h^2 + c^2 \end{aligned} \qquad (4.20)$$

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):

$$c^2 = \sigma_{e_{ij}e_{ij'}} / \sigma^2_p \qquad (4.21)$$

As an aside, note that this equation for the intra-class correlation also holds for repeated own records. In that case, $a_{jj'}=1$, $c^2=\sigma^2_{pe}/\sigma^2_p$, and thus $t = h^2 + \sigma^2_{pe}/\sigma^2_p = (\sigma^2_g+\sigma^2_{pe})/\sigma^2_p$, 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:

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

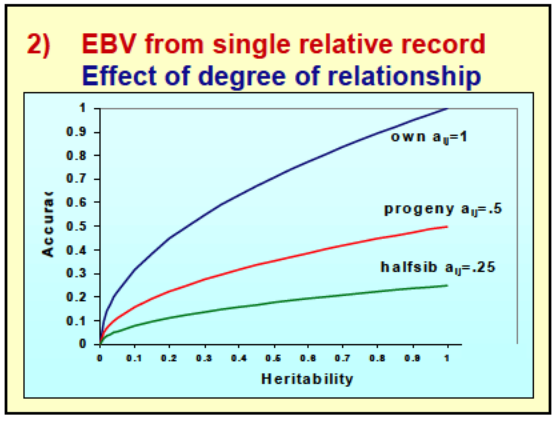
The covariance is:

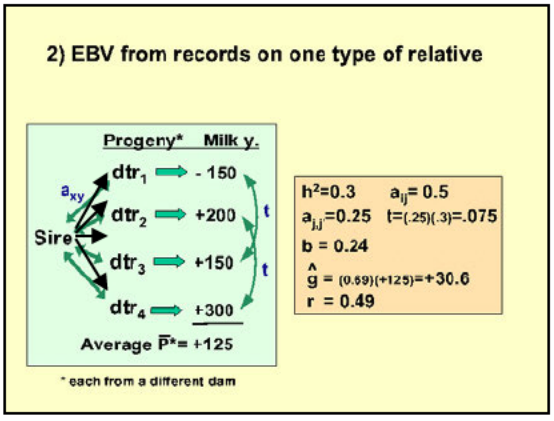
$$\sigma_{g\bar{x}} = a_{ij}\sigma^2_g \qquad (4.23)$$

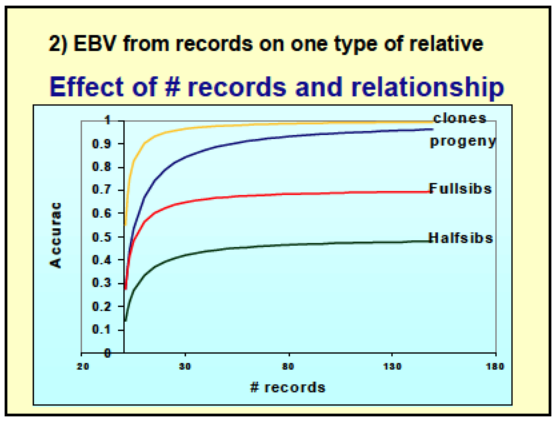
Thus,
$$b_{g\bar{x}} = \frac{m a_{ij} \sigma_g^2}{\sigma_p^2\left((m-1)t+1\right)} = a_{ij}\frac{mh^2}{(m-1)t+1} \qquad (4.24)$$

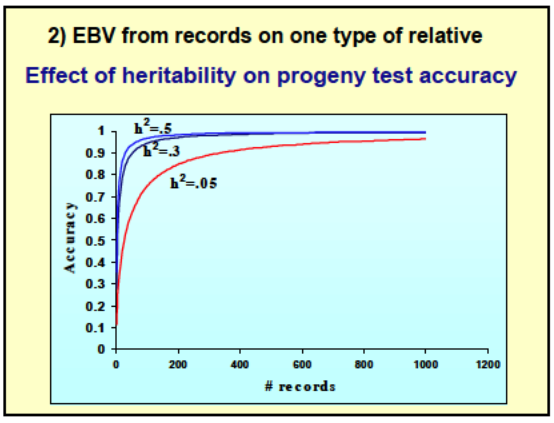
And, accuracy of selection is given by,
$$r = corr_{g\bar{x}} = a_{ij}\sqrt{\frac{mh^2}{(m-1)t+1}} \qquad (4.25)$$

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









4.4 EBV from Multiple Sources - Selection Index

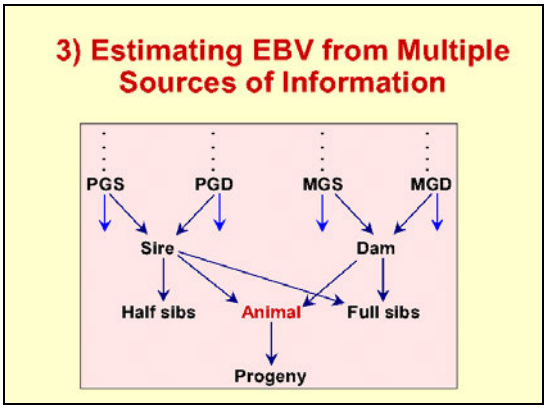
When records are available from multiple sources, e.g. records on the animal itself, its dam, half-sibs, progeny, etc., it will obviously be most beneficial to use all records to estimate the breeding value. This can be achieved by extending the simple regression methods described in the previous to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

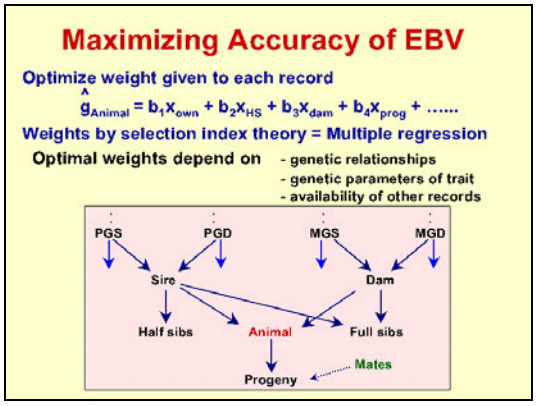
$$\hat{g}_i = b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \qquad (4.26)$$

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.





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

$$H = v_1 g_1 + v_2 g_2 + \ldots + v_n g_n \quad = \mathbf{v'g} \tag{4.27}$$

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, $\mathbf{v'} = [\, v_1 \,,\, v_2 \,,\, \ldots \,,\, v_n]$ and $\mathbf{g'} = [\, g_1 \,,\, g_2 \,,\, \ldots \,,\, g_n]$.

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,

$$I = b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \quad = \mathbf{b'x} \tag{4.28}$$

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, $\mathbf{b'} = [\, b_1 \,,\, b_2 \,, \ldots \,,\, b_m]$ and $\mathbf{x'} = [\, x_1 \,,\, x_2 \,,\, \ldots \,,\, x_m]$. 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 (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 $\mathbf{v} = [1]$. In this case the selection index is equal to the EBV for the trait evaluated:

$$I = \hat{g} = b_1 x_1 + b_2 x_2 + \ldots + b_m x_m \tag{4.29}$$

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 $\mathbf{v} = [1]$.

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

$$S_H = b_{H.I}(I - \bar{I}) \tag{4.30}$$

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

$$S_H = b_{HI}\, i\sigma_I = \frac{\sigma_{HI}}{\sigma_I^2} i\sigma_I = i\sigma_{HI}/\sigma_I \tag{4.31}$$

Thus for any given intensity of selection, i, response in H is maximized when σ_{HI}/σ_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

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

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 σ_{HI}/σ_I, subject to $b_{HI} = 1$.

Considering first the maximization of σ_{HI}/σ_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

$$\sigma_I^2 = b_1^2\sigma_{p11} + b_2^2\sigma_{p22} + \ldots + 2b_1b_2\sigma_{p12} + 2b_1b_3\sigma_{p13} \ldots = \sum_{k=1}^{m} \sum_{l=1}^{m} b_k\, b_l \sigma_{p_{ki}} \tag{4.33}$$

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

$$\sigma_{HI} = b_1v_1\sigma_{g11} + b_1v_2\sigma_{g12} + \ldots + b_mv_n\sigma_{g_{mn}} = \sum_{k=1}^{m} \sum_{l=1}^{m} b_k\, v_i \sigma_{g_{ki}} \tag{4.34}$$

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

then $\log M = \log \sigma_{HI} - \log \sigma_I$

or $\log M = \log \sigma_{HI} - \tfrac{1}{2} \log \sigma_I^2$

and substituting from (4.33) and (4.34):

$$\log M = \log(\sum\sum b_k v_i \sigma_{g_{ki}}) - \tfrac{1}{2}\log(\sum\sum b_k b_l \sigma_{p_{ki}}) \qquad (4.35)$$

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:

$$\frac{\delta \log M}{\delta b_k} = 0 \qquad \text{for } k = 1 \text{ to } m.$$

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

$$\frac{\delta \log M}{\delta b_k} = \frac{\sum_{i=1}^{n} v_i \sigma_{g_{ki}}}{\sigma_{HI}} - \frac{\sum_{l=1}^{m} b_l \sigma_{p_{kl}}}{\sigma_I^2}$$

Hence, M is maximal when $$\sum_{l=1}^{m} b_l \sigma_{p_{kl}} = \frac{\sigma_I^2}{\sigma_{HI}} \sum_{i=1}^{n} v_i \sigma_{g_{ki}} \qquad (4.36)$$

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,

$$\sum_{l=1}^{m} b_l \sigma_{p_{kl}} = \sum_{i=1}^{n} v_i \sigma_{g_{ki}} \qquad (4.37)$$

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

i.e.

$$\sum_{l=1}^{m} b_l \sigma_{p_{1l}} = \sum_{i=1}^{n} v_i \sigma_{g_{1i}}$$

$$\sum_{l=1}^{m} b_l \sigma_{p_{2l}} = \sum_{i=1}^{n} v_i \sigma_{g_{2i}}$$

$$\vdots \qquad \vdots$$

$$\sum_{l=1}^{m} b_l \sigma_{p_{ml}} = \sum_{i=1}^{n} v_i \sigma_{g_{mi}}$$

If we write these equations in their expanded form, i.e.

$$\begin{aligned}
b_1\sigma_{p_{1.1}} + b_2\sigma_{p_{1.2}} \ \ldots \ + b_m\sigma_{p_{1.m}} &= v_1\sigma_{g_{1.1}} + v_2\sigma_{g_{1.2}} \ \ldots \ + v_n\sigma_{g_{1.n}} \\
b_1\sigma_{p_{2.1}} + b_2\sigma_{p_{2.2}} \ \ldots \ + b_m\sigma_{p_{2.m}} &= v_1\sigma_{g_{2.1}} + v_2\sigma_{g_{2.2}} \ \ldots \ + v_n\sigma_{g_{2.n}} \\
&\ \ \vdots \qquad\qquad\qquad\quad \vdots \\
b_1\sigma_{p_{m.1}} + b_2\sigma_{p_{m.2}} \ \ldots \ + b_m\sigma_{p_{m.m}} &= v_1\sigma_{g_{m.1}} + v_2\sigma_{g_{m.2}} \ \ldots \ + v_n\sigma_{g_{m.n}}
\end{aligned}$$

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

$$\mathbf{Pb = Gv} \tag{4.38}$$

where: **b** = column vector of *m* selection index coefficients
P = *m* x *m* matrix of phenotypic covariances among the observations in the index,
G = *m* x *n* matrix of genetic covariances among the *m* index observations and the *n* traits in the aggregate genotype
v = 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, $\mathbf{P}^{-1}\,\mathbf{P} = \mathbf{I}$, the solution to obtaining **b** can be obtained by pre-multiplying both sides of (4.38) by $\mathbf{P}^{-1}$ to obtain,

$$\mathbf{b = P^{-1}Gv} \tag{4.39}$$

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 Var(*H-I*). Thus we wish to minimize

$$\begin{aligned}
E(H - I)^2 &= E[I - H)'\,(I - H)] \\
&= E[I - H)'\,(I - H)'] \\
&= E[(\mathbf{b'x - v'g})(\mathbf{x\,b - g'v})] \\
&= E[(\mathbf{b'xx'b - b'xg'v - v'gx\,b + v'gg'v}]
\end{aligned}$$

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:

$$\begin{aligned}
E(\mathbf{b'xx'b}) &= \mathbf{b'}E(\mathbf{xx'})\mathbf{b} = \mathbf{b'Pb}, \\
E(\mathbf{b'xg'v}) &= \mathbf{b'}E(\mathbf{xg'})\mathbf{v} = \mathbf{b'Gv}, \\
E(\mathbf{v'gx\,b}) &= \mathbf{v'G\,b} = \mathbf{b'Gv} \qquad \text{since } \mathbf{v'G'b} \text{ is a scalar}
\end{aligned}$$

and $\quad E(\mathbf{v'gg'v}) = \mathbf{v'}E(\mathbf{gg'})\mathbf{v} = \mathbf{v'Cv}$

Therefore, to minimize $M = \mathbf{b'Pb - 2b'Gv + v'Cv}$

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

Therefore $\mathbf{Pb = Gv}$

Hence, $\mathbf{b = P^{-1}Gv}$ 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, σ_I^2, is easily found as

$$\begin{aligned}\sigma_I^2 &= Var(b_1x_1 + b_2x_2 \ldots b_mx_m) \\ &= b_1^2\sigma_{p_1}^2 + b_2^2\sigma_{p2}^2 + \ldots + 2b_1b_2\sigma_{p12} + 2b_1b_3\sigma_{p13}\end{aligned}$$

or in matrix notation: $\sigma_I^2 = Var(\mathbf{b'x}) = \mathbf{b'}\ Var(\mathbf{x})\mathbf{b} = \mathbf{b'Pb}$ (4.41)

Following the same argument as for σ_I^2, $\sigma_H^2\ Var(\mathbf{v'g}) = \mathbf{v'}\ Var(\mathbf{g})\mathbf{v} = = \mathbf{v'Cv}$ (4.42)

where **C** is an *n* x *n* matrix of genetic covariances among the traits in the aggregate genotype.

Similarly, it follows that $\sigma_{HI} = Cov(\mathbf{b'x}, \mathbf{v'g}) = \mathbf{b'}\ Cov(\mathbf{x,g})\mathbf{v} = \mathbf{b'Gv}$ (4.43)

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

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

and from equations (4.41) and (4.43), $\mathbf{b'Pb = b'Gv}$ (4.45)

Thus, for the optimal index, equation (4.44) for accuracy simplifies to:

$$r_{HI} = \frac{\sigma_I}{\sigma_H} = \sqrt{\frac{\mathbf{b'Pb}}{\mathbf{v'Cv}}} = \sqrt{\frac{\mathbf{b'Gv}}{\mathbf{v'Cv}}} \tag{4.46}$$

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 $H = g$, $\mathbf{v} = [1]$, and $\sigma_H^2 = \sigma_g^2$. This simplifies derivations to:

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

and from equation (4.46) $r_{HI} = r_{g,g} = \sqrt{\dfrac{\mathbf{b'G}}{\sigma_\mathbf{g}^2}}$ (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 $I = \hat{g} = b_1 x_1$ and the aggregate genotype is $H = g$

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

In this case, $\mathbf{P} = \sigma_{\bar{x}}^2$ and $\mathbf{G} = \sigma_{g\bar{x}}$

Hence, $\mathbf{b} = b = \mathbf{P}^{-1}\mathbf{G} = (\sigma_{\bar{x}}^2)^{-1}\sigma_{g\bar{x}} = \sigma_{g\bar{x}} / \sigma_{\bar{x}}^2$

The accuracy of selection, given by (4.48), is $r_{HI} = r_{g,g} = \sqrt{\dfrac{\mathbf{b'G}}{\sigma_\mathbf{g}^2}} = \sqrt{\dfrac{b\mathbf{G}}{\mathbf{C}}} = \dfrac{\sigma_{g\bar{x}}}{\sigma_{\bar{x}}\sigma_g}$

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.

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

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
- σ_{p_k} = phenotypic standard deviation of trait k
- σ_{g_k} = additive genetic standard deviation of trait k
- $r_{p_{kl}}$ = phenotypic correlation between traits k and l
- $r_{g_{kl}}$ = genetic correlation between traits k and l
- a = additive genetic relationship within a group
- a_{ij} = additive genetic relationship between individual(s) in groups i and j
- a_{hj} = additive genetic relationship between the individual in the breeding goal (h) and individuals in group j

P-matrix

diagonal:

• Variance of *m* records of a given type

$$\frac{1+(m-1)t}{m}\sigma_p^2 \qquad (=\sigma_p^2 \text{ for } m=1) \tag{4.49}$$

with t = repeatability for repeated records

$t = ah^2 + c^2$ for multiple individuals

off-diagonal:

• Covariance between mean of *m* records on different traits (*k* and *l*) for the same group:

$$\frac{r_{p_{kl}}\sigma_{p_k}\sigma_{p_l}+(m-1)a\,r_{g_{kl}}\sigma_{g_k}\sigma_{g_l}}{m} \qquad (= r_{p_{kl}}\sigma_{p_k}\sigma_{p_l} \text{ for } m=1) \tag{4.50}$$

• Covariance between (mean of) record(s) on same trait *k* for different groups (*i* and *j*):

$$(a_{ij}\,h_k^2 + c_k^2)\sigma_{p_k}^2 \tag{4.51}$$

• Between records on different traits (*k* and *l*) in different groups (*i* and *j*):

$$a_{ij}r_{g_{kl}}\sigma_{g_k}\sigma_{g_l} \tag{4.52}$$

G-matrix

• Covariance of the genetic value for trait *k* on the breeding goal animal (*h*) with records on trait *l* for group *j*

$$a_{hj}r_{g_{kl}}\sigma_{g_k}\sigma_{g_l} \qquad (=a_{hj}\sigma_{g_k}^2 \text{ if } k=l) \tag{4.53}$$

C-matrix

Diagonal:

• Variance of genetic value for trait *k*

$$\sigma_{g_k}^2 \tag{4.54}$$

Off-diagonal:

• Covariance between genetic values for traits *k* and *l* on breeding goal animal

$$r_{g_{kl}}\sigma_{g_k}\sigma_{g_l} \tag{4.55}$$

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,

$$I = \hat{g} = b_1x_1 + b_2x_2$$

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,

$$\mathbf{P} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1x_2} \\ \sigma_{x_1x_2} & \sigma^2_{x_2} \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} \sigma_{x_1 g} \\ \sigma_{x_2 g} \end{bmatrix} \tag{4.56}$$

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:
$$\mathbf{P} = \begin{bmatrix} 1 & \tfrac{1}{2}h^2 \\ \tfrac{1}{2}h^2 & \dfrac{1+(m-1)\tfrac{1}{2}h^2}{m} \end{bmatrix} \sigma^2_p = \begin{bmatrix} 1 & .25 \\ .25 & .4 \end{bmatrix} \sigma^2_p \tag{4.57}$$

and:
$$\mathbf{G} = \begin{bmatrix} h^2 \\ \tfrac{1}{2}h^2 \end{bmatrix} \sigma^2_p = \begin{bmatrix} .5 \\ .25 \end{bmatrix} \sigma^2_p \tag{4.58}$$

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

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

Hence, the selection index would be

$$I = \hat{g} = 0.4074\, x_1 + 0.3704\, x_2$$

The accuracy of this index or EBV is given by

$$r_{HI} = r_{g,g} = \sqrt{\frac{\mathbf{b'G}}{\sigma^2_{\mathbf{g}}}} = \sqrt{\frac{\begin{bmatrix} .4074 \\ .3704 \end{bmatrix}' \begin{bmatrix} .5 \\ .25 \end{bmatrix} \sigma^2_p}{0.5\,\sigma^2_p}} = 0.77 \tag{4.59}$$

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 $S = i r_{g,g}\, \sigma_g$, and i and σ_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:

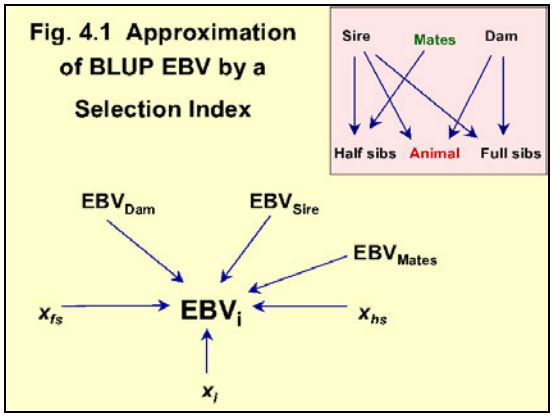
1) Develop a selection index based only on those relatives providing the greatest amount of information, rather than all possible relatives as in the animal model. For example, when records on parents, full and half sibs, and progeny are accounted for, information on more distant relatives may only provide 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 $(m\text{-}1)n$ 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}
- $\bar{\hat{g}}_m$ = 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:

$$I_i = \hat{g}_i = b_1 x_i + b_2 x_{fs} + b_3 x_{hs} + b_4 \hat{g}_s + b_5 \hat{g}_d + b_6 \bar{\hat{g}}_m \tag{4.60}$$

$$\mathbf{P} = \begin{bmatrix} \sigma^2_{x_i} & \sigma_{x_i x_{fs}} & \sigma_{x_i x_{hs}} & \sigma_{x_i \hat{g}_s} & \sigma_{x_i \hat{g}_d} & \sigma_{x_i \bar{\hat{g}}_m} \\ & \sigma^2_{x_{fs}} & \sigma_{x_{fs} x_{hs}} & \sigma_{x_{fs} \hat{g}_s} & \sigma_{x_{fs} \hat{g}_d} & \sigma_{x_{js} \bar{\hat{g}}_m} \\ & & \sigma^2_{x_{hs}} & \sigma_{x_{hs} \hat{g}_s} & \sigma_{x_{hs} \hat{g}_d} & \sigma_{x_{hs} \bar{\hat{g}}_m} \\ & & & \sigma^2_{\hat{g}_s} & \sigma_{\hat{g}_s \hat{g}_d} & \sigma_{\hat{g}_s \bar{\hat{g}}_m} \\ & & & & \sigma^2_{\hat{g}_d} & \sigma_{\hat{g}_d \bar{\hat{g}}_m} \\ & & & & & \sigma^2_{\bar{\hat{g}}_m} \end{bmatrix} \tag{4.61}$$

$$\mathbf{G} = \begin{bmatrix} \sigma_{g_i x_i} & \sigma_{g_i x_{fs}} & \sigma_{g_i x_{hs}} & \sigma_{g_i \hat{g}_s} & \sigma_{g_i \hat{g}_d} & \sigma_{g_i \bar{\hat{g}}_m} \end{bmatrix} \tag{4.62}$$

$$\mathbf{P} = \begin{bmatrix} 1 & {}^1/_2\,h^2 + c^2 & {}^1/_4\,h^2 & {}^1/_2\,r_s^2 h^2 & {}^1/_2\,r_d^2 h^2 & 0 \\ & \dfrac{1+(n-2)({}^1/_2\,h^2 + c^2)}{n-1} & {}^1/_4\,h^2 & {}^1/_2\,r_s^2 h^2 & {}^1/_2\,r_d^2 h^2 & 0 \\ & & {}^1/_4\,h^2 + \dfrac{{}^1/_4\,h^2 + c^2}{m-1} + \dfrac{1-{}^1/_2\,h^2 - c^2}{n(m-1)} & {}^1/_2\,r_s^2 h^2 & 0 & \dfrac{{}^1/_2\,r_m^2 h^2}{m-1} \\ & & & r_s^2 h^2 & 0 & 0 \\ & & & & r_d^2 h^2 & 0 \\ & & & & & \dfrac{r_m^2 h^2}{m-1} \end{bmatrix} \sigma_p^2 \quad (4.63)$$

$$\mathbf{G} = \begin{bmatrix} h^2 & {}^1/_2\,h^2 & {}^1/_4\,h^2 & {}^1/_2\,r_s^2\,h^2 & {}^1/_2\,r_d^2\,h^2 & 0 \end{bmatrix} \sigma_p^2 \quad (4.64)$$

With
$$x_{hs} = \left(\sum_{k=1}^{m-1} \sum_{l=1}^{n} \frac{x_{kl}}{n} \right) / (m\text{-}1) \quad (4.65)$$

Where
$$x_{kl} = {}^1/_2\,g_s + {}^1/_2\,g_{d_k} + g_{ms_{kl}} + c_{kl} + e_{kl} \quad (4.66)$$

Thus
$$x_{hs} = {}^1/_2\,g_s + \frac{\sum_{k=1}^{m-1} \left({}^1/_2\,g_{d_k} + c_k \right)}{m-1} + \frac{\sum_{k=1}^{m-1} \sum_{l=1}^{n} \left(g_{ms_{kl}} + e_{kl} \right)}{n(m-1)} \quad (4.67)$$

And
$$\sigma^2_{x_{hs}} = {}^1/_4\sigma_g^2 + \frac{{}^1/_4\sigma_g^2 + c^2\sigma_p^2}{m-1} + \frac{{}^1/_2\sigma_g^2 + \sigma_e^2}{n(m-1)} \quad (4.68)$$

Also,
$$\sigma^2_{\hat{g}} = r^2_{g,\hat{g}}\,\sigma_g^2 \quad (4.69)$$

And
$$\sigma_{x_i\,\hat{g}_s} = \sigma_{({}^1/_2\,g_s + {}^1/_2\,g_d + g_{m_i} + e_i, \hat{g}_s)} = \sigma_{({}^1/_2 g_s, \hat{g}_s)} = {}^1/_2\sigma_{g_s,\hat{g}_s} = {}^1/_2\,r_s^2\sigma_g^2 \quad (4.70)$$

As before, index weights can be derived as: $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}$

And accuracy as: $r_{g,\hat{g}} = \sqrt{\mathbf{b}'\mathbf{P}\mathbf{b}/\sigma_g^2}$

Because elements of the **P** and **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).

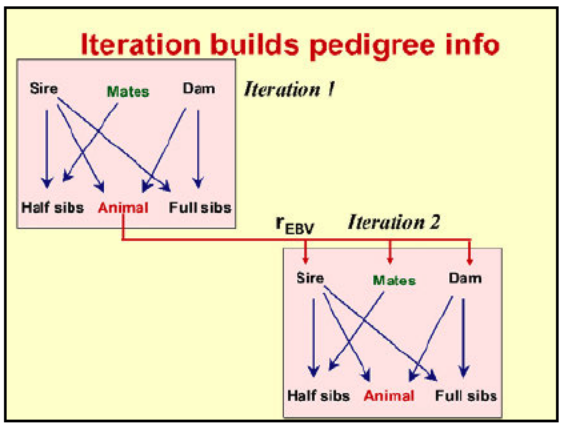
Iterative Procedure to Compute Accuracy of EBV
Building-up pedigree information

$EBV_i = b' [x_i , x_{fs} , x_{hs} , \hat{g}_{dam} , \hat{g}_{sire} , \hat{g}_{mates}]'$

$b = P^{-1}G \quad r_{\hat{g},g} = \sqrt{b'Pb}/\sigma_g$

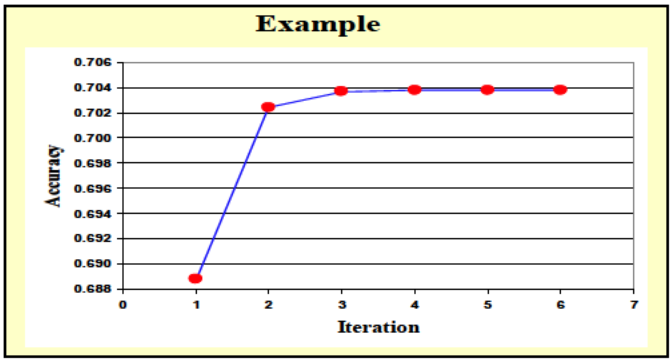
1) Set accuracy of $\hat{g}_{sire}=\hat{g}_{dam}=\hat{g}_{mates}= h$ (own record)
2) Set up index (P, G) and derive accuracy $r_{\hat{g},g}=\sqrt{b'Pb}/\sigma_g$
3) Set accuracy of $\hat{g}_{sire}=\hat{g}_{dam}=\hat{g}_{mates}$ equal to $r_{\hat{g},g}$
4) Repeat steps 2 and 3 until accuracy converges

Needs adaptation if selection after progeny



Example

Iteration 1	σ_p^2= 100		n= 10	n offspring per dam –> n-1 full sibs of individual					
	h^2= 0.25		m= 20	m dams per sire with n offspring each					
	c^2= 0			–> (m-1)n halfsibs consisting of (m-1) fullsib families					
STARTING VALUE	σ_g^2= 25								
FOR ACCURACY OF	σ_e^2= 75								
PARENTAL EBV									
$r_e=r_d=r_m$= 0.5000	P =	x_i	x_{fs}	x_{hs}	g_{sire}	g_{dam}	g_{mates}		
(start with r=h)	x_i	***100.00***	12.50	6.25	3.13	3.13	0.00	$b=P^{-1}G$=	0.175
	x_{fs}	12.50	***22.22***	6.25	3.13	3.13	0.00		0.291
	x_{hs}	6.25	6.25	***7.04***	3.13	0.00	0.16		0.464
	g_{sire}	3.13	3.13	3.13	***6.25***	0.00	0.00		0.035
	g_{dam}	3.13	3.13	0.00	0.00	***6.25***	0.00		0.267
	g_{mates}	0.00	0.00	0.16	0.00	0.00	***0.33***		-0.232
									0.6888
	G =	25.00	12.50	6.25	3.13	3.13	0.00	Acc =	
Iteration 2									
$r_e=r_d=r_m$= 0.6888	P =	x_i	x_{fs}	x_{hs}	g_{sire}	g_{dam}	g_{mates}		
from previous iteratio	x_i	***100.00***	12.50	6.25	5.93	5.93	0.00	$b=P^{-1}G$=	0.169
	x_{fs}	12.50	***22.22***	6.25	5.93	5.93	0.00		0.234
	x_{hs}	6.25	6.25	***7.04***	5.93	0.00	0.3		0.500
	g_{sire}	5.93	5.93	5.93	***11.86***	0.00	0.00		0.048
	g_{dam}	5.93	5.93	0.00	0.00	***11.86***	0.00		0.298
	g_{mates}	0.00	0.00	0.31	0.00	0.00	***0.62***		-0.250
								Acc =	0.7024
	G =	25.00	12.50	6.25	5.93	5.93	0.00		



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).

Chapter 5

Selection-Induced Gametic Phase Disequilibrium The Bulmer Effect

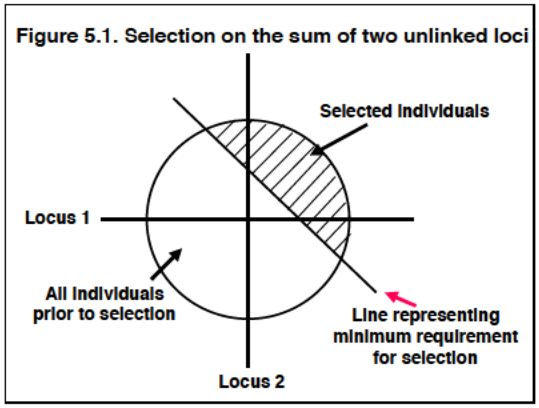
In the previous chapters, genetic variance was assumed constant over generations. Selection, however, has an impact not only on the mean of the population but also on genetic variance. Changes in genetic variance affect the amount of change that can be made in future generations.

The objective of this section is to model the effect of selection on genetic variance and to incorporate its effects in derivation of selection indexes and response to selection. As in the previous chapters, the basis of these models will be the infinitesimal genetic model, in which the trait is assumed to be affected by a large number of unlinked loci with small effect.

5.1 Effects of Selection on Genetic Variance

Pearson (1903), in his discussions on conditional variances early this century, noted that truncating a distribution affected both the mean and variance of the population. Anecdotally, founding animal breeders such as Lush, Falconer, and Henderson are said to have recognized that this could have implications for animal breeding since truncation selection could reduce genetic variance among parents from that observed before selection. Bulmer (1971, 1976, 1981) was the first to publish an examination of this effect of selection on genetic variance, and, consequently, the effect is often referred to as the "Bulmer Effect" or the effect of linkage disequilibrium. A more appropriate term is Falconer's "gametic phase disequilibrium" (See Falconer and Mackay, 1996, for an explanation of this term.)

To explain gametic phase disequilibrium, we will look at a situation where two unlinked genes affect a trait in an additive manner and each gene has a large number of alleles and both genes make the same contribution to genetic variance in the trait. Animals from a previously unselected population are selected on the sum of the genetic effects at both loci as illustrated in Figure 5.1.



When an animal has a high value at locus 1, it has a high chance of being selected, irrespective of the value at locus 2. Similarly, an animal with a high value at locus 2 will have a high value of being selected irrespective of the value at locus 1. However, animals with a moderately high value at locus 1 will only be selected when the value at locus 2 is also at least moderately high. As a consequence of this selection, effects at the two loci are negatively correlated in the selected individuals. In other words, the effects at the two loci in the selected individuals are no longer uncorrelated, i.e. selection has introduced ***gametic phase disequilibrium***.

Genetic variance in the trait is equal to variance of the sum of the gene effects at the two loci:

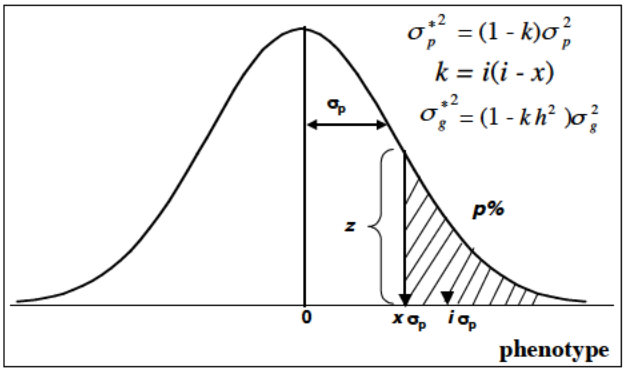
$$\sigma_g^2 = \sigma_{g_1}^2 + \sigma_{g_2}^2 + \sigma_{g_1 g_2} \qquad (5.1)$$

where $\sigma_{g_i}^2$ is the variance due to effects at locus *i* and $\sigma_{g_1 g_2}$ is the covariance between the effects at the two loci. Prior to selection, $\sigma_{g_1 g_2}$ is equal to zero, which reflects that the genes are in (linkage) equilibrium. Selection introduced a negative covariance and, as can be seen from equation (5.1), it is this negative covariance or disequilibrium between the two loci that reduces the genetic variance in the group of selected individuals.

It is important to recall that in the infinitesimal genetic model, individual genes are not recognized. The reduction in variance among selected individuals can also be derived from normal distribution theory, an approach that we will follow from here on. Nevertheless, it is important to keep in mind that the underlying mechanism that creates the reduction in genetic variance is the negative disequilibrium that is created between loci.

The effects of selection on genetic variance will first be described for a situation where animals are selected on their phenotype; a situation which is often referred to as "mass selection". The distribution of phenotypes prior to selection will have a standard deviation of σ_p, but as is clear from Figure 5.2, the standard deviation will be considerably less than among the proportion, p, of animals that are selected for breeding.

Figure 5.2 Mass Selection



The group of selected animals represents one tail of the distribution of the phenotypic distribution. If σ_p^2 is the phenotypic variance in the population before selection, *k* the factor by which the variance is reduced, and a subscript * is used to denote parameters after selection, the variance, σ_p^{*2}, in the selected individuals is:

$$\sigma_p^{*2} = (1 - k)\,\sigma_p^2 \qquad (5.2)$$

Factor *k* depends on intensity of selection (Pearson, 1903). When selection is by truncation of a normal distribution, then:

$$k = i(i - x) \qquad (5.3)$$

where *i* is the selection intensity and *x* is the standardized truncation point to the normal distribution corresponding to *i*, expressed in standard deviation units.

For genetic improvement, the question is what effect does selection on phenotype have on genetic variance of the trait. Again, from standard normal distribution theory it follows that with truncation selection on trait *y* the variance of a correlated trait *x* in the selected group, σ_x^{*2}, is given by

$$\sigma_x^{*2} = (1 - k\,r_{xy}^2)\,\sigma_x^2 \qquad (5.4)$$

where r_{xy} is the correlation between traits *x* and *y*.

Covariances between variables are similarly affected by selection. For example, the genetic covariance between *w* and *x* after selection on *y* is

$$\sigma_{wx}^* = \sigma_{wx} - k\frac{\sigma_{wy}\sigma_{xy}}{\sigma_y^2} \qquad (5.5)$$

Note that equation (5.4) for genetic variance is just a special case of (5.5) when $w = x$.

For mass selection, genetic variance among the selected individuals can be deduced as follows:

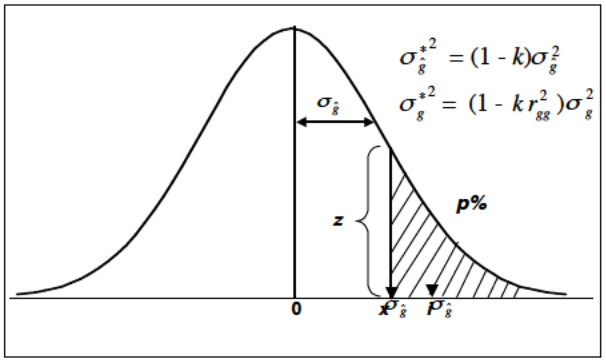
$$\begin{aligned}\sigma_g^{*2} &= (1 - k\,r_{gy}^2)\,\sigma_g^2 \qquad (5.6)\\ &= (1 - k\,h^2)\,\sigma_g^2\end{aligned}$$

where σ_g^2 is the genetic variance before selection. The correlation between additive genetic value *g* and phenotypic value, *y*, is *h*, the square of the heritability. The phenotypic variance is reduced by a factor *k* and the proportion h^2 of σ_g^2 is reduced by that same factor.

The formulae to calculate the reduction in genetic variance will now be generalized to a situation where selection is based on estimated breeding value $\hat{g}$ (Figure 5.3). Genetic variance among selected individuals can be derived using the correlation between the true genetic value *g* and the EBV, $\hat{g}$, which is equal to $r_{g,g}$. When selection is on $\hat{g}$, the variance in $\hat{g}$ among the selected animals is $\sigma_g^{*2} = (1 - k)\,\sigma_g^2$ and from (5.4) it follows that the genetic variance among the selected animals is

$$\sigma_g^{*2} = (1 - k\,r_{gg}^2)\,\sigma_g^2 \qquad (5.7)$$

Figure 5.3 EBV Selection



Referring back to Chapter 2, genetic variance of a population prior to selection can be partitioned into the parental and Mendelian sampling components as:

$$\sigma^2_g = \tfrac{1}{4}\sigma^2_{g_s} + \tfrac{1}{4}\sigma^2_{g_d} + \sigma^2_{g_m} \qquad (5.8)$$

This can now be modified to give genetic variance after selection among sires and dams. Using (5.6), genetic variance in the selected sires and dams can be calculated, where $\sigma^{*2}_{g_s}$ is the genetic variance among the selected sires and $\sigma^{*2}_{g_d}$ is the genetic variance among the selected dams. This leads to:

$$\sigma^2_g = \tfrac{1}{4}\sigma^{*2}_{g_s} + \tfrac{1}{4}\sigma^{*2}_{g_d} + \sigma^2_{g_m} \qquad (5.9)$$

This can be generalized to predict the genetic variance in generation t+1 from the variance among the parents selected in generation t

$$\sigma^2_{g_{(t+1)}} = \tfrac{1}{4}\sigma^{*2}_{g_{s(t)}} + \tfrac{1}{4}\sigma^{*2}_{g_{d(t)}} + \sigma^2_{g_m} \qquad (5.10)$$

Note that only the parental contributions to variation are affected by selection. The variance generated by Mendelian sampling, $\sigma^2_{g_m}$, is unaffected by selection and is equal to $\tfrac{1}{2}\sigma^2_{g_{(o)}}$ where $\sigma^2_{g_{(o)}}$ is the genetic variance in the unselected and non-inbred base population. The intuitive reasoning for this is that Mendelian sampling variance represents variation created by sampling one of a pair of parental alleles at each locus. This sampling process is unaffected by selection. Mendelian sampling variance is, however, affected by inbreeding, which will be discussed later.

Based on this, the following general recursive equation can be developed to predict genetic variance among progeny:

$$\sigma^2_{g_{(t+1)}} = \tfrac{1}{4}(1 - k_s r^2_{s_{(t)}})\sigma^2_{g_{(t)}} + \tfrac{1}{4}(1 - k_d r^2_{d_{(t)}})\sigma^2_{g_{(t)}} + \tfrac{1}{2}\sigma^2_{g_{(o)}} \qquad (5.11)$$

where k_s and k_d are based on selection intensities among males and females, and $r_{s_{(t)}}$ and $r_{d_{(t)}}$ are the respective accuracies of selection in generation t.

5.2 Prediction of Genetic Variance and Response for Mass Selection

In Table 5.1, the genetic variance is given for different (4) generations of mass selection in males and females. Generation 0 is assumed to be unselected, h^2=½, and $\sigma^2_e = \sigma^2_{g_{(o)}}$ =100. Truncation selection is used in both males and females and 5% of animals with highest phenotype are selected. In that case: i = 2.063 and x = 1.645, which based on equation (5.3) results in $k = i(i-x)$ = 2.063(2.063-1.645) = 0.862. Using equation (5.6), genetic variance among selected parents (sires and dams) is (1-0.862x½)100 = 56.9.

From equation (5.11) it follows that genetic variance in generation 1 is equal to ¼x56.9 + ¼x56.9 + ½x50 = 78.45. Selection reduced genetic variance to 78.45. In the base population, σ^2_e was 100 and the level of this variance is not affected by selection. Heritability in generation 1 is now 78.45/(100+78.45)=0.44. With this new level of h^2, variance among parents selected in generation 1 can be calculated using (5.6) and variance in generation 2 using (5.11).

Table 5.1 Effect of truncation selection with p=5% in males and females (i=2.063, x=1.645) during 5 generations (t = 0 to 4) on additive genetic variance $\sigma^2_{g_{(t)}}$ and average additive genetic merit of individuals ($\bar{g}_{(t)}$). Heritability in generation 0 was ½ (no inbreeding).

t	$\sigma^2_{g_{(t)}}$	$h^2_{(t)}$	$\bar{g}_{(t)}$	$\bar{g}_{(t)} - \bar{g}_{(t-1)}$
0	100	0.50	50.0	0
1	78	0.43	64.6	14.6
2	74	0.43	76.7	12.1
3	74	0.42	88.3	11.6
4	73	0.42	99.8	11.5
5	73	0.42	111.3	11.5
Selection stopped (random selection from here on)				
6	87	0.47	111.3	0
7	93	0.48	111.3	0
8	97	0.49	111.3	0
9	98	0.49	111.3	0
10	99	0.49	111.3	0

From Table 5.1 it can be seen that genetic variance reaches an equilibrium after three generations of selection. Genetic variance is equal to 74 and does not decrease further although selection is continued. This is referred to as the asymptotic genetic variance. When this is reached, the amount of gametic phase disequilibrium created by selection of individuals is equal to the amount of gametic phase disequilibrium which is broken down during meiosis (Mendelian sampling). When selection is stopped after four generations, no new gametic phase disequilibrium is created in the parents and the variance reduction is halved each generation as a result of Mendelian sampling. After 10 generations, genetic variance is back to its original level.

Response to selection from one generation to the next can be predicted as derived in chapter 1, but using parameters that apply to the parental generation:

$$\bar{g}_{(t+1)} = \bar{g}_{(t)} + ih_{(t)}\sigma_{g_{(t)}} \qquad (5.12)$$

The mean of the population changes as a result of selection. After five generations of selection the population level has increased by 111.3 units (Table 5.10. The greatest genetic gain was realized in generation 1 because this was the generation with the highest h^2 and genetic variance. Response in subsequent generations is reduced both because of a reduction in genetic variance, as well as a result of a reduction in accuracy of selection. The population remains at the same level after selection has stopped.

Genetic variance in the population is reduced by 26% after one round of selection but this is the result of a much larger, i.e. 52%, reduction in variance among selected sires and dams. This results from the fact that variance due to Mendelian sampling is not affected by selection and consequently remains 50. Another way to look at this is to consider variation within and between full sib families. Without selection, between and within family variances are both equal to 50. With selection, variation between full sib families is equal to $\frac{1}{4}\sigma^{*2}_{g_{s(t)}} + \frac{1}{4}\sigma^{*2}_{g_{d(t)}}$, while the within full sib family genetic variance is equal to $\sigma^2_{g_m} = \frac{1}{2}\sigma^2_{g_{(o)}}$. In generation 1, the between full-sib genetic variance is equal to ¼x56.9+¼x56.9=28.45, while the within full-sib variance remains equal to 50. This demonstrates that selection has changed the ratio of within and between family genetic variance. An implication of this is that using a reduced heritability in deriving selection index weights is not the correct way to deal with changes in genetic variance resulting from selection because this assumes that all components of genetic variance are affected in the same way, which is not true, as we have seen in equation (5.11). Mass selection is a special case in which we only use the observations on the individual and forms an exception to this rule.

5.2.1 Asymptotic Genetic Variance and Response to Selection

The previous enables recursive prediction of changes in variance and response to selection. Both variance and response reach steady state or asymptotic values after a number of generations. For the case of mass selection (and BLUP selection as we will see later), these steady state parameters can also be derived directly, as will be demonstrated below.

Starting with recursive equation (5.11): $\sigma^2_{g_{(t+1)}} = \frac{1}{4}(1-k_s r^2_{s_{(t)}})\sigma^2_{g_{(t)}} + \frac{1}{4}(1-k_d r^2_{d_{(t)}})\sigma^2_{g_{(t)}} + \frac{1}{2}\sigma^2_{g_{(o)}}$, steady state parameters (denoted by subscript $_{(L)}$) can be derived by setting $\sigma^2_{g_{(L)}} = \sigma^2_{g_{(t+1)}} = \sigma^2_{g_{(t)}}$, $r_{s_{(L)}} = r_{s_{(t)}}$, and $r_{d_{(L)}} = r_{d_{(t)}}$, which results in the following steady-state equation:

$$\sigma^2_{g_{(L)}} = \frac{1}{4}(1-k_s r^2_{s_{(L)}})\sigma^2_{g_{(L)}} + \frac{1}{4}(1-k_d r^2_{d_{(L)}})\sigma^2_{g_{(L)}} + \frac{1}{2}\sigma^2_{g_{(o)}} \qquad (5.12)$$

This equation can be solved if an equation can be developed that expresses accuracy of selection at the limit, $r_{s_{(L)}}$ and $r_{d_{(L)}}$, in terms of $\sigma^2_{g_{(L)}}$ and base population parameters. This is possible for mass selection and, as will be shown later, also for selection on BLUP EBV but not in general for selection on other types of selection indexes.

For mass selection, $r_{s_{(t)}} = r_{d_{(t)}} = h_{(t)}$, and assuming for simplicity equal selected fractions in both sexes, thus $k_s=k_d=k$, equation (5.11) simplifies to the following the recursive equation:

$$\sigma^2_{g_{(t+1)}} = \tfrac{1}{2}(1 - k\, h^2_{(t)})\,\sigma^2_{g_{(t)}} + \tfrac{1}{2}\,\sigma^2_{g_{(o)}} \qquad (5.13)$$

and at the limit, from (5.12): $\sigma^2_{g_{(L)}} = \tfrac{1}{2}(1 - k\, h^2_{(L)})\,\sigma^2_{g_{(L)}} + \tfrac{1}{2}\,\sigma^2_{g_{(o)}}$ (5.14)

with $h^2_{(L)} = \sigma^2_{g_{(L)}} / (\sigma^2_{g_{(L)}} + \sigma^2_e)$ (5.15)

Using $\sigma^2_e = \dfrac{1 - h^2_{(0)}}{h^2_0}\,\sigma^2_{g_{(o)}}$ (5.16)

steady state heritability can be solved in terms of the base population heritability as:

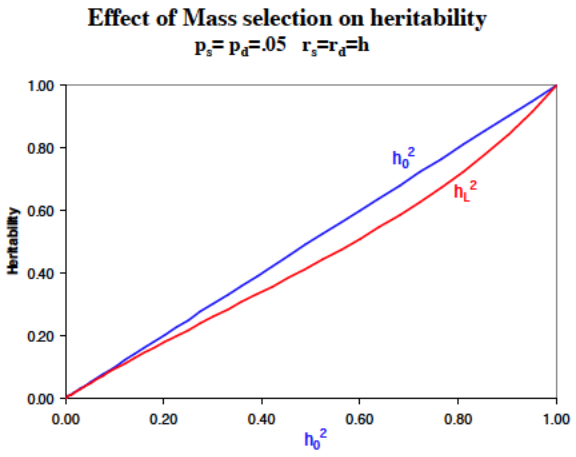
$$h^2_{(L)} = \frac{h^2_{(0)}}{1 + (1 - h^2_{(0)})\,k h^2_{(L)}} = \frac{-1 + \sqrt{1 + 4h^2_{(0)}k(1 - h^2_{(0)})}}{2k(1 - h^2_{(0)})} \qquad (5.17)$$

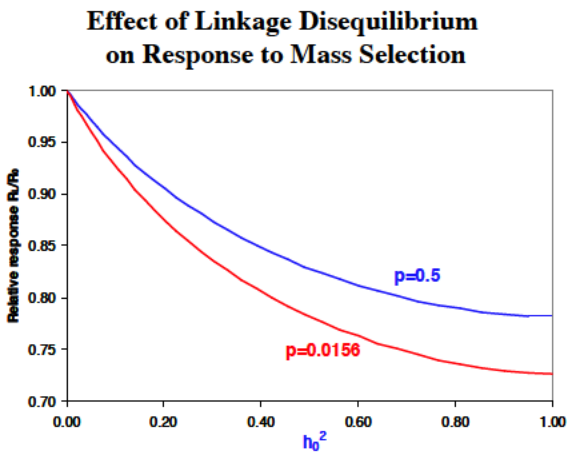
Substituting into equation (5.14) gives the following expression for the steady state genetic variance in terms of base population parameters:

$$\sigma^2_{g_{(L)}} = \frac{2\sigma^2_{g_{(0)}}(1 - h^2_{(0)})}{1 - 2h^2_{(0)} + \sqrt{1 + 4h^2_{(0)}k(1 - h^2_{(0)})}} \qquad (5.18)$$

An expression for the response to mass selection in the limit relative to response in the initial generation is:

$$R_{(L)}/R_{(1)} = \frac{i h_{(L)} \sigma_{g_{(L)}}}{i h_{(0)} \sigma_{g_{(0)}}} = \sqrt{\frac{h^2_{(L)}}{h^2_{(0)}(1 + k\, h^2_{(L)})}} \qquad (5.19)$$





5.3. Incorporating Gametic Phase Disequilibrium in the Selection Index

Because selection affects genetic variances and co-variances, it also affects elements of the **P** and **G** matrices that are needed to derive the optimal weights for selection indexes. In this section we will illustrate how changes in genetic parameters can be incorporated in selection index derivations and will evaluate their impact on accuracy of the index.

In general, because selection affects between and within-family variances differentially, derivation of elements of the **P** and **G** matrices must be based on the partitioning of the genetic value of individuals into parental and Mendelian sampling components:

$$g_{offspring} = \tfrac{1}{2}\, g_s + \tfrac{1}{2}\, g_d + g_m \tag{5.20}$$

and, from equation (5.10), genetic variance in the offspring generation, t, must be partitioned into:

$$\sigma^2_{g_{(t)}} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} \tag{5.21}$$

with from (5.7)

$$\sigma^{*2}_{g_{s(t-1)}} = (1 - k_s\, r^2_{s_{(t-1)}})\,\sigma^2_{g_{(t-1)}}$$

$$\sigma^{*2}_{g_{d(t-1)}} = (1 - k_d\, r^2_{d_{(t-1)}})\,\sigma^2_{g_{(t-1)}}$$

$$\sigma^2_{g_m} = \tfrac{1}{2}\,\sigma^2_{g_{(o)}}$$

As an example consider the situation where selection of sires and dams is on an index using phenotype of the individual and the mean performance of m full sibs. The index for selection in generation t will then take the form,

$$\hat{g}_{(t)} = b_{1_{(t)}} x_1 + b_{2_{(t)}} x_2 \tag{5.22}$$

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 the matrices needed to derive the index for generation t, $\mathbf{P}_{(t)}$ and $\mathbf{G}_{(t)}$ will take the form,

$$\mathbf{P}_{(t)} = \begin{bmatrix} \sigma^2_{x_1} & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma^2_{x_2} \end{bmatrix}, \qquad \mathbf{G}_{(t)} = \begin{bmatrix} \sigma_{x_1 g} \\ \sigma_{x_2 g} \end{bmatrix} \tag{5.23}$$

Elements can be derived as follows:

From equation (5.21):

$$\sigma^2_{x_1} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} + \sigma^2_e \tag{5.24}$$

$$\sigma^2_{x_2} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + (\sigma^2_{g_m} + \sigma^2_e)/m \tag{5.25}$$

$$\sigma_{x_1 x_2} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} \tag{5.26}$$

$$\sigma_{x_1 g} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} \tag{5.27}$$

and

$$\sigma_{x_2 g} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} \tag{5.28}$$

In generation 0, prior to selection, the above equations simplify to those derived in section 4.4.2.2.1.

For a trait with $h^2 = 0.5$, $\sigma^2_{g_{(0)}} = 25$, $\sigma^2_{p_{(0)}} = 50$, and m=5 full-sibs, we get the following:

$$\mathbf{P}_{(0)} = \begin{bmatrix} 50 & 12.5 \\ 12.5 & 20 \end{bmatrix} \qquad \mathbf{G}_{(0)} = \begin{bmatrix} 25 \\ 12.5 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_{(0)} = \mathbf{P}_{(0)}^{-1}\mathbf{G}_{(0)} = \begin{bmatrix} .4074 \\ .3704 \end{bmatrix}$$

Accuracy is $$r_{(0)} = \sqrt{\frac{\mathbf{b}'_{(0)}\mathbf{G}_{(0)}}{\sigma^2_{g_{(0)}}}} = 0.77$$

When in generation 0 only the 5% of sires and dams with the highest EBV are used to produce offspring, $k = 0.863$ and

$$\sigma^{*2}_{g_{s(1)}} = \sigma^{*2}_{g_{d(1)}} = (1 - k\, r^2_{(0)})\,\sigma^2_{g_{(0)}} = (1 - 0.863 \times 0.77^2)25 = 12.21$$

and $$\sigma^2_{g_{(t)}} = \tfrac{1}{4}\,\sigma^{*2}_{g_{s(t-1)}} + \tfrac{1}{4}\,\sigma^{*2}_{g_{d(t-1)}} + \sigma^2_{g_m} = 18.61$$

Using these values to derive elements of the **P** and **G** matrices for t=1 we get:

$$\mathbf{P}_{(1)} = \begin{bmatrix} 43.61 & 6.11 \\ 6.11 & 13.61 \end{bmatrix} \qquad \mathbf{G}_{(1)} = \begin{bmatrix} 18.61 \\ 6.11 \end{bmatrix} \quad \text{and} \quad \mathbf{b}_{(1)} = \mathbf{P}_{(1)}^{-1}\mathbf{G}_{(1)} = \begin{bmatrix} .3883 \\ .2746 \end{bmatrix}$$

Accuracy is $$r_{(1)} = \sqrt{\frac{\mathbf{b}'_{(1)}\mathbf{G}_{(1)}}{\sigma^2_{g_{(1)}}}} = 0.69$$

Using the recursive equations, this accuracy can be used to predict response to selection from t=1 to t=2 and to derive the genetic variance and selection in t=2.

Note that, compared to generation 0, selection reduced the variance among sires and dams and, as a consequence, the relative importance of observations on full-sibs is lower for the index used for selection in t=1 and the relative importance of observations on the individual has increased.

The reduced importance of full-sib information can also be illustrated by comparing accuracy of the index to the accuracy of selecting on own phenotype alone, which is equal to $h_{(0)}$ and $h_{(1)}$ for t=0 and t=1, respectively. Based on this, the efficiency of the index excluding information from the full-sibs is 0.71/0.77 = 0.92 and 0.65/0.69 = 0.95 before and after one round of selection.

5.4 Incorporating Gametic Phase Disequilibrium in BLUP EBV

The previous section described methods to incorporate the effect of selection on genetic variance components in derivation of selection indexes based on recursive equations. In principle, these

methods can also be applied to the selection indexes described in section 4.5, method 1, to approximate BLUP EBV. Examples are in Wray and Hill (1989) and Villaneuva et al. (1993).

For BLUP EBV, however, an alternative method can be used to incorporate the Bulmer effect, which facilitates direct derivation of steady state parameters. This method is based on the second approach for approximating BLUP EBV described in section 4.5 and utilizes the important property of BLUP EBV that their prediction error variance (PEV) does not depend on selection, but only on the amount of information used, with information defined as the number and type of records that is available on the individual itself and its relatives. This was described by Henderson (1975), using the argument that PEV's are based on the inverse of the coefficient matrix, which depends on the design matrices, the matrix of additive genetic relationships, and genetic parameters in the base population:

$$\sigma^2_{\varepsilon} = \text{Var}(\varepsilon) = \text{Var}(\hat{\boldsymbol{g}} - \boldsymbol{g}) = \mathbf{C}_{22} \qquad (5.29)$$

where $\boldsymbol{\varepsilon}$, $\hat{\boldsymbol{g}}$, and $\boldsymbol{g}$ are vectors of prediction errors, EBV, and BV, respectively, and $\mathbf{C}_{22}$ is the part of the inverse of the coefficient matrix of the mixed model equations that corresponds to animal breeding values. Elements of $\mathbf{C}_{22}$ do not depend on selection. Therefore, the PEV of a particular animal with a particular amount of information in an unselected population is the same as if that animal was in a selected population (but with selection accounted for through ancestor information). Thus, to get the PEV of an EBV, the mixed model equations can be set up ignoring the effect of selection on genetic variance and solved for. The same applies to approximations of BLUP EBV using the selection index methods described in section 4.5. Thus, the variance of prediction errors can be derived as:

$$\sigma^2_{\varepsilon_{(0)}} = (1 - r^2_{(0)})\,\sigma^2_{g_{(o)}} \qquad (5.30)$$

where the subscript 0 (t=0) refers to parameters derived for an unselected population, and $r_{(0)}$ is the accuracy of the BLUP EBV, derived using an index that ignores the effect of selection, following section 4.5.

Although selection doesn't affect the PEV, and, therefore, remains equal to $\sigma^2_{\varepsilon_{(0)}}$, PEV can also be derived based on the accuracy and genetic variance in the selected population as:

$$\sigma^2_{\varepsilon_{(t)}} = (1 - r^2_{(t)})\,\sigma^2_{g_{(t)}} \qquad (5.31)$$

Thus, using the property that PEV is unaffected by selection:

$$\sigma^2_{\varepsilon_{(t)}} = \sigma^2_{\varepsilon_{(0)}}$$
$$(1 - r^2_{(t)})\,\sigma^2_{g_{(t)}} = (1 - r^2_{(0)})\,\sigma^2_{g_{(o)}}$$

which, solving for $r^2_{(t)}$ results in:

$$r^2_{(t)} = 1 - (1 - r^2_{(0)})\,\sigma^2_{g_{(o)}} / \sigma^2_{g_{(t)}} \qquad (5.32)$$

This equation expresses the accuracy of EBV in a selected population in terms of the accuracy of EBV in an unselected population and the ratio of genetic variance in the unselected and selected population. Equation (5.32) holds for any generation and for any group of individuals.

Together with the recursive equation (5.11) for genetic variance:

$$\sigma^2_{g_{(t+1)}} = \tfrac{1}{4}(1 - k_s\, r^2_{s_{(t)}})\,\sigma^2_{g_{(t)}} + \tfrac{1}{4}(1 - k_d\, r^2_{d_{(t)}})\,\sigma^2_{g_{(t)}} + \tfrac{1}{2}\,\sigma^2_{g_{(o)}}$$

equation (5.32) provides a recursive system to derive genetic variance, accuracy of selection, and response to selection, as illustrated in Table 5.2 for selection on BLUP EBV that are described in section 4.5, method 2. Note that it is assumed that full pedigree information is available in generation zero.

Table 5.2. Recursive prediction of genetic variance, accuracy, and response with selection on BLUP EBV. Selected fractions are 0.2 and 0.5 for males and females, respectively, for a trait with heritability 0.25 and phenotypic variance 100. Selection is on BLUP EBV from a hierarchical mating structure with 20 mates per sire and 10 offspring per dam. Accuracy in generation zero is derived in section 4.5.

t	$\frac{(i_s+i_d)}{2}$	k_s	k_d	$\sigma_{g(0)}^2$	$\sigma_{g(t)}^2$ from t-1	$r_{(0)}$	$r_{(t)}=\sqrt{\frac{(1-(1-r_{(0)}^2)\sigma_{g(0)}^2}{\sigma_{g(t)}^2}}$	$\bar{g}_{(t+1)} = \bar{g}_{(t)} + {}^1/_2(i_s+i_d)r_{(t)}\sigma_{g(t)}$	$R_{(t)} = \bar{g}_{(t+1)} - \bar{g}_{(t)}$	$\sigma^{*2}_{gs(t)} = (1-r_{(t)}^2 k_s)\sigma_{g(t)}^2$	$\sigma^{*2}_{gd(t)} = (1-r_{(t)}^2 k_d)\sigma_{g(t)}^2$	$\sigma_{g(t+1)}^2 = {}^1/_2\sigma^{*2}_{gs(t)} + {}^1/_2\sigma^{*2}_{gd(t)} + {}^1/_2\sigma_{g(0)}^2$
0	1.1	0.78	0.64	25	25.00	0.704	0.704	3.871	3.871	15.326	17.074	20.600
1	1.1	0.78	0.64	25	20.60	0.704	0.623	6.979	3.108	14.363	15.490	19.963
2	1.1	0.78	0.64	25	19.96	0.704	0.607	9.961	2.982	14.224	15.261	19.871
3	1.1	0.78	0.64	25	19.87	0.704	0.604	12.924	2.963	14.204	15.228	19.858
4	1.1	0.78	0.64	25	19.86	0.704	0.604	15.884	2.960	14.201	15.223	19.856
5	1.1	0.78	0.64	25	19.86	0.704	0.604	18.843	2.960	14.200	15.223	19.856

Table 5.2 shows that, similar to mass selection, the impact of the Bulmer effect reaches a steady state after 5 generations of selection.

5.4.1 Asymptotic Genetic Variance and Response to Selection

Equations (5.32) and (5.11) can also be used to directly derive steady state parameters, following Dekkers (1992). Assuming for simplicity equal selection in males and females, using equation (5.32), accuracy at the limit is:

$$r^2_{(L)} = 1 - (1-r^2_{(0)})\ \sigma^2_{g_{(o)}} / \sigma^2_{g_{(L)}} \qquad (5.33)$$

Simplifying equation (5.11) for equal selection among males and females, genetic variance at the limit is:

$$\sigma^2_{g_{(L)}} = ½(1-k\ r^2_{(L)})\sigma^2_{g_{(L)}} + ½\,\sigma^2_{g_{(o)}} \qquad (5.34)$$

Rearranging results in:

$$\sigma^2_{g_{(L)}} = \sigma^2_{g_{(o)}} /(1-k\ r^2_{(L)}) \qquad (5.35)$$

and substituting equation (5.33) gives an equation that expresses genetic variance at the limit in terms of parameters for t=0:

$$\sigma^2_{g_{(L)}} = [1+k(1-r^2_{(0)})]\,\sigma^2_{g_{(o)}} /(1+k) \qquad (5.36)$$

Equations (5.36) and (5.33) can then be used to derive response at the limit as:

$$R_{(L)} = i\ r_{(L)}\ \sigma_{g_{(L)}}$$

Response at t=0 is:

$$R_{(0)} = i\ r_{(0)}\ \sigma_{g_{(o)}}$$

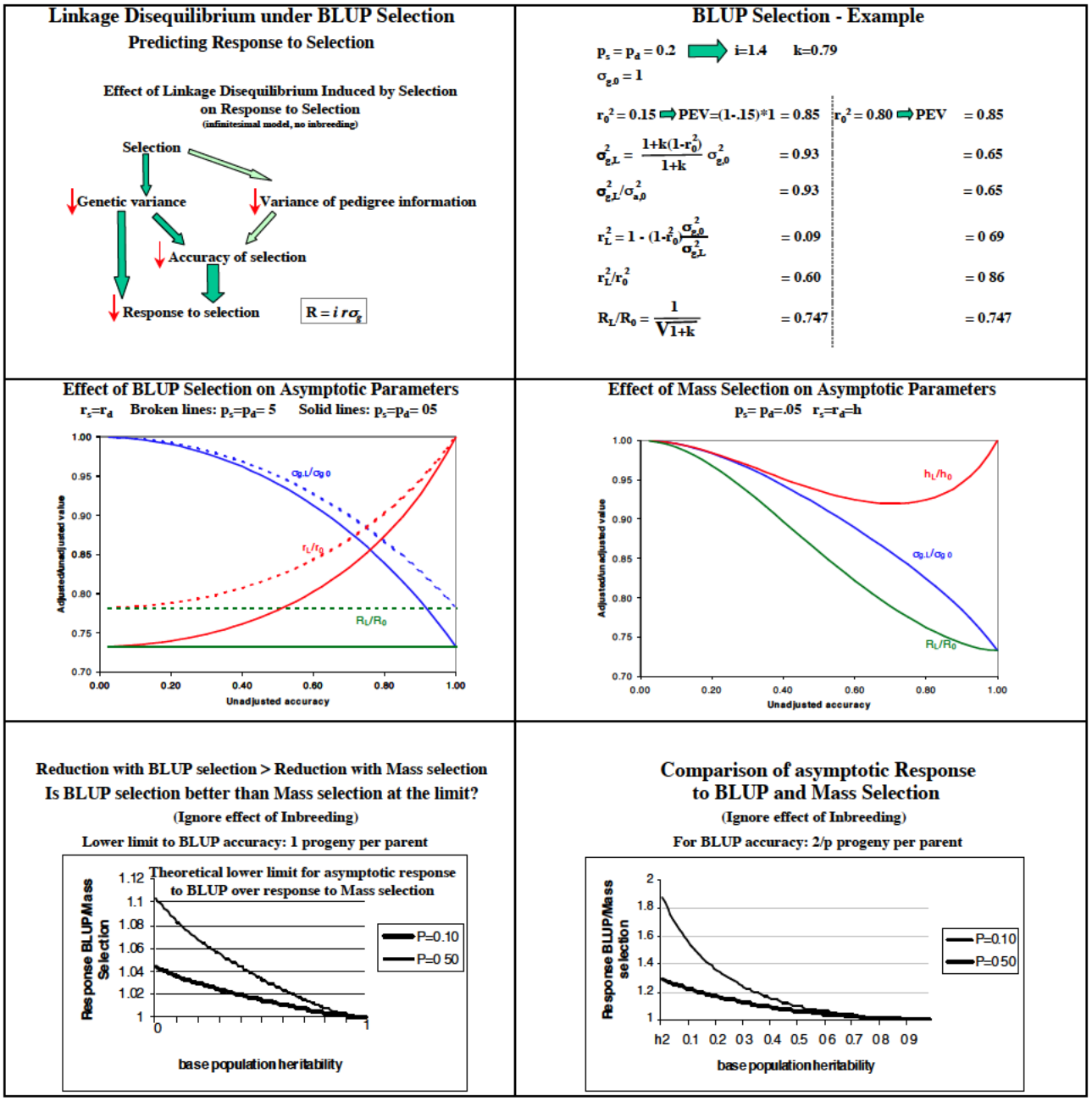
Therefore, response at the limit relative to response without accounting for the effect of selection on genetic variance under BLUP selection is equal to:

$$R_{(L)}/R_{(0)} = r_{(L)}\,\sigma_{g(L)} / r_{(0)}\,\sigma_{g(o)}$$

which, using equations (5.36) and (5.33) simplifies to:

$$R_{(L)}/R_{(0)} = \frac{1}{\sqrt{1+k}} \qquad (5.37)$$

Therefore, the reduction in response under BLUP selection depends only on selection intensity, and not on initial accuracy or heritability, as is the case for mass selection.



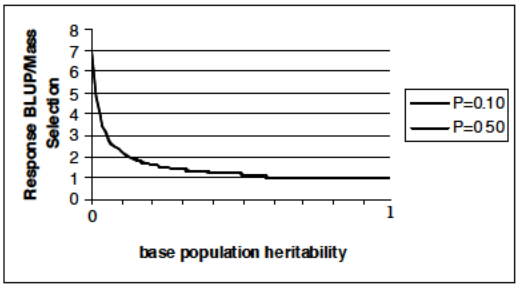
When selection intensities and initial selection accuracy's are different for the two sexes, similar procedures can be used to derive the following results (Dekkers, 1992):

$$R_{(L)}/R_{(0)} = \frac{i_s\sqrt{2\frac{r^2_{s_{(0)}}}{r^2_{d_{(0)}}} - k_d(\frac{r^2_{s_{(0)}}}{r^2_{d_{(0)}}} - 1)} + i_d\sqrt{2 + k_s(1 - \frac{r^2_{s_{(0)}}}{r^2_{d_{(0)}}})}}{(i_s\frac{r_{s_{(0)}}}{r_{d_{(0)}}} + i_d)\sqrt{2 + k_s + k_d}} \qquad (5.38)$$

When initial selection accuracy's are equal for both sexes, this equation simplifies to:

$$R_{(L)}/R_{(0)} = \sqrt{\frac{2}{2 + k_s + k_d}} \qquad (5.39)$$

Comparison of asymptotic Response to BLUP and Mass Selection
(Ignore effect of Inbreeding)
For BLUP accuracy: many progeny per parent



Example Programs with Four Selection Paths

Breeding plan		I	II	III	IV
Pathway	p_{SS}	.05	.03	03	.03
	p_{SD}	.20	.10	.10	.10
	p_{DS}	.05	.01	01	.01
	p_{DD}	.90	.80	80	.80
r_s		.85	.85	85	.85
r_d		.65	.65	50	.75
$\sigma^2_{S_{g,L}}/\sigma^2_{g,0}$		.736	.732	.773	.699
$\sigma^2_{D_{g,L}}/\sigma^2_{g,0}$		.774	.759	.783	.741
$r_{S,L}/r_{S,0}$		.929	.927	.942	.914
$r_{D,L}/r_{D,0}$		.775	.753	.408	.853
R_L/R_0		.759	.743	685	.752

Relative reductions in response to BLUP due to LD range from 22 to 32% for most practical programs

5.5 Selection Across Multiple Age Groups

The recursive equations developed in the previous sections can be expanded to selection across multiple age groups. Following the notation and derivations of Chapter 2, the genetic mean in year $t+1$ can be predicted as: $\bar{g}_{(t+1)} = \frac{1}{2}\bar{g}^*_{s_{(t)}} + \frac{1}{2}\bar{g}^*_{d_{(t)}}$

where $\bar{g}_{s_{(t)}}$ is the mean genetic value of sires selected at time t, which can be derived as a weighted average of genetic means of selected sires from each age group i at time t:

$$\bar{g}^*_{s_{(t)}} = \frac{1}{P_s}\sum p_{si}\, w_{si}\, \bar{g}^*_{si_{(t)}}$$

with $$\bar{g}^*_{si_{(t)}} = \bar{g}_{si_{(t)}} + i_{si}\, r_{si_{(t)}}\, \sigma_{gi_{(t)}}$$

and $$\bar{g}^*_{d_{(t)}} = \frac{1}{P_d}\sum p_{di}\, w_{di}\, \bar{g}^*_{di_{(t)}}$$

with $$\bar{g}^*_{di_{(t)}} = \bar{g}_{di_{(t)}} + i_{di} r_{di_{(t)}}\, \sigma_{gi_{(t)}}$$

From equation (5.7), genetic variance among selected sires from age group i at time t is equal to:

$$\sigma^{*2}_{g_{si(t)}} = (1 - k_{si}\, r^2_{si_{(t)}})\, \sigma^{\ 2}_{g_{si(t)}}$$

where k_{si} is the variance reduction factor corresponding to the selection intensity among sires of age group i: $k_{si} = i_{si}\,(i_{si} - x_{si})$

Genetic variance among all selected sires at time t is the pooled genetic variance of selected sires within each age group, augmented by the genetic variance between age groups:

$$\sigma^{*2}_{g_{s(t)}} = \frac{1}{P_s} \sum p_{si}\, w_{si}\, \sigma^{*2}_{g_{si(t)}} + \frac{1}{P_s} \sum p_{si}\, w_{si}\, (\bar{g}^{*}_{si_{(t)}} - \bar{g}^{*}_{s_{(t)}})^2 \qquad (5.40)$$

Similarly for dams:

$$\sigma^{*2}_{g_{d(t)}} = \frac{1}{P_d} \sum p_{di}\, w_{di}\, \sigma^{*2}_{g_{di(t)}} + \frac{1}{P_d} \sum p_{di}\, w_{di}\, (\bar{g}^{*}_{di_{(t)}} - \bar{g}^{*}_{d_{(t)}})^2 \qquad (5.41)$$

And genetic variance at time t+1 can be computed using equation (5.8) as:

$$\sigma^{\ 2}_{g_{(t\ 1)}} = ¼\, \sigma^{*2}_{g_{s(t)}} + ¼\, \sigma^{*2}_{g_{d(t)}} + ½\, \sigma^{\ 2}_{g_0}$$

5.6 Effects of Sample size and Inbreeding

There are two additional factors that affect genetic variance in future generations under the infinitesimal model: sample size and inbreeding. Models to incorporate these effects will be presented in the following sections.

5.6.1 Effect of finite population size on genetic variance

Expected variances derived in the previous sections apply to infinite population sizes. When selecting n individuals out of a population, in addition to the effect of selection on genetic variance, variance is expected to be reduced further by a factor $(1-\frac{1}{n})$. Thus, extending equation (5.7):

$$\sigma^{*2}_{g} = (1-\frac{1}{n})(1 - k\, r^2_{\hat{g}g})\, \sigma^2_g \qquad (5.42)$$

This adjustment is needed because the variances predicted in the previous sections are expected population variances rather than expected sampling variances. Recalling from statistics, sample variance is estimated by dividing sums of squares by n, whereas population variance is estimated by dividing sums of squares by n-1. Thus, to convert an estimate of population variance to an estimate of sample variance, the population variance estimate must be multiplied by $(n\text{-}1)/n = (1-\frac{1}{n})$. It is clear that the impact of this adjustment will be minor for n>50.

5.6.2 Effect of Inbreeding on Genetic Variance

The coefficient of inbreeding of an individual is equal to the probability that two alleles drawn at random from a locus at that individual are identical by descent. Inbreeding, thereby, reduces the variance contributed by Mendelian sampling by a parent by a factor $(1\text{-}F_i)$, where F_i is the

coefficient of inbreeding of the parent. Averaging over all sires and dams that are used for breeding, Mendelian sampling variance contributed to the next generation then is equal to:

$$\sigma^2_{g_{m(t+1)}} = (1 - \tfrac{1}{2}(\overline{F}_{s(t)} + \overline{F}_{d(t)}))\tfrac{1}{2}\sigma^2_{g_{(o)}} \qquad (5.43)$$

where $\overline{F}_{s(t)}$ and $\overline{F}_{d(t)}$ are mean coefficients of inbreeding of sires and dams selected at time *t*.

5.7 Multiple Stage Selection

Breeding goal: $H = v_1g_1 + v_2g_2 + v_3g_3 + \ldots\ldots + v_ng_n$ **= v'g**

Information sources: $X_1, X_2, X_3, X_4, \ldots\ldots, X_m$

Selection index: $I = b_1X_1 + b_2X_2 + b_3X_3 + \ldots\ldots + b_mX_m$

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

Selection on I maximizes response to selection in H

- **requires all animals to be measured for all traits**

Multiple-stage selection:

Stage 1: select on $I_1 = b_1X_1 + b_2X_2 + \ldots. + b_kX_k$ **$= b_1'X_1$**

Stage 2: select on $I_2 = b_1X_1 + b_2X_2 + b_3X_3 + \ldots. + b_mX_m$ **$= b_2'X$**

Only animals that are selected in stage 1 have to be evaluated for information sources $X_{k+1}, \ldots., X_m$

➔ **Cost savings**
➔ **Opportunities to increase population size for early stages**

Optimal index weights:

I_1 : **$b_1 = P_{11}^{-1} G_1 v$** **$P_{11} = Var(X_1)$**

$G_1 = Cov(X_1, g)$

I_2 : **$b_2 = P^{-1} G v$**

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix} \qquad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$$

Optimal index weights for index I_2 are not affected by Stage 1 selection on I_1, provided that all data that is included in I_1 is also included in I_2.

(Cunningham 1975 Theor. Appl. Genet. 46:55-61)

But accuracy and response to selection on I_2 is affected by selection on I_1

Stage 1: **accuracy of I_1:** $\mathbf{r_1} = \sqrt{\dfrac{\mathbf{b_1'G_1v}}{\mathbf{v'Cv}}}$ **C = Var(g)**

trait response vector: $S_{g,1} = i_1 \dfrac{\mathbf{b_1'G_1}}{\sqrt{\mathbf{b_1'P_{11}b_1}}}$

Stage 2: **accuracy of I_1:** $\mathbf{r_2} = \sqrt{\dfrac{\mathbf{b_2'G^*v}}{\mathbf{v'C^*v}}}$

trait response vector: $S_{g,2} = i_2 \dfrac{\mathbf{b_2'G^*}}{\sqrt{\mathbf{b_2'P^*b_2}}}$

Total response vector stage 1+2: $S_g = S_{g,1} + S_{g,2}$

P^*, G^*, and C^* are P, G, and C matrices adjusted for selection on I_1

Matrix equivalent of covariance adjustment for selection on w:

$$\sigma_{xy}^{\ *} = \sigma_{xy} - k\frac{\sigma_{wx}\sigma_{wy}}{\sigma_w^2} \qquad k = i(i\text{-}x) \qquad x \textbf{ = truncation pt.}$$

Vectors w, x, y ; selection on b'w

$$\mathrm{Cov}(\mathbf{x},\mathbf{y})^* = \mathrm{Cov}(\mathbf{x},\mathbf{y}) - k\,\frac{\mathrm{Cov}(\mathbf{x},\mathbf{b'w})\,\mathrm{Cov}(\mathbf{b'w},\mathbf{y})}{\mathrm{Var}(\mathbf{b'w})}$$

$$= \mathrm{Cov}(\mathbf{x},\mathbf{y}) - k\,\frac{\mathrm{Cov}(\mathbf{x},\mathbf{w})\mathbf{bb'}\mathrm{Cov}(\mathbf{w},\mathbf{y})}{\mathbf{b'}\mathrm{Var}(\mathbf{w})\mathbf{b}}$$

Stage 1 selection on b'w = $b_1'X_1$

$$P^* = Var(X)^* = Cov(X,X)^* = P - k \frac{Cov(\underline{X},\underline{X}_1)\underline{b}_1 \underline{b}_1' Cov(\underline{X}_1,\underline{X})}{\underline{b}_1' Var(\underline{X}_1)\underline{b}_1}$$

$$= P - k \frac{\begin{bmatrix}\underline{P}_{11}\\ \underline{P}_{21}\end{bmatrix}\underline{b}_1 \underline{b}_1' \begin{bmatrix}\underline{P}_{11} & \underline{P}_{21}\end{bmatrix}}{\underline{b}_1' \underline{P}_{11} \underline{b}_1}$$

$$G^* = Cov(X,g)^* = G - k \frac{Cov(\underline{X},\underline{X}_1)\underline{b}_1 \underline{b}_1' Cov(\underline{X}_1,\underline{g})}{\underline{b}_1' Var(\underline{X}_1)\underline{b}_1}$$

$$= G - k \frac{\begin{bmatrix}\underline{P}_{11}\\ \underline{P}_{21}\end{bmatrix}\underline{b}_1 \underline{b}_1' \underline{G}_1}{\underline{b}_1' \underline{P}_{11} \underline{b}_1}$$

$$C^* = Var(g)^* = Cov(g,g)^* = C - k \frac{Cov(\underline{g},\underline{X}_1)\underline{b}_1 \underline{b}_1' Cov(\underline{X}_1,\underline{g})}{\underline{b}_1' Var(\underline{X}_1)\underline{b}_1}$$

$$= C - k \frac{\underline{G}_1' \underline{b}_1 \underline{b}_1' \underline{G}_1}{\underline{b}_1' \underline{P}_{11} \underline{b}_1}$$

Multi-stage selection with availability of multi-trait EBV:

- EBV for all m traits available at every stage (different accuracies)

- **select on complete index at every stage**

$$I = v_1\hat{g}_1 + v_2\hat{g}_2 + \ldots + v_n\hat{g}_n$$

5.7.1 Optimization of proportions selected at each stage

Total proportion selected over s stages = p

Proportion selected at stage i = p_i

$$p = \prod_{i=1}^{s} p_i$$

a_i = cost of traits measured at stage i

$$\text{Total cost} = TC = a_1 + \sum_{i=2}^{s} a_i \prod_{j=1}^{i-1} p_j$$

Maximize gain in breeding goal per unit of cost: $Q = \Delta H/TC$

Or Maximize Profit from sale of breeding stock:

Profit = N p k b ΔH – TC

N = total # animals in breeding program
p = percent selected
k = # times breed is multiplied before distribution
b = slope of supply-demand curve
= extra returns from sale of animal per extra unit genetic worth of animal

Maximize proportion selected at each stage – requires numerical integration (Ducrocq and Colleau, 1989).

References:
Cunningham (1975), Theor. Appl. Genet. 46:55
Ducrocq and Colleau (1989) Genet. Sel. Evol. 21:185
Xu and Muir (1991) Genetics 129:963
Xu and Muir (1992) Theor. Appl. Genet. 83:451
Xu, Martin, and Muir (1995) J. Anim. Sci. 73:699

Chapter 6

Economic Selection Indexes

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 that 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 which adequately defines aggregate genotype over the range of genetic values encountered (see Chapter 7). If the function is a linear function, then the **aggregate genotype**, H, can be written as

$$H = v_1 g_1 + v_2 g_2 \ldots v_n g_n \qquad = \mathbf{v'g} \qquad (6.1)$$

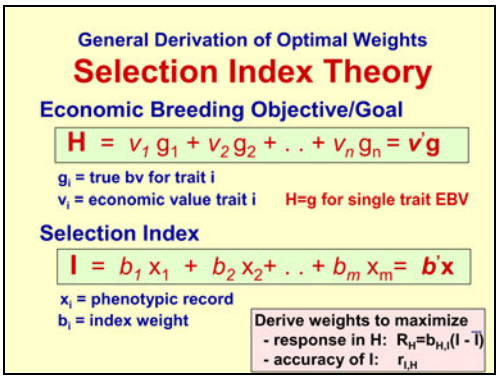
where g_i is the additive genetic value of the i^{th} trait and v_i is the economic value of genetically improving that trait. Note that v_i is a partial economic weight, that is, it is the economic value of genetically improving the i^{th} trait, when all other traits remain unchanged.

In vector notation, $\mathbf{v'} = [\, v_1 \,,\, v_2 \,,\, \ldots \,,\, v_n]$ and $\mathbf{g'} = [\, g_1 \,,\, g_2 \,,\, \ldots \,,\, g_n]$.

In practice, additive genetic values are not known. However we can record each individual's performance for a number of traits. Observations on these traits can then be combined into a **selection index**, I of the form,

$$I = b_1 x_1 + b_2 x_2 \ldots b_m x_m = \mathbf{b'x} \qquad (6.2)$$

where x_i is an observation on the i^{th} trait and b_i is the selection index coefficient (or weight) for that trait. In vector notation: $\mathbf{b'} = [\, b_1 \,,\, b_2 \,,\, \ldots \,,\, b_m]$ and $\mathbf{x'} = [\, x_1 \,,\, x_2 \,,\, \ldots \,,\, x_m]$.



The problem is then to estimate the selection index coefficients, b_i, such that selection of individuals on their **selection index** value, I, maximizes response in the **aggregate genotype**, H. The selection index methods described in Chapter 4 can be used to derive such indexes based on:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{v} \qquad (6.3)$$

It is worth noting at this point that the traits recorded and which appear in the index do not need to be, and often are not, the same traits as those that appear in the aggregate genotype. As a crude example, consider a terminal sire line of pigs. Assume that the profitability of this line is (approximately) a linear function of carcass weight and lean percentage at 120 days of age. Thus the aggregate genotype would be written as:

$$H = v_1\, g_{(\text{carcass weight})} + v_2\, g_{(\text{lean percentage})}$$

Neither carcass weight nor lean percentage can be recorded directly in live pigs, but we could record live weight at 120 days and ultrasonically estimated back fat depth at 120 days. The selection index would then take the form,

$$I = b_1\, x_{(\text{live weight})} + b_2\, x_{(\text{ultrasonic back fat depth})}$$

In this case, neither of the two traits recorded actually appear in the aggregate genotype, but both can be expected to be closely related to the traits in the aggregate genotype.

6.1 Predicting Response to Selection and Related Parameters

If we were interested in a practical problem of what weight to give a series of observations on a particular population of animals for selection, and we were certain of our phenotypic and genetic parameters contained in **P** and **G** and were sure of our economic weights in **v**, then we might stop here. But curiosity alone would likely prompt us to ask what the predicted variance of the index would be, what would be the predicted response of the aggregate genotype to selection on the index, and what would be the predicted correlated response of each of the traits in the aggregate genotype? And, mere curiosity aside, such predictions are essential when comparing different possible indexes, for assessing whether predicted responses of individual traits are likely to be acceptable to the users of the index or to their customers, and for determining whether it is worthwhile to record data on a given trait for inclusion in the index.

In section 4.4.1.2 the variance of the index, σ_I^2, was derived as $\sigma_I^2 = \mathbf{b'Pb}$ (6.4)

It is important to realize that this equation holds whatever values of **b** are used, i.e. not only for the optimal **b** derived using selection index theory, but for any arbitrary vector **b**. However, if **b** is the optimal set of index coefficients, then $\mathbf{b}=\mathbf{P}^{-1}\mathbf{G}\mathbf{v}$ **,** which when substituted into (6.4) gives:

$$\sigma_I^2 = \mathbf{b'P}\,\mathbf{P}^{-1}\,\mathbf{G}\mathbf{v} = \mathbf{b'}\,\mathbf{G}\mathbf{v} \qquad (6.5)$$

This second form is most often quoted as the variance of a selection index. But it should always be remembered that this holds only for the "optimal" selection index. Since you may wish to explore the consequences of using sub-optimal indexes, it is probably safer to use the form given in equation (6.4), which holds for any selection index, optimal or sub-optimal.

Applying the general equation for response to selection, to selection on I to improve H, predicted genetic superiority for the aggregate genotype, H, to selection on the index, I, is given by:

$$S_H = i\, r_{HI}\, \sigma_H \tag{6.6}$$

As given in 4.4.1.2, the variance of the breeding goal can be derived as:

$$\sigma_H^2 = \mathbf{v'Cv} \tag{6.7}$$

where $\mathbf{C}$ is an n x n matrix of genetic covariances among the traits in the aggregate genotype.

Similarly, it follows that $\sigma_{HI} = \mathbf{b'Gv}$ (6.8)

Hence, $$r_{HI} = \frac{\sigma_{HI}}{\sigma_I\, \sigma_H} = \frac{\mathbf{b'Gv}}{\sqrt{\mathbf{b'Pb\ v'Cv}}} \tag{6.9}$$

and $$S_H = i \frac{\mathbf{b'Gv}}{\sqrt{\mathbf{b'Pb}}} \tag{6.10}$$

Note again, all the above derivations, (6.6) to (6.10), apply to any selection index (that is any set of b values) not just the optimum index. However, if we are dealing with the optimal index, where $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gv}$, then as noted at equations (6.4) and (6.5) $\mathbf{b'Pb} = \mathbf{b'Gv}$ and substituting into equation (6.10) gives:

$$S_H = i\ \sqrt{\mathbf{b'Gv}} \tag{6.11}$$

We could have also obtained this optimal selection response directly, as $S_H = i\, b_{HI}\, \sigma_I$

and recalling that for the optimal selection index: $b_{HI} = 1$, hence: $S_H = i\, \sigma_I$

and substituting in the variance of the optimal index defined at (6.5): $S_H = i\ \sqrt{\mathbf{b'Gv}}$ (6.12)

The accuracy of an index is defined as its correlation with the aggregate genotype, i.e. r_{HI}, as estimated at (2.17). For an optimal index, noting the equivalence of equations (2.13) and (2.14), the expectation for the accuracy of the index can also be written as

$$r_{HI} = \sqrt{\frac{\mathbf{b'Gv}}{\mathbf{v'Cv}}} \tag{6.13}$$

The expected change in the additive genetic value of the i^{th} trait in the aggregate genotype due to selection on the index, S_{g_i}, can be found as a correlated response to selection on I as:

$$S_{g_i} = b_{g_iI}\, S_I = b_{g_iI}\, i\sigma_I = i\, \frac{\sigma_{g_iI}}{\sigma_I^2}\sigma_I = i\, \frac{\sigma_{g_iI}}{\sigma_I} \tag{6.14}$$

Where $$\sigma_{g_iI} = \mathrm{cov}(g_i, \mathbf{b'x}) = \mathbf{b'}\mathrm{cov}(g_i, \mathbf{x}) = \mathbf{b'G}_i \tag{6.15}$$

where $\mathbf{G}_i$ denotes the i^{th} column of $\mathbf{G}$: $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_i, \ldots, \mathbf{G}_n]$

Hence for all indexes: $$S_{g_i} = i \frac{\mathbf{b'G}_i}{\sqrt{\mathbf{b'Pb}}} \quad (6.16)$$

For an optimum index, the solution is also $$S_{g_i} = i \frac{\mathbf{b'G}_i}{\sqrt{\mathbf{b'Gv}}} \quad (6.17)$$

It also follows directly from (6.17), that the vector of genetic responses of each of the traits in the aggregate genotype, can be found as: $$\boldsymbol{S}_g = [S_{g_1}, \ldots, S_{g_i}, \ldots, S_{g_n}] = i \frac{\mathbf{b'G}}{\sqrt{\mathbf{b'Pb}}} \quad (6.18)$$

These derivations of the principal parameters defining a selection index for all indexes and for optimal indexes are summarized in Table 6.1

Table 6.1 Summary of selection index formulae for any index and for optimal indexes.

Parameters	Derivation	
	Any Index	**Optimal Index**
$\mathbf{b}$	Arbitrary	$\mathbf{P^{-1}Gv}$
σ_I^2	$\mathbf{b'Pb}$	$\mathbf{b'Gv}$
σ_H^2	$\mathbf{v'Cv}$	$\mathbf{v'Cv}$
σ_{HI}	$\mathbf{b'Gv}$	$\mathbf{b'Gv}$
r_{HI}	$\frac{\mathbf{b'Gv}}{\sqrt{\mathbf{b'Pb\ v'Cv}}}$	$\sqrt{\frac{\mathbf{b'Gv}}{\mathbf{v'Cv}}} = \sqrt{\frac{\mathbf{b'Pb}}{\mathbf{v'Cv}}} = \frac{\sigma_I}{\sigma_H}$
S_H	$i \frac{\mathbf{b'Gv}}{\sqrt{\mathbf{b'Pb}}}$	$i\sqrt{\mathbf{b'Gv}} = i\sqrt{\mathbf{b'Pb}} = i\sigma_I$
S_{g_i}	$i \frac{\mathbf{b'G}_i}{\sqrt{\mathbf{b'Pb}}}$	$i \frac{\mathbf{b'G}_i}{\sqrt{\mathbf{b'Gv}}}$
$\boldsymbol{S}_g$	$i \frac{\mathbf{b'G}}{\sqrt{\mathbf{b'Pb}}}$	$i \frac{\mathbf{b'G}}{\sqrt{\mathbf{b'Gv}}}$

The correlation between index values of two relatives, *i* and *j*, which is needed to, e.g., compute selection intensities as in section 3.6, can be computed as:

$$t = \text{corr}(I_i, I_j) = \text{corr}(\mathbf{b'x}_i, \mathbf{b'x}_j) = \frac{\mathbf{b'}\,\text{cov}(\mathbf{x}_i, \mathbf{x}_j)\mathbf{b}}{\mathbf{b'Pb}} = \frac{\mathbf{b'Rb}}{\mathbf{b'Pb}}$$

where **R** is a matrix with covariances between information sources on the two relatives (De Boer and Van Arendonk, 1989).

6.2 Example of Economic Index Selection

A beef breed is to be used as a terminal sire and economic analysis has shown that three key traits are post-weaning gain (PWG) with an economic value of 370 $/kg/d, ultrasonic back fat depth (BF), with an economic value under the local payment system of -20 $/mm, and feed intake, with an economic value of –50 $/kg/d. Only PWG and BF are recorded in the bull-testing program. Genetic and economic parameters are summarized in Figure 6.1.

Figure 6.1. Example Derivation of Economic Index
Beef Cattle Terminal Sire Line

	Post Weaning Gain (PWG)	Back Fat (BF)	Feed Intake (FI)
Econ value	370 $/kg/d	-20 $/mm	-50 $/kg/d
h^2	0.40	0.44	0.20
σ_p	0.23 kg/d	0.15 mm	0.5 kg/d

r_p \ r_g	PWG	BF	FI
PWG		0.32	0.70
BF	0.18		0.48
FI	0.23	0.40	

Select on index of own phe-notype for PWG + BF

Matrices **P**, **G**, and **C** have the following elements:

$$\mathbf{P} = \begin{bmatrix} \sigma^2_{p_{PWG}} & \sigma_{p_{PWG}, p_{BF}} \\ \sigma_{p_{PWG}, p_{BF}} & \sigma^2_{p_{BF}} \end{bmatrix} \qquad (6.19)$$

$$\mathbf{G} = \begin{bmatrix} \sigma_{p_{PWG},g_{PWG}} & \sigma_{p_{PWG},g_{BF}} & \sigma_{p_{PWG},g_{FI}} \\ \sigma_{p_{BF},g_{PWG}} & \sigma_{p_{BF},g_{BF}} & \sigma_{p_{BF},g_{FI}} \end{bmatrix} \tag{6.20}$$

$$\mathbf{C} = \begin{bmatrix} \sigma^2_{g_{PWG}} & \sigma_{g_{PWG},g_{BF}} & \sigma_{g_{PWG},g_{FI}} \\ \sigma_{g_{BF},g_{PWG}} & \sigma^2_{g_{BF}} & \sigma_{g_{BF},g_{FI}} \\ \sigma_{g_{FI},g_{PWG}} & \sigma_{g_{FI},g_{BF}} & \sigma^2_{g_{FI}} \end{bmatrix} \tag{6.21}$$

The elements of **P**, **G**, and **C** can be found by recalling that $r_{12} = \dfrac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}}$

hence $\sigma_{12} = r_{12}\sigma_1\sigma_2$

and that $\sigma_g^2 = h^2\,\sigma_p^2$

giving $\mathbf{P} = \begin{bmatrix} 0.0529 & 0.00621 \\ 0.00621 & 0.0225 \end{bmatrix}$, $\mathbf{G} = \begin{bmatrix} 0.02116 & 0.004632 & 0.02277 \\ 0.004632 & 0.0099 & 0.01068 \end{bmatrix}$

and $\mathbf{C} = \begin{bmatrix} 0.02116 & 0.004632 & 0.02277 \\ 0.004632 & 0.0099 & 0.01068 \\ 0.02277 & 0.01068 & 0.05 \end{bmatrix}$

The vector of economic weights is: $\mathbf{v} = \begin{bmatrix} 370 \\ -20 \\ -50 \end{bmatrix}$

Hence, the index weights would be $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{v} = \begin{bmatrix} 123.6 \\ 9.52 \end{bmatrix}$

The index accuracy would be $r_{HI} = \sqrt{\dfrac{\mathbf{b'Gv}}{\mathbf{v'Cv}}} = 0.6214$

The expected genetic superiority for the breeding goal to 1 standard deviation selection on the index is:

$$\boldsymbol{S}_H = \frac{\mathbf{b'Gv}}{\sqrt{\mathbf{b'Pb}}} = 28.722$$

Expected genetic superiorities for PWG, BF, and FI to one standard deviation selection on the index are:

$$\boldsymbol{S}_g = \frac{\mathbf{b'G}}{\sqrt{\mathbf{b'Gv}}} = \begin{bmatrix} 0.0926 \\ 0.0232 \\ 0.1015 \end{bmatrix}$$

Thus, one standard deviation of selection on the index would be expected to yield bulls with an average breeding value of +0.0926 kg/day for PWG, + 0.0232 mm for BF, and +.1015 kg/day more FI. Also, their expected breeding value for profit is +$28.72.

6.3 Economic Indexes Based on Estimated Breeding Values

So far we have dealt with economic selection indexes that were based on phenotypic records. In these cases, the main interest is to predict the individual's breeding value for the aggregate genotype. In practice, however, the sources of information that are available to develop the economic index are EBV for individual traits, rather than phenotypic records:

$$I = b_1 \hat{g}_1 + b_2 \hat{g}_2 + \ldots + b_m \hat{g}_m = \mathbf{b}' \hat{\mathbf{g}}_I$$

This leads to a step-wise procedure for development of economic selection indexes, in which the first step consists of predicting BV for individual traits and the second step of combining the resulting EBV into an economic index. An advantage of this step-wise approach is that it allows different breeders to put different emphasis on traits in the aggregate genotype, while utilizing the most accurate EBV for the component traits.

To derive the optimal weight that should be placed on EBV, consider a vector of available phenotypic records **x** that can be subdivided into subvectors $\mathbf{x}_i$ which correspond to phenotypic records on trait i for i=1,…, m:

$$\mathbf{x}' = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_m]$$

and a breeding goal with n traits: $H = v_1 g_1 + v_2 g_2 + \ldots + v_n g_n = \mathbf{v}'\mathbf{g}_H$

Then, an economic selection index based on the full set of phenotypic records **x** can be derived as:

$$I_F = \mathbf{b}_F'\mathbf{x}$$

with the vector of index weights derived using standard selection index theory from:

$$\mathbf{b}_F = \mathbf{P}^{-1}\mathbf{Gv}$$

Note that this index maximizes accuracy of predicting H given the available phenotypic records.

In the two-step approach, the objective is to develop the following type of index:

$$I_S = b_1 \hat{g}_1 + b_2 \hat{g}_2 + \ldots + b_m \hat{g}_m = \mathbf{b}_S' \hat{\mathbf{g}}_I$$

Consider first the situation where the traits in the index and the breeding goal are the same. Thus, $m=n$ and $\mathbf{g}_H = \mathbf{g}_I = \mathbf{g}$.

The first step is to derive EBV for each individual trait i based on all available data $\mathbf{x}$: $\hat{g}_i = \mathbf{b}_i'\mathbf{x}$

Index weights for this EBV can be derived as: $\mathbf{b}_i = \mathbf{P}^{-1}\mathbf{G}_i$

where $\mathbf{G}_i$ is the i^{th} column of matrix $\mathbf{G}$, i.e. a vector representing the genetic covariances between the observations in $\mathbf{x}$ and the i^{th} trait in the aggregate genotype g_i:

$$\mathbf{G} = [\mathbf{G}_1; \mathbf{G}_2; \ldots ; \mathbf{G}_n]$$

Because all EBV are based on the same data vector $\mathbf{x}$, the selection index equations for the individual EBV can be combined to directly estimate the vector of EBV as:

$$\hat{\mathbf{g}} = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n]' = [\mathbf{b}_1; \mathbf{b}_2; \ldots ; \mathbf{b}_n]'\mathbf{x}$$

Combining the selection index equations that are used to derive each set of index weights $\mathbf{b}_i$ we get:

$$\hat{\mathbf{g}} = \{\mathbf{P}^{-1}[\mathbf{G}_1; \mathbf{G}_2; \ldots ; \mathbf{G}_n]\}'\mathbf{x} = (\mathbf{P}^{-1}\mathbf{G})'\mathbf{x}$$

Using these EBV to develop the economic index we get:

$$I_S = \mathbf{b}_S' \hat{\mathbf{g}} = \mathbf{b}_S'(\mathbf{P}^{-1}\mathbf{G})'\mathbf{x} = (\mathbf{P}^{-1}\mathbf{G}\,\mathbf{b}_S)'\mathbf{x}$$

Note that, if we set $\mathbf{b}_S = \mathbf{v}$ then $I_S = (\mathbf{P}^{-1}\mathbf{G}\mathbf{v})'\mathbf{x} = \mathbf{b}_F'\mathbf{x} = I_F$ because $\mathbf{b}_F = \mathbf{P}^{-1}\mathbf{G}\mathbf{v}$ based on the one-step approach

Note that this proves that the optimal index can be obtained by weighting the EBV by the economic weight for that trait in the breeding goal.

Note, however, that this only holds if the same traits are included in the breeding goal and the index and if the EBV are based on all available data on all traits, i.e. multiple-trait EBV are used.

If the traits in the breeding goal and index are not the same, i.e. $m \neq n$ and $\mathbf{g}_H \neq \mathbf{g}_I$ then the index weights can be derived by partitioning the breeding goal into a component that is related to the traits in the index and an uncorrelated residual:

$$H = \mathbf{v}'\mathbf{g}_H = \mathbf{v}'(\mathbf{b}'_{\mathbf{g}_H\mathbf{g}_I}\,\mathbf{g}_I + \mathbf{e})$$

where $\mathbf{b}_{\mathbf{g}_H\mathbf{g}_I}$ is a vector of regression coefficients for the regression of vector $\mathbf{g}_H$ on $\mathbf{g}_I$. Vector $\mathbf{b}_{\mathbf{g}_H\mathbf{g}_I}$ can be derived using standard multiple regression methods as:

$$\mathbf{b}_{\mathbf{g}_H\mathbf{g}_I} = \mathbf{C}_I^{-1}\,\mathbf{C}_{IH}$$

where $\mathbf{C}_I$ is the genetic variance/covariance matrix among the traits which appear in the index and $\mathbf{C}_{IH}$ is the genetic covariance matrix between traits in the index and traits in the aggregate genotype.

The residual vector **e** is by definition uncorrelated to traits in the index and does, therefore, not need to be considered when deriving the index. Thus, considering only the first term, the new breeding goal is one for which the assumption $\mathbf{g}_H=\mathbf{g}_I$ holds but with a new set of economic values:

$$H^* = \mathbf{v}^{*\prime}\mathbf{g}_I = \mathbf{v}'\,\mathbf{b}'_{\mathbf{g}_H\mathbf{g}_I}\,\mathbf{g}_I$$

Thus

$$\mathbf{v}^* = \mathbf{b}_{\mathbf{g}_H\mathbf{g}_I}\,\mathbf{v}$$

and the optimal weights on the EBV for index $I_S = \mathbf{b}_S'\hat{\mathbf{g}}$ can be derived as:

$$\mathbf{b}_S = \mathbf{C}_I^{-1}\,\mathbf{C}_{IH}\,\mathbf{v}$$

An alternative derivation of this same result was given by Schneeberger et al. (1992).

Note that when $\mathbf{g}_H=\mathbf{g}_I$ i.e. when the same traits are included in H and I, $\mathbf{C}_I = \mathbf{C}_{IH}$ and $\mathbf{b}_S$ simplifies to the original result: $\mathbf{b}_S = \mathbf{v}$.

It is interesting to note that the index weights for indexes that contain EBV depend only on the economic values and genetic parameters. They do not depend on the accuracy of the individual EBV. Thus, the same index can be used for all animals.

It is important to realize, however, that these results only hold if all observations on all traits are used to estimate the BV for all traits. This would be the case when multiple-trait genetic evaluation models are used. The previous equations do not hold when trait EBV are derived from single-trait evaluation models, for which only data on trait i are used to estimate $\hat{g}_i$. The reason is that single-trait evaluation methods do not consider covariances between trait records, whereas covariances are considered in the multiple-trait evaluation methods. However, if the single trait EBV have high accuracy, index weights can still be approximated by the economic values because, with high accuracy, correlated trait information has little impact on EBV and, thus, single-trait EBV approximate multi-trait EBV.

If single trait EBV do not have high accuracy, approximate weights on EBV must be derived using selection index procedures using: $\mathbf{b}_S = \mathbf{P}_S^{-1}\,\mathbf{G}_S\mathbf{v}$

where $\mathbf{P}_S$ = matrix with (co-)variances among single trait EBV, $\hat{g}_i$

$\mathbf{G}_S$ = matrix with covariances of EBV $\hat{g}_i$ with true BV of traits in the breeding goal

Elements of $\mathbf{P}_S$ and $\mathbf{G}_S$ must now be approximated by specifying a vector $\mathbf{x}_i$ of sources of phenotypic records for each EBV $\hat{g}_i$. Standard selection index theory can then be used to derive each EBV as: $\hat{g}_i = \mathbf{b}_i'\mathbf{x}_i$ with $\mathbf{b}_i = \mathbf{P}_{ii}^{-1}\mathbf{G}_{ii}$

where $\mathbf{P}_{ii}$ is the (co-)variance matrix for records in $\mathbf{x}_i$, and $\mathbf{G}_{ii}$ is the vector of covariances of g_i with $\mathbf{x}_i$.

Then, the covariance between two EBV can be computed as:

$$\sigma_{g_i,g_j} = \text{cov}(\mathbf{b}_i'\mathbf{x}_i, \mathbf{b}_j'\mathbf{x}_j) = \mathbf{b}_i' \text{ cov}(\mathbf{x}_i, \mathbf{x}_j)\mathbf{b}_j = \mathbf{b}_i' \mathbf{P}_{ij} \mathbf{b}_j$$

where $\mathbf{P}_{ij}$ is the (co-)variance matrix between records in $\mathbf{x}_i$ and $\mathbf{x}_j$
and the covariance between and EBV and a true BV can be computed as:

$$\sigma_{g_i,g_j} = \text{cov}(\mathbf{b}_i'\mathbf{x}_i, \mathbf{g}_j) = \mathbf{b}_i' \text{ cov}(\mathbf{x}_i, \mathbf{g}_j) = \mathbf{b}_i' \mathbf{G}_{ij}$$

where $\mathbf{G}_{ij}$ is a vector with covariances of $\mathbf{x}_i$ with g_j.

Note that in contrast to indexes based on multi-trait EBV, weights on single-trait EBV depend on the accuracy of the EBV and on the sources of information that contribute to each animal's EBV. Index weights will, therefore, differ from animal to animal. Depending on parameters and the range of accuracies, it may however be possible to use a single index for all animals, with the index derived based on average amounts of information. Methods to evaluate the loss in accuracy when using a single index will be described in the next section.

One issue that we have not yet discussed is how to derive the accuracy of economic selection indexes that are based on EBV. When the index is based on single-trait EBV, accuracy can be derived using standard selection index theory as: $r_{HI} = \sqrt{\frac{\mathbf{b}_S'\mathbf{G}_S\mathbf{v}}{\mathbf{v}'\mathbf{C}\mathbf{v}}}$

For the index based on multi-trait EBV, although index weights do not depend on the amount of information that goes into each EBV, the accuracy of the index does. For multi-trait EBV indexes, accuracy must, therefore, be derived by specifying the sources of information in the multiple-trait vector of observations **x**. Then, accuracy can be obtained by deriving the accuracy of the full index: $r_{HI} = \sqrt{\frac{\mathbf{b}_F'\mathbf{G}\mathbf{v}}{\mathbf{v}'\mathbf{C}\mathbf{v}}}$

This is illustrated in the following example where selection is based on observations on growing pigs for growth rate (GR, g/d) on the individual (x_{GR}) and the mean of 5 full sibs (for $\bar{x}_{GR}$) and own performance for feed intake (x_{FI}, g/d). The phenotypic standard deviation is 100 for GR and 200 for FI, and the h^2 for both traits is 0.25. The phenotypic correlation is 0.6 while the genetic correlation is 0.8.

The aggregate genotype is $H = 0.2\, g_{GR} - 0.05\, g_{FI}$

The selection index weights for prediction of the aggregate genotype using the full index with all three sources of information (x_{GR}, $\bar{x}_{GR}$, x_{FI}) can be calculated as:

$$\mathbf{b}_F = \mathbf{P}^{-1}\mathbf{G}\mathbf{v} = \begin{bmatrix} 10000 & 1250 & 12000 \\ 1250 & 3000 & 2000 \\ 12000 & 2000 & 40000 \end{bmatrix}^{-1} \begin{bmatrix} 2500 & 4000 \\ 1250 & 2000 \\ 4000 & 10000 \end{bmatrix} \begin{bmatrix} 0.2 \\ -0.05 \end{bmatrix} = \begin{bmatrix} 0.02874 \\ 0.04011 \\ -0.00313 \end{bmatrix}$$

Weights $\mathbf{b}_1$ for the prediction of the breeding value for growth rate using all data are:

$$\mathbf{b}_1 = \mathbf{P}^{-1}\mathbf{G}_1 = \begin{bmatrix} 10000 & 1250 & 12000 \\ 1250 & 3000 & 2000 \\ 12000 & 2000 & 40000 \end{bmatrix}^{-1} \begin{bmatrix} 2500 \\ 1250 \\ 4000 \end{bmatrix} = \begin{bmatrix} 0.1702 \\ 0.3239 \\ 0.0327 \end{bmatrix}$$

Similarly, weights for feed intake using all data, $\mathbf{b}_2$, are:

$$\mathbf{b}_2 = \mathbf{P}^{-1}\mathbf{G}_2 = \begin{bmatrix} 10000 & 1250 & 12000 \\ 1250 & 3000 & 2000 \\ 12000 & 2000 & 40000 \end{bmatrix}^{-1} \begin{bmatrix} 4000 \\ 2000 \\ 10000 \end{bmatrix} = \begin{bmatrix} 0.1061 \\ 0.4935 \\ 0.1935 \end{bmatrix}$$

As demonstrated previously, the multiple-trait EBV can be combined into an economic index to predict the aggregate genotype with index weights equal to the economic values:

$$I_S = 0.2\, \hat{g}_{GR} - 0.05\, \hat{g}_{FI}$$

Substituting the multiple-trait index weights $\mathbf{b}_1$ and $\mathbf{b}_2$ gives:

$$I_S = 0.2 \begin{bmatrix} 0.1702 \\ 0.3239 \\ 0.0327 \end{bmatrix}' \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} - 0.05 \begin{bmatrix} 0.1061 \\ 0.4935 \\ 0.1935 \end{bmatrix}' \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} = \begin{bmatrix} 0.02874 \\ 0.04011 \\ -0.00312 \end{bmatrix}' \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix} = \mathbf{b}_F' \begin{bmatrix} x_{GR} \\ \bar{x}_{GR} \\ x_{FI} \end{bmatrix}$$

proving that the index based on multiple-trait EBV is equivalent to the full index.

Let us now look at a situation where breeding values for GR and FI are predicted with a univariate model, i.e. the EBV for GR is based on observations for GR only and similarly the EBV for FI is based entirely on observations for FI. The following index weights result:

$$\mathbf{b}_1 = \mathbf{P}_{11}^{-1}\mathbf{G}_{11} = \begin{bmatrix} 10000 & 1250 \\ 1250 & 3000 \end{bmatrix}^{-1}\begin{bmatrix}2500\\1250\end{bmatrix} = \begin{bmatrix}0.2088\\0.3297\end{bmatrix}$$

Since the single-trait EBV for FI is based only on own performance, the index weight is equal to heritability: $\mathbf{b}_2 = \mathbf{P}_{22}^{-1}\mathbf{G}_{22} = [40000]^{-1}\,[10000] = [0.25]$

Combining these single-trait EBV into an economic index to predict *H* results in:

$$\begin{aligned}\mathbf{b}_S = \mathbf{P}_S^{-1}\mathbf{G}_S\mathbf{v} &= \begin{bmatrix}\sigma^2_{\hat{g}_{GR}} & \sigma_{\hat{g}_{GR},\hat{g}_{FI}} \\ \sigma_{\hat{g}_{GR},\hat{g}_{FI}} & \sigma^2_{\hat{g}_{GR}}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{\hat{g}_{GR},g_{GR}} & \sigma_{\hat{g}_{GR},g_{FI}} \\ \sigma_{\hat{g}_{FI},g_{GR}} & \sigma_{\hat{g}_{FI},g_{FI}}\end{bmatrix}\mathbf{v} \\ &= \begin{bmatrix}\mathbf{b}_1'\mathbf{P}_{11}\mathbf{b}_1 & \mathbf{b}_1'\mathbf{P}_{12}\mathbf{b}_2 \\ \mathbf{b}_2'\mathbf{P}_{21}\mathbf{b}_1 & \mathbf{b}_2'\mathbf{P}_{22}\mathbf{b}_2\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{b}_1'\mathbf{G}_{11} & \mathbf{b}_1'\mathbf{G}_{12} \\ \mathbf{b}_2'\mathbf{G}_{21} & \mathbf{b}_2'\mathbf{G}_{22}\end{bmatrix}\mathbf{v} \\ &= \begin{bmatrix}934.184 & 791.25 \\ 791.25 & 2500\end{bmatrix}^{-1}\begin{bmatrix}934.125 & 1494.6 \\ 1000 & 2500\end{bmatrix}\begin{bmatrix}0.2\\-0.05\end{bmatrix} = \begin{bmatrix}0.1292\\-0.0109\end{bmatrix}\end{aligned}$$

These weights on single-trait EBV differ from those on multiple-trait EBV, which are equal to the economic values. Multiplying the weights out gives:

$$I_S = 0.1292\begin{bmatrix}0.2088\\0.3297\\0\end{bmatrix}'\begin{bmatrix}x_{GR}\\\bar{x}_{GR}\\x_{FI}\end{bmatrix} - 0.0109\begin{bmatrix}0\\0\\0.25\end{bmatrix}'\begin{bmatrix}x_{GR}\\\bar{x}_{GR}\\x_{FI}\end{bmatrix} = \begin{bmatrix}0.02698\\0.04261\\-0.00272\end{bmatrix}'\begin{bmatrix}x_{GR}\\\bar{x}_{GR}\\x_{FI}\end{bmatrix} = \mathbf{b}_F^{*\prime}\begin{bmatrix}x_{GR}\\\bar{x}_{GR}\\x_{FI}\end{bmatrix}$$

These weights are different from those for the optimal index and, thus, the index derived from single-trait EBV is not optimal. The reason for the suboptimality is that, when deriving the single-trait EBV, relative weights put on records are determined while only considering a single trait and not the whole aggregate genotype. In our present example, this resulted in a weight of 0.2088 on x_{GR} and of 0.3297 on $\bar{x}_{GR}$, a ratio of 0.2088/0.3297=0.633. This same ratio of weights is still present in the overall index (0.02698/0.04261=0.633). In the optimal index, however, the ratio of weights on x_{GR} versus $\bar{x}_{GR}$ is 0.02874/0.04011=0.7165.

Although the single-trait EBV index is not optimal, an important question is how much we would lose in accuracy when using this index instead of the optimal index. This can be studied by evaluating the accuracy of both indexes.

Accuracy of the multiple-trait EBV index must be derived from the full index with the three sources of information: $r_{HI} = \sqrt{\frac{\mathbf{b}_F'\mathbf{Gv}}{\mathbf{v}'\mathbf{Cv}}} = 0.5518$

The variance of the aggregate genotype $\mathbf{v'Cv}$ equals 45 in this example.

The accuracy of the single-trait EBV index can be derived in two ways:

1) as the accuracy of a sub-optimal index with the three information sources:

$$r_{HI} = \frac{\mathbf{b}_F^{*\prime}\mathbf{Gv}}{\sqrt{\mathbf{b}_F^{*\prime}\mathbf{Pb}_F^{*}\,\mathbf{v}'\mathbf{Cv}}}$$

2) or as the accuracy of an optimal index based on the two single-trait EBV

$$r_{HI} = \sqrt{\frac{\mathbf{b}_S'\mathbf{G}_S\mathbf{v}}{\mathbf{v}'\mathbf{Cv}}}$$

Both result in accuracy to be equal to 0.5511, which is only slightly smaller than the accuracy of the optimal index, which was 0.5518. Thus, in this case, use of single-trait EBV had limited impact on efficiency.

As an example of the use of principles similar to those outlined above, Veerkamp et al. (1995) investigated sensitivity to economic values and genetic parameters of a sire selection index of type and production traits for a breeding goal with production traits and longevity. They also investigated the efficiency of using index weights derived for an index with multiple-trait EBV in indexes that included single-trait rather than multiple-trait EBV and in indexes that included EBV from two multiple-trait evaluations, one for production traits and one for type traits. Results showed that losses in efficiency were less than 1% for EBV based on at least 50 daughter records, which confirms the robustness of selection indexes for progeny-tested dairy sires.

6.4 Sensitivity of Selection Indexes to Estimates of Variances and Covariances

As described so far, the selection index provides a method to maximize selection response for a given aggregate genotype when a given set of observations are available. It is assumed that the variances and covariances that make up the elements of **P**, **G** and **C** are known without error.

In practice, elements of **P**, **G** and **C** are estimated with error, and are often obtained from several different sources. Sales and Hill (1976) deal in detail with the ways in which errors in the estimates of variances and covariances affect efficiency of selection. For the present purposes, it is sufficient to note that, when the elements of **P** and **G** are not known without error, the selection index is often relatively insensitive to errors in estimates of these elements. However, it is always wise to examine how sensitive the resulting index is to the assumed elements of **P** and **G**. (Note that the elements of **C** do not affect selection index coefficients nor the prediction of response to selection.)

Bouchez and Goffinet (1990) propose a method for eliminating traits (or other sources of information) from the index based on maximizing an approximate estimate of the mean square error of prediction of the index, allowing for the inaccuracy of estimates of variances and covariances among traits. Their method deals with the situation where parameters are estimated from the same population as is being selected. Their method does not appear to increase selection response but involving fewer traits, is claimed to be more robust than a full index. An alternative approach would be to use Bayesian methods to simultaneously derive genetic parameters and selection indexes. Bayesian methods allow uncertainty about parameters to be included in development of the index. Literature estimates of genetic parameters can be included through specification of proper priors. Bayesian analyses are, however, typically complex in nature and an alternative series of simpler tests and modifications is proposed here.

6.4.1 Testing Consistency of Variance/Covariance Matrixes

Whenever a selection index is constructed, the **P** and **C** matrixes should be tested for being positive definite. A non-positive definite variance/covariance matrix may indicate that one or more traits in the index are linear functions of combinations of other traits, or that certain elements of the matrix are not possible given that the other elements are correct. For example, consider the phenotypic covariances among three traits with standardized variances such that, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. Then, if $\sigma_{12} = 0.9$ and $\sigma_{13} = 0.9$, common sense would tell us that σ_{23} must also take a relatively high value. A value of $\sigma_{23} = -0.5$ would clearly be impossible. It can be shown in this case that if

$$\mathbf{P} = \begin{bmatrix} 1 & .9 & .9 \\ .9 & 1 & \sigma_{23} \\ .9 & \sigma_{23} & 1 \end{bmatrix}$$

P becomes non-positive definite when $\sigma_{23} \leq 0.62$. At this point **P** also becomes singular and an inverse to **P** (i.e. $\mathbf{P}^{-1}$) cannot be found.

All matrix programming languages carry sub-routines to calculate the eigenvalues of a matrix. There are as many eigenvalues as there are rows and columns of the matrix (i.e. an *n* x *n* matrix has *n* eigenvalues). If one or more of the eigenvalues is negative, the matrix is non-positive definite.

Matrices which are close to being non-positive definite should also be examined closely because such matrixes are very close to being singular and are only just within the allowable parameter space, indicating possible inconsistencies in parameter values.

As noted by Sales and Hill (1976) and Hayes and Hill (1980), matrixes are more likely to become inconsistent (non-positive definite) as the number of traits considered increases, the size of the data base used to obtain estimates decreases, and when estimates are obtained from different sources. This in turn can lead to increasing instability and uncertainty over the resulting indexes and their responses. In general, this argues for keeping the number of traits in both the

index and the aggregate genotype down to the minimum number consistent with achieving effective genetic progress.

The **P** and **C** matrixes referred to above are those for observations in the index and traits in the aggregate genotype respectively. Often the elements of **P** are themselves derived from population phenotypic and genetic parameters. For example, observations in the selection index might be full-sib mean performance for a series of traits. Then, the variances and covariances among these observations would be derived from the phenotypic and genetic variances and covariances among single phenotypic records. We can denote $\mathbf{P}_O$ and $\mathbf{G}_O$ as the phenotypic and genetic variance covariance matrixes among single phenotypic records of traits appearing in the selection index, which we used to derive the elements of **P**. Then, if our records are to be biologically meaningful, $\mathbf{P}_O$ and $\mathbf{G}_O$ should also be within the allowable parameter space and should thus be tested for being positive-definite.

Similarly, if our model is $y = g + e$

then $\mathbf{E}_O = \mathbf{P}_O - \mathbf{G}_O$

is the implied environmental variance/covariance matrix among the traits in the index, given $\mathbf{P}_O$ and $\mathbf{G}_O$. Again, to be consistent and biologically meaningful, $\mathbf{E}_O$ should also be positive definite.

Variance/covariance matrixes can be biologically and economically inconsistent despite being positive definite. For example, a matrix for the three traits, milk yield (*M*), fat yield (*F*) and fat concentration (*f*) could be positive definite but not conform to the expectation that, since $f = \frac{F}{M}$, the variance of *f* and its covariances with *F* and *M* are entirely dependent on the variances and covariances among *F* and *M*. This is because the test of positive definiteness examines possible linear combinations of traits and does not recognize the possibility of there being a pre-specified linear or non-linear relationship, such as *f* being a ratio of *F/M*. In general, any trait that is a direct function of other traits in the index (or the aggregate genotype) will carry little or no additional information and as such is usually best omitted.

The economic weight of a trait is defined as the marginal change in economic value given a genetic change in that trait while holding all other traits in the aggregate genotype constant. Thus any trait in the aggregate genotype which is entirely a function of other traits in the aggregate genotype has an indeterminate economic value since it is impossible to make genetic change in that trait without making genetic change in other traits. As an example, again considering the traits *M*, *F* and *f*, where $f = \frac{F}{M}$, it is quite clear that there cannot be a change in *f* without a change in either *F* or *M* or both. Thus, if *F* and *M* appear in the aggregate genotype, *f* should not appear.

It is surprising how often these basic checks for consistency of variance/covariance components are not made in examples appearing in the literature, sometimes causing quite misleading results and conclusions. It is thus best not only to check over your own parameters for consistency, but

also to check parameters in published papers, unless the authors have quite clearly stated that the appropriate checks have been made.

In many cases, by elimination of unnecessary and less useful traits in the index, matrixes will become positive definite. Where matrixes remain non-positive definite, Hayes and Hill (1981) present a mathematical method ("bending") for altering parameters to obtain positive definite matrixes. This method assumes there is no additional information available about parameter estimates and as such is unlikely to be generally appropriate. An alternative is to adjust individual parameter estimates based on an assessment of reliability of the estimates. An example and further discussion is given in the following section. But there is certainly room for further investigation of this problem.

6.4.2 Tests of Sensitivity of Indexes to Assumed Variances and Covariances

If the various variance/covariance matrixes pass the test of being positive definite and being consistent in other ways, there is still a concern that the index derived may be sensitive to the assumed values of the variance/covariance components. A test of sensitivity would be to ask what proportion of the maximum response would we expect in the aggregate genotype if we used one set of variances to derive our index coefficients when another set of covariances was the correct set.

If we use the subscript u to describe the result of a particular set of parameters used and t to describe the set defined as true, then

$$\mathbf{b}_u = \mathbf{P}_u^{-1}\mathbf{G}_u\,\mathbf{v}$$

and

$$\mathbf{b}_t = \mathbf{P}_t^{-1}\mathbf{G}_t\,\mathbf{v}.$$

If $\mathbf{P}_t$ and $\mathbf{G}_t$ are the true variance/covariance matrixes, then, by definition, selection using index coefficients, $\mathbf{b}_t$, will give the maximum expected response in the aggregate genotype H. Thus use of $\mathbf{b}_u$ will give less response than $\mathbf{b}_t$. We can then define the efficiency of using $\mathbf{b}_u$ instead of $\mathbf{b}_t$ as the ratio of genetic superiority of selected individuals in H when selecting using $\mathbf{b}_u$ compared to when selecting using $\mathbf{b}_t$, i.e. $E_{ut} = \dfrac{S_{H_u}}{S_{H_t}}$

S_{H_u} can be found from the equation in Table 6.1 that defines response to selection from an arbitrary index and S_{H_t} from the equation which defines response to selection from the optimal index, so that

$$S_{H_u} = i\frac{\mathbf{b'}_u\,\mathbf{G}_t\,\mathbf{v}}{\sqrt{\mathbf{b'}_u\,\mathbf{P}_t\,\mathbf{b}_u}} \quad \text{and} \quad S_{H_t} = i\sqrt{\mathbf{b'}_t\,\mathbf{G}_t\,\mathbf{v}}$$

and thus

$$E_{ut} = \frac{\mathbf{b'}_u\,\mathbf{G}_t\,\mathbf{v}}{\sqrt{\mathbf{b'}_u\,\mathbf{P}_t\,\mathbf{b}_u}}\;\frac{1}{\sqrt{\mathbf{b'}_t\,\mathbf{G}_t\,\mathbf{v}}}$$

The process would be to define a parameter set to be our best possible estimate based on the evidence available and to use this as $\mathbf{G}_t$ and $\mathbf{P}_t$. Alternative sets would then be defined and the question asked of each alternative set, "if the alternative set were indeed true, how well would the index based on our best estimates perform?" There will likely be some difficulty in deciding what are realistic alternative sets of parameters. Clearly, if parameters are well estimated, there would be little relevance in examining widely different parameter sets. Whereas, if we are very uncertain of our estimated parameter set, we may well wish to examine widely different alternative sets. Also, the problem becomes worse the larger the number of elements in **P** and **G**, since the number of possible combinations of alternative parameter sets rapidly expands with increasing dimensions of **P** and **G**.

If there is insufficient evidence to say that the parameters used are the "best available" or "most likely", it may be worth considering an alternative set of parameters if one can be found that gives an index that is less sensitive to the parameters being incorrect than use of the original set of parameters. This could be judged by creating a matrix of relative efficiencies where each parameter set was tested against all other parameter sets, i.e.

$$\mathbf{E} = \begin{matrix} \\ \\ \\ \end{matrix} \overset{\textit{'true' parameters}}{\begin{bmatrix} E_{11} & E_{12} & E_{1n} \\ E_{21} & E_{22} & E_{2n} \\ \\ E_{n1} & E_{n2} & E_{nn} \end{bmatrix}} \textit{used parameters}$$

where the rows correspond to parameter sets used and the columns to the parameter sets assumed to be correct, so that: $E_{ij} = \dfrac{S_{H_{u_i}}}{S_{H_{t_j}}}$ where i and j are the used and true sets of parameters, resp.

The diagonal elements of **E** are 1.0, all other elements are <1.0. The parameter set could then be chosen for which the elements of the corresponding row of **E** were as close to 1.0 as possible, i.e. the parameter set which showed the least sensitivity to being incorrect in relation to the alternative parameter sets examined.

These processes of investigation are rather subjective. Indeed there is scope for research into the general area of how best to obtain estimates of population parameters, how to perform sensitivity analyses and how to modify parameters for particular applications in animal breeding. Bayesian methods provide opportunities to address these issues.

There is also the question of what action to take if our results are indeed sensitive to parameter values and what constitutes sensitivity? For example, if our best estimated parameter sets give response efficiencies E_u in the range of 0.85-1.0 for alternative parameter sets, would this be considered as sensitive or insensitive? There is no hard and fast rule, though by analogy with advances in methods of genetic evaluation (which improve accuracy of evaluation and hence response), potential losses in response of one or two percent may be of little concern whereas losses of 5% or more at the very least warrant careful investigation.

6.4.2.1 A Numerical Example of Testing Index Sensitivity to Variance/Covariance Estimates

Imagine a species such as swine or beef cattle, selected for growth performance, perhaps as a terminal sire. Assume that the traits in the aggregate genotype have been determined to be slaughter weight (SW) and feed intake during growth to slaughter (FI), that both these traits are recorded, and that a phenotypic selection index is used. When both traits are scaled so that σ_p = 1, economic weights of SW and FI were found to be 2 and -1 monetary units per phenotypic s.d. (Note that FI has a negative economic weight, reflecting that feed costs money; increased FI decreases profit.) Heritabilities of SW and FI have been estimated from several data sets and found to be 0.4 for both traits. Phenotypic and genetic correlations have been estimated only once and found to be 0.8 and 0.8. Thus, since selection is on phenotypic records of the individual, $\mathbf{P} = \begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix}$, $\mathbf{G} = \begin{bmatrix} .4 & .32 \\ .32 & .4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

Note that the off-diagonal element of *G*, $\sigma_{g_{SW.FI}}$, is obtained by recalling that $r_{g_{ij}} = \dfrac{\sigma_{g_{ij}}}{\sqrt{\sigma^2_{g_i}\sigma^2_{g_j}}}$,

hence $\sigma_{g_{SW.FI}} = 0.8\sqrt{0.4 * 0.4}$.

From these parameters and selection intensity *i* we obtain, $\mathbf{b} = \begin{bmatrix} .8 \\ -.4 \end{bmatrix}$ and $S_H = 0.5367\, i$.

In the present case, we might decide that heritabilities are well estimated, coming as they do from several published data sets, but we are uncertain about the estimated correlations. In general, phenotypic correlations are much more accurately estimated than genetic correlations. So we conclude that we are most uncertain about our genetic correlation, which we believe may in reality lie anywhere between 0.65 and 0.95.

We then investigate the following four situations for the correlations, all other parameters remaining constant:

1. $r_{g_{SW,FI}} = 0.80$ and $r_{p_{SW,FI}} = 0.8$
2. $r_{g_{SW,FI}} = 0.65$ and $r_{p_{SW,FI}} = 0.8$
3. $r_{g_{SW,FI}} = 0.95$ and $r_{p_{SW,FI}} = 0.8$
4. $r_{g_{SW,FI}} = 0.80$ and $r_{p_{SW,FI}} = 0.7$.

To calculate E_{21} we will need to determine the index weight corresponding to situation 2.

The **G** matrix in that case is: $\mathbf{G} = \begin{bmatrix} .4 & .26 \\ .26 & .4 \end{bmatrix}$

yielding new estimates of $\mathbf{b} = \begin{bmatrix} 1.233 \\ -0.867 \end{bmatrix}$

Response to selection when using the index weights for situation 2 while parameters for situation 1 are the true parameters is: $S_{H_{u_2,t_1}} = i \dfrac{\mathbf{b'}_2 \, \mathbf{G}_1 \, \mathbf{v}}{\sqrt{\mathbf{b'}_2 \, \mathbf{P}_1 \, \mathbf{b}_2}} = 0.5122 \, i$

The efficiency of using alternative parameter set 1 while the original set 1 is true is estimated as:

$$E_{21} = \frac{S_{H_{U_2,t_1}}}{S_{H_{t_1}}} = 0.5122/0.5367 = 0.9545$$

Similarly, efficiencies for all the other combinations of parameter sets can be calculated, which results in:

$$\mathbf{E} = \begin{bmatrix} 1 & 0.9545 & 0.8480 & 0.9738 \\ 0.9545 & 1 & 0.6513 & 0.8746 \\ 0.8480 & 0.6513 & 0 & 0.9258 \\ 0.9751 & 0.8646 & 0.9444 & 1 \end{bmatrix}$$

Two things are worth pointing out in these results. In the first place, uncertainty over the true value of the genetic correlation between SW and FI seems to be important in terms of response to selection since an index using a genetic correlation of 0.8 is only 84.8% effective if r_g is as high as 0.95, although it is 95.5% efficient if r_g is as low as 0.65. We conclude that, in terms of finding an optimal index for this selection program, we need to have a more accurate estimate of the genetic correlation between SW and FI than we have at present. In particular, it will be important to know whether this genetic correlation is higher than our current estimate of 0.8. A reduction of the phenotypic correlation from 0.8 to 0.7 has a smaller effect on efficiency than a similar change in the genetic correlation.

Secondly, the efficiency is symmetric about the diagonal for the first three-parameter sets but not for the last set. In the first three parameter sets **G** changed but **P** remains constant. For those cases it has been shown that the efficiency criteria is equal to the correlation between indices. Given that similarity, this symmetry is no surprise. Parameter set 4 resulted in a different **P** compared to the other three sets and as a consequence, for example, efficiency E_{41} is not the same as E_{14}.

Predicted optimal economic responses (S_H) for the first three parameter sets corresponding to r_g = 0.65, 0.8, and 0.95 were 0.749, 0.537, and 0.422 economic units per standard deviation of selection intensity. Predicted absolute economic responses to selection increased by 77.5% as r_g decreased from 0.95 to 0.65.

From experience, it seems to be a general phenomenon that altering variances and covariances has a much larger effect on the prediction of absolute economic response to selection than on the efficiency of one index versus another. Since the prediction of absolute economic response is used in assessment of the economic cost-benefits of selection programs, having good estimates of variances and covariances will often be of even more importance for cost-benefit assessments than for the optimization of a particular program in terms of the index to be used.

6.4.3 Uncertainty Over Economic Weights

As will be discussed later, economic weights are rarely known without error. Indeed, there is often considerable uncertainty of what the true economic weight might be, arising from uncertainties of biological and management models used and uncertainty about the values of different traits in future production systems and markets. In many cases we will therefore have a "best estimate" set of economic weights and several, perhaps many, alternatives, covering alternative present and future scenarios.

As with investigations of uncertain variances and covariances, we can carry out analogous investigations for uncertain economic weights. In this case the subscripts u and t refer to the economic weights used and those deemed to be "true". Then, efficiency of a set of index weights derived for the used set of economic weights when compared to that for the "true" set of economic weights is given by:

$$E_{ut} = \frac{S_{H_u}}{S_{H_t}} = \frac{\mathbf{b}'_u \mathbf{G} \mathbf{v}_t}{\sqrt{\mathbf{b}'_u \mathbf{P} \mathbf{b}_u}} \frac{1}{\sqrt{\mathbf{b}'_t \mathbf{G} \mathbf{v}_t}}$$

where $\mathbf{b}_u = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}_u$
and $\mathbf{b}_t = \mathbf{P}^{-1} \mathbf{G} \mathbf{v}_t$.

As with uncertainties over variances and covariances, E_{ut} could be derived for a single set of "most likely" or "best estimate" economic weights, compared to a variety of alternative economic weights. Or, a matrix of efficiencies, **E**, could be determined, comparing every set of economic weights with every other set. Economic weights do not affect **P** and, as shown in the previous section, **E** is symmetric about the diagonal, so that $E_{ij} = E_{ji}$.

6.5 The Value of Including Traits in the Selection Index

Recording of animals takes time and effort and consequently costs money. Some traits may be considerably more difficult and costly to record than others. Thus a key question in the design of breeding programs is which traits and types of relatives should be recorded? It is often relatively straightforward to identify a number of traits and types of relatives potentially useful to record. A selection index could then be constructed and the question asked, how much does each observation contribute to response in the aggregate genotype? Then, the economic benefits of including that observation in terms of enhanced response can be evaluated against the cost of recording and a decision taken on whether or not to collect that information. Consider the usual selection index problem with n observations in the index and m traits in the aggregate genotype, but define the index as including all traits that might possibly be included. Then, the efficiency of a reduced index without observation i can be defined as the ratio of economic (aggregate genotype) response for the reduced index to that with the full index, or:

$$E_i = \frac{S_{H_i^*}}{S_H}$$

where $S^*_{H_i}$ is the response in the aggregate genotype for selection on the index without observation i, $S_{H_i^*}$, and S_H is the response with the full index I,

$$S_{H_i^*} = i\sqrt{\mathbf{b}_i^{*}\mathbf{G}_i^{*}\mathbf{v}} = i\sigma_{I_i^*}$$

where $\mathbf{b}_i^*$ is $\mathbf{b}_i^* = \mathbf{P}_i^{*-1}\mathbf{G}_i^*\mathbf{v}$

and $\mathbf{P}_i^*$ and $\mathbf{G}_i^*$ are reduced forms of **P** and **G** corresponding with the reduced index. Matrix $\mathbf{P}_i^*$ is found by deleting the i^{th} row and column from **P**, and $\mathbf{G}_i^*$ by deleting the i^{th} row from **G**.

Similarly, the efficiency of indexes with more than one observation deleted can easily be found.

For an index with observations *i* and *j* missing: $E_{ij} = \dfrac{S_{H_{ij}^*}}{S_H}$

where $S_{H_{ij}^*} = \sqrt{\mathbf{b}_{ij}^{*}\mathbf{G}_{ij}^{*}\mathbf{v}}$

and $\mathbf{b}_{ij}^* = \mathbf{P}_{ij}^{*-1}\mathbf{G}_{ij}^*\mathbf{v}$

and $\mathbf{P}_{ij}^*$ is **P** with the i^{th} and j^{th} rows and the i^{th} and j^{th} columns removed and $\mathbf{G}_{ij}^*$ is **G** with the i^{th} and j^{th} rows removed.

The efficiency, given a certain aggregate genotype and set of population parameters, is directly related to changes in r_{HI}. The contribution of each individual observation in the index to the accuracy can be calculated. Cunningham (1969) has described a method to derive the decrease in accuracy which is given below.

Assume an index with *n* sources of information, $I = \mathbf{b'x}$ and an aggregate genotype with *m* traits, $H = \mathbf{v'g}$. Solving the set of equations, $\mathbf{Pb} = \mathbf{Gv}$, gives the vector **b**. The variance of the index is **b'Pb**. When, for instance, the first observation is ignored, we obtain a new index, I_1^*, with a vector with *n*-1 weighting factors, $\boldsymbol{\beta}$. The weighting factors are a solution of solving: $\mathbf{P}_1^*\boldsymbol{\beta} = \mathbf{G}_1^*\mathbf{v}$ where, as before, $\mathbf{P}_1^*$ is the *n*-1 by *n*-1 variance-covariance matrix of the observations in I_1^*, and $\mathbf{G}_1^*$ is the *n*-1 by *m* matrix obtained from **G** by ignoring the first row.

Based on the original equation, this can be visualized as:

$$\mathbf{Pb} = \mathbf{Gv}$$

$$\begin{bmatrix} P_{11} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P_1^*} \end{bmatrix}\begin{bmatrix} b_1 \\ \mathbf{b_0} \end{bmatrix} = \begin{bmatrix} G_1 \\ \mathbf{G_1^*} \end{bmatrix}\mathbf{v} = \begin{bmatrix} R_1 \\ \mathbf{R_0} \end{bmatrix}$$

where P_{11}, b_1, and R_1 are scalars representing the first elements in **P**, **Gv**, and **b**, respectively. Matrix **Q** is a column vector representing the remaining elements of the first column in **P**. In most cases, $\boldsymbol{\beta} \neq \mathbf{b}_0$.

The variance of the reduced index I_1^* is: $\sigma^2_{I_1^*} = \mathbf{\beta' P_1^* \beta}$

and its correlation with the aggregate genotype H is: $r_{HI_1^*} = \sqrt{\dfrac{\mathbf{\beta' P_1^* \beta}}{\sigma^2_H}}$

Efficiency of the index ignoring observation i is equal to $E_{I_i^*,I} = \dfrac{S_{H_i^*}}{S_H} = \dfrac{r_{H,I_i^*}}{r_{H,I}} = \dfrac{\sigma_{I_i^*}}{\sigma_I}$, which depends entirely on the difference in variances of the indices, $\mathbf{\beta' P_1^* \beta}$ and $\mathbf{b'Pb}$. To determine the magnitude of this difference, let $\mathbf{W}$ be the inverse matrix of $\mathbf{P}$: $\mathbf{W} = \mathbf{P}^{-1} = \begin{bmatrix} W_{11} \ldots\ldots\ldots\ldots W_{1n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ W_{n1} \ldots\ldots\ldots\ldots W_{nn} \end{bmatrix}$

It can be shown (Cunningham, 1969) that: $\sigma^2_I - \sigma^2_{I_1^*} = \dfrac{b_1^2}{W_{11}}$

In general, ignoring the i^{th} observation in the index: $\sigma^2_I - \sigma^2_{I_i^*} = \dfrac{b_i^2}{W_{ii}}$

The advantage of this method is that in computing the decrease in variance, no new index has to be derived; information for the computation is available from computations for the original index. Variance $\sigma^2_{I_1^*}$ can be derived from the equation above as: $\sigma^2_{I_i^*} = \sigma^2_I - \dfrac{b_i^2}{W_{ii}}$

Efficiency of the index ignoring observation i can also be derived from the above results as:

$$E_{I_i^*,I} = \frac{S_{H_i^*}}{S_H} = \frac{r_{H,I_i^*}}{r_{H,I}} = \sqrt{\frac{\sigma^2_I - \dfrac{b_i^2}{W_{ii}}}{\sigma^2_I}}$$

The method can be extended to compute the reduction of variance of the index when simultaneously ignoring more than one trait. The variance of an index, from which the i^{th} until the j^{th} observation is ignored, is $\sigma^2_{I^*_{i..j}} = \sigma^2_I - \mathbf{b}'_{i\,j} \mathbf{W}^{-1}_{i\,j} \mathbf{b}'_{i\,j}$

where $\mathbf{b}_{i..j}$ is the vector including the i^{th} until the j^{th} weighting factor of the original index, and $\mathbf{W}_{i..j}$ is the diagonal submatrix of $\mathbf{P}^{-1}$ corresponding with the observations that were removed from the index.

6.5.1 Example of A Reduced Index

Consider the example in section 6.5.2.1 for selection on slaughter weight (SW) and feed intake (FI). How much do inclusion of SW and FI in the index contribute to selection response?

The full index was given by $\mathbf{P} = \begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix}$, $\mathbf{G} = \begin{bmatrix} .4 & .32 \\ .32 & .4 \end{bmatrix}$ with $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

Giving $\mathbf{b} = \begin{bmatrix} .8 \\ -.4 \end{bmatrix}$, $\sigma_I^2 = .288$, and $S_H = 0.537\, i$

An index without SW would be found from

$\mathbf{P}_1^* = [1]$, $\mathbf{G}_1^* = [.32 \quad .4]$, and $\mathbf{v} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$

giving $\mathbf{b}_1^* = 0.24$ and $S_{H_1^*} = 0.24\, i$

and hence $E_1 = \dfrac{S_{H_1^*}}{S_H} = 0.447$

Similarly, removing FI from the original index, gives

$\mathbf{P}_2^* = [1]$, $\mathbf{G}_2^* = [.4 \quad .32]$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

giving $\mathbf{b}_2^* = 0.48$, $S_{H_2^*} = 0.48\, i$

and $E_2 = 0.894$.

These results could also have been obtained from **b** and $\mathbf{P}^{-1}$, which is equal to:

$$\mathbf{P}^{-1} = \mathbf{W} = \begin{bmatrix} 2.778 & -2.222 \\ -2.222 & 2.778 \end{bmatrix}$$

Thus: $$E_1 = \sqrt{\frac{\sigma_I^2 - \frac{b_1^2}{W_{11}}}{\sigma_I^2}} = \sqrt{\frac{.288 - \frac{.8^2}{2.778}}{.288}} = 0.447$$

and: $$E_2 = \sqrt{\frac{\sigma_I^2 - \frac{b_2^2}{W_{22}}}{\sigma_I^2}} = \sqrt{\frac{.288 - \frac{(-.4)^2}{2.778}}{.288}} = 0.894$$

Thus, not recording SW would reduce economic response by (1 - 0.447) 100 = 55.3%, while not recording FI would reduce economic response by only 10.6% compared to recording both traits. Since recording FI is likely to be more expensive than SW, it would seem well worthwhile to ask in detail how much it will cost to record FI in the breeding program and what the expected extra 10.6% economic response is worth. The final answer will depend on such parameters as the number of animals in our breeding program, the number of animals benefiting from the extra genetic response, the cost of recording individual animals, and the time between incurring costs (recording) and obtaining returns (selling or producing from genetically improved animals).

6.6 The Value of Traits in the Aggregate Genotype

With equation (6.11): $S_H = i\sqrt{\mathbf{b'Gv}}$, response in the aggregate genotype as a result of selection on the index can be calculated. Response in the aggregate genotype is an important criterion in comparing different indexes. In addition, it is also interesting to look at responses in the individual traits in the aggregate genotype. This will give information on which traits have contributed to response in the aggregate genotype and one might want to look at the direction and size of change in each of the traits. In the example of beef cattle, it is interesting to see whether response to selection is due to a change in slaughter weight, feed intake, or a combination of these two traits. In this section, we will illustrate the relationship between changes in the aggregate genotype and in individual traits. Furthermore, we will look at comparing response to selection for various situations and try to give some guidelines on the criteria to use in such a comparison.

The expected change in individual traits (g_j) of the aggregate genotype as a result of selection on the selection index (I) can be computed using the covariance between g_j and I, as shown in (6.18). Recall that the expected change in g_j (expressed in units of measurement) as a consequence of selection on I is:

$$S_{g_j} = i\frac{\mathbf{b'G}_j}{\sigma_I}$$

where $\mathbf{G}_j$ denotes the j^{th} column of $\mathbf{G}$, and that the vector of responses in each trait is equal to:

$$\boldsymbol{S}_g = i\,\frac{\mathbf{b'G}}{\sigma_I}$$

Response in the aggregate genotype, S_H in monetary units (assuming that elements of $\mathbf{v}$ are expressed in monetary units), can be calculated as:

$$S_H = i\frac{\mathbf{b'Gv}}{\sigma_I}$$

which is equal to:

$$S_H = i\frac{\mathbf{b'G}}{\sigma_I}\mathbf{v} = i\,\boldsymbol{S}_g\mathbf{v} = i\sum\frac{\mathbf{b'G}_j}{\sigma_I}\mathrm{v}_j$$

The above equation demonstrates that response in the aggregate genotype is the sum of responses in individual traits multiplied by their economic weights. The contribution of response in g_j to overall response (C_j) can be calculated as:

$$C_j = \frac{S_{g_j}\mathrm{v}_j}{S_H}$$

With this criterion we can evaluate the relative contribution of each trait in the aggregate genotype. To illustrate this, we return to the example of 6.5.2.1, use parameter set 1 and set selection intensity equal to 1. We have seen that response in the aggregate genotype, S_H, equals 0.537. Using (6.18), response in slaughter weight is 0.358 and response in feed intake is 0.179, both expressed in phenotypic standard deviations. The economic weight is 2 and -1 for SW and FI, respectively. Using these weights, response in monetary units in SW and FI is equal to \$0.716 and \$-0.179, respectively. The contribution of SW to overall response *(C_{SW})* is equal to 0.716/0.537 = 1.33. In words, response in SW accounts for 133% of the monetary response in the aggregate genotype. At first sight, it looks strange that the contribution exceeds 100%. When we have a closer look at the example, we see that we are trying to improve two traits

which are positively correlated, of which one has a positive and the other has a negative economic weight. Given the parameters we have used, the index results in an increase in slaughter weight, which is associated with a change in feed intake. The latter change results in a reduction in the financial gain. This is reflected by the negative contribution of 33% of FI to the overall response.

From the example it is clear that C_j can take values that are smaller than zero or bigger than one. The advantage of looking at values for C_j instead of S_{g_j} is particularly the unit that is used, i.e. money vs units of measurement.

Different indexes can be compared using accuracy of selection, as long as the parameters for the aggregate genotype (e.g. genetic variances and co-variances) and the economic weights have not changed. Comparison of indexes in other cases is less straightforward. Let us, for example, return to the example and consider a second index for a situation where feed intake was excluded from the aggregate genotype and for two different values of r_g, all other factors including the index remaining unaffected. Results for that and the original aggregate genotype are summarized in Table 6.2.

Table 6.2 Results of selection using a selection index including slaughter weight (SW) and feed intake (FI) for two different values of the genetic correlation between SW and FI r_g and economic weight of FI v_2 using a selection intensity of 1[a)].

	$r_g = 0.8$		$r_g = 0.65$	
	$v_2 = -1$	$v_2 = 0$	$v_2 = -1$	$v_2 = 0$
r_{HI}	0.632	0.632	0.765	0.652
σ_I	0.537	0.800	0.750	0.825
σ_H	0.849	1.265	0.980	1.265
S_H	0.537	0,800	0.750	0.835
S_{SW}	0.358	0.4	0.358	0.412
S_{FI}	0.179	0.32	-0.035	0.175
C_{SW}	1.33	1	0.953	1
C_{FI}	-0.33	0	0.047	0
Set	1	2	3	4

[a] The index in all cases consists of observations on FI and SW on the individual animal.

Let us first look at the situation where r_g is 0.8. The accuracy of r_{HI} was identical for both sets of economic weights but the variance of the aggregate genotype differed largely. We are looking at a situation with two highly correlated traits with opposite signs for their economic weights. Setting the economic weight equal to zero for one of the traits resulted in a higher variance of the aggregate genotype. The response in the aggregate genotype and slaughter weight increased when v_2 was set equal to zero. Correlated response in FI for $v_2 = 0$ amounted to 0.32 phenotypic standard deviations, which was much larger than for $v_2 = -1$. Changing the economic weight of FI did not affect r_{HI}. This is not a result that one would find in general, but one which is specific for the parameter set used in this example. There are two factors that contributed to this. In the first place, both traits had the same heritability and standard deviation. Schaeffer (1984) showed that the benefit of including a correlated trait in prediction of the EBV for a trait depends on the absolute difference between the phenotypic and genetic correlation. The correlated trait does not

contribute when the phenotypic correlation is equal to the genetic correlation, which is the case for $r_g = 0.8$. This means that $v_2 = 0$ refers to an index with observations on SW only for the correlation of 0.8. In that case one also expects to find $r_{HI} = \sqrt{0.4} = 0.632$.

Different results are found when r_g is lowered to 0.65. For $v_2 = -1$, a positive response of 0.358 units in SW is associated with a reduction in FI of 0.035 units, i.e. a change in the desired direction for both traits in the aggregate genotype. As to be expected, correlated response in FI is positive when $v_2 = 0$ but it is smaller in size than when the genetic correlation was 0.8.

In comparing the results in Table 6.2, it is obvious that there is not a single criterion which can be used. The criteria to look at very much depend on the question one wants to answer. It is important to realize that variation in the aggregate genotype differs between all four cases studied. This is different from a situation where one varies the observations to be included in the index. In that case one generally works with a constant σ_H^2 and can use the accuracy r_{HI}. In comparing the four situations in Table 6.2, the efficiency criteria introduced in section 6.5.2 might be very useful. Efficiencies for the different combinations of parameters sets, using the numbering of sets as given in Table 6.2, can be summarized as:

$$E = \begin{bmatrix} 1 & 0.954 & 0.894 & 0.976 \\ 0.954 & 1 & 0.720 & 0.867 \\ 0.894 & 0.720 & 1 & 0.970 \\ 0.976 & 0.867 & 0.970 & 1 \end{bmatrix}$$

This illustrates that when r_g is 0.8, loss in selection response amounts to (1 - 0.954)*100% = 4.6% when FI was ignored in the aggregate genotype, while the true economic value of FI is -1.

6.7 Non-Linear Indexes

The discussion of selection indexes so far has assumed that the aggregate genotype is a linear function of additive genetic values. In practice, the aggregate genotype will often not be linear. In general, however, linear indexes can be found which closely approximate most non-linear descriptions of economic value. These will be discussed in relation to derivation of economic weights in Chapter 7.

6.8 Constrained and Desired Gains Indexes

In the selection indexes discussed so far, response to selection of traits in the aggregate genotype is determined entirely by the economic weight of that trait, the information available in the index, and the phenotypic and genetic variances and covariances among traits. It is, however, possible to construct indexes in which the rate of genetic change in one or more traits is pre-determined. For example, it is possible to maximize an index for genetic change in one set of traits subject to other traits being constrained to zero genetic change. Alternatively, the change

in one trait might be desired to be twice the change in another, with all other traits being allowed to change as the index dictates.

Methods for achieving these types of constrained or desired gain indexes were reviewed by Brascamp (1984). Linear programming (see later Chapter) can also be used to achieve constrained and desired gain indexes (Keller and Gibson, 1990; Toro, 1992).

Following Brascamp (1984), separate the vector of traits in the aggregate genotype $H = \mathbf{v}'\mathbf{g}$ into two sets of traits: $\mathbf{g} = \begin{bmatrix} \mathbf{g_0} \\ \mathbf{g_1} \end{bmatrix}$

where $\mathbf{g_0}$ includes traits for which progress will be maximized according to $\mathbf{v_0}'\mathbf{g_0}$, i.e. the economic part of the breeding goal

and $\mathbf{g_1}$ includes traits for which changes are constrained to relative changes = $\boldsymbol{\delta}$.

Correspondingly, partition matrix **G** as: $\mathbf{G} = \begin{bmatrix} \mathbf{G_0} \\ \mathbf{G_1} \end{bmatrix}$

Let the constrained index be: $I^* = \mathbf{b}^{*\prime}\mathbf{X}$ with variance $\sigma^2_{I^*} = \mathbf{b}^{*\prime}\mathbf{P}\mathbf{b}^*$

For simplicity, and without loss of generality, set $\sigma^2_{I^*} = 1$

Then, from Table 6.1, response in $\mathbf{g_1}$ to selection on this index is proportional to $\mathbf{b}^{*\prime}\mathbf{G_1}$

Thus, the problem of finding the constrained index can be formulated as the following constrained optimization problem:

$$\underset{\mathbf{b}^*}{Max}\, \mathbf{b}^{*\prime}\mathbf{G_0}\mathbf{v_0} \quad \text{subject to } \mathbf{b}^{*\prime}\mathbf{G_1} = \alpha\boldsymbol{\delta} \quad \text{and} \quad \mathbf{b}^{*\prime}\mathbf{P}\mathbf{b}^* = 1$$

This problem can be solved using the Lagrange multiplier method, as described in Brascamp (1984) and Weller (1994).

Although there is considerable literature on the construction and properties of such indexes, at no point does there appear in the literature a sound reasoning for applying these indexes in economic animal breeding. Gibson and Kennedy (1991) attempted to provide a rationale for why such indexes should not be used in animal breeding, on the grounds that an economic index can always be found to at least equal and usually out-perform a constrained index. They showed that constrained or desired gains indexes have implied, or ***pseudo economic values*** for all traits, including for those that are constrained. The vector of pseudo economic values, **v***, is the set of economic values that would have resulted in the index weights **b***, using the selection index equations. Values for the pseudo economic values can be derived as follows:

Using the selection index equations, set $\mathbf{P}\mathbf{b}^* = \mathbf{G}\mathbf{v}^*$ and solve for **v*** as follows:

$$\mathbf{G}'\mathbf{P}\mathbf{b}^* = \mathbf{G}'\mathbf{G}\mathbf{v}^* \quad \rightarrow \quad \mathbf{v}^* = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{P}\mathbf{b}^*$$

6.8.1 Example of Desired Gains Index

Figures 6.2 and 6.3 illustrate the potential impact of putting a constraint on genetic progress for a given trait. The example is of selection for an aggregate genotype in pigs consisting of growth, feed intake, and backfat thickness. Economic values are 0.178 \$/g/d for growth, -0.05 \$/g/d for feed intake, and –0.0415 \$/mm for backfat thickness:

$$H = 0.178\, g_{growth} - 0.05\, g_{feed\ intake} - 0.0415\, g_{backfat}$$

The index includes phenotype on the individual and its sire for all three traits. Based on these economic values, unit intensity of selection on the optimal index results in an increase backfat thickness by 0.7316 mm (Figure 6.2). In many breeding programs, such an increase is undesirable and, thus it seems reasonable to develop a constrained index in which genetic gain for backfat thickness is constrained to zero. Results for such an index are in Figure 6.3. For this index, the relative pseudo economic values are:

$$\mathbf{v^* = (G'G)^{-1}G'Pb^*} = \begin{bmatrix} 0.02659 \\ -0.007469 \\ -0.585 \end{bmatrix} \text{ compared to } \mathbf{v} = \begin{bmatrix} 0.178 \\ -0.05 \\ -0.0415 \end{bmatrix} \text{ for the optimal index.}$$

Note that economic values in **v*** are relative economic values. When expressed relative to an economic value of 0.178 for growth rate, which is its true economic value, implied economic values under the constrained index become: $\mathbf{v^*} = \begin{bmatrix} 0.178 \\ -0.05 \\ -3.916 \end{bmatrix}$

Thus, the implied economic value for fat thickness relative to growth is almost 100 times as large as its true economic value.

Efficiency of the constrained index in terms of response in the true aggregate genotype can be evaluated using the procedures described in sections 6.5.2 and 6.5.3. Alternatively, with genetic superiorities in the individual traits already computed for the desired gains index, response in H can be computed following section 6.7 as:

$$S_H^* = \boldsymbol{S}_g^* \mathbf{v} = \begin{bmatrix} 13.65, -67.37, 0 \end{bmatrix} \begin{bmatrix} 0.178 \\ -0.05 \\ -0.0415 \end{bmatrix} = \$5.80$$

In contrast, response for the optimal index is (see Figure 6.2) $S_H = \$7.39$

Thus, efficiency is: $E = \dfrac{S_H^*}{S_H} = \dfrac{5.80}{7.39} = 0.785,$ a 21.5% loss.

Figure 6.2 Selind program output for optimal economic index for pig selection example

```
Listing of used matrices

Listing of P-matrix
      7744         968   1.144e+04         220       47.52       27.74
       968        7744         220   1.144e+04       27.74       47.52
 1.144e+04         220       4e+04        5000          56       29.66
       220   1.144e+04        5000       4e+04       29.66          56
     47.52       27.74          56       29.66           4         1.1
     27.74       47.52       29.66          56         1.1           4

Listing of G-matrix
      1936         440       55.47
       968         220       27.74
       440       1e+04       59.33
       220        5000       29.66
     55.47       59.33         2.2
     27.74       29.66         1.1

Listing of C-matrix
      1936         440       55.47
       440       1e+04       59.33
     55.47       59.33         2.2

Listing of P-inverse-matrix
 0.0002572  -6.187e-05  -7.397e-05   2.736e-05   -0.001678  -0.0004214
-6.187e-05   0.0002572   2.736e-05  -7.397e-05  -0.0004214   -0.001678
-7.397e-05   2.736e-05   4.751e-05  -1.344e-05   0.0001265  -1.121e-05
 2.736e-05  -7.397e-05  -1.344e-05   4.751e-05  -1.121e-05   0.0001265
 -0.001678  -0.0004214   0.0001265  -1.121e-05      0.2886    -0.06352
-0.0004214   -0.001678  -1.121e-05   0.0001265    -0.06352      0.2886

Index based on economic weights
-----------------------------------------------------
|               trait |   ec weight |  genetic sup |
-----------------------------------------------------
|              growth |       0.178 |        30.88 |
|         feed intake |       -0.05 |       -38.39 |
|           fat thick |     -0.0415 |       0.7316 |
-----------------------------------------------------
-------------------------------------------------------------------------
|              animal ,               trait |     B-value |     rvi (%) |
-------------------------------------------------------------------------
|     own performance ,              growth |     0.08516 |        30.5 |
|                sire ,              growth |     0.01686 |       1.018 |
|     own performance ,         feed intake |    -0.03579 |       28.88 |
|                sire ,         feed intake |   -0.007106 |      0.9787 |
|     own performance ,           fat thick |       1.095 |       3.882 |
|                sire ,           fat thick |      0.1251 |     0.04967 |
-------------------------------------------------------------------------
Standard deviation of index  : 7.3861
      of aggregate genotype  : 8.8283
Correlation between index and aggregate genotype  : 0.83663
```

Figure 6.3 Selind program output for index for pig improvement example with gain for backfat thickness restricted to zero

```
RESULTS SECOND RUN WITH D AS DESIRED GAINS INPUT

Listing of Gtranspose*P-inverse*G
1252  -136.6  38.44
-136.6  5162  27.62
38.44  27.62  1.395

Listing of S
0.05653  0.011  -1.776
0.011  0.002356  -0.3498
-1.776  -0.3498  56.6
```

Pseudo relative economic values based on: $\mathbf{v^* = (G'G)^{-1}G'Pb^*}$

trait	ec weight	genetic sup	TRUE EC. VALUES
growth	0.02659	13.65	0.178
feed intake	-0.007469	-67.37	-0.05
fat thick	-0.585	2.313e-15	-0.0415

```
ECONOMIC RESPONSE = (0.178)*(13.65)+(-0.05)*((-67.37)+(-0.0415)*(0) = 5.80

      Compared to 7.3861 for the optimal economic index (21.5% reduction!!)
```

animal ,	trait	b-value	rvi (%)
own performance ,	growth	0.009932	30.1
sire ,	growth	0.002312	1.395
own performance ,	feed intake	-0.004965	44.49
sire ,	feed intake	-0.001173	1.95
own performance ,	fat thick	-0.1071	2.685
sire ,	fat thick	-0.04554	0.4801

```
Standard deviation of index  : 0.86603
      of aggregate genotype  : 1.139
Correlation between index and aggregate genotype  : 0.76033
```

For H with implied economic values

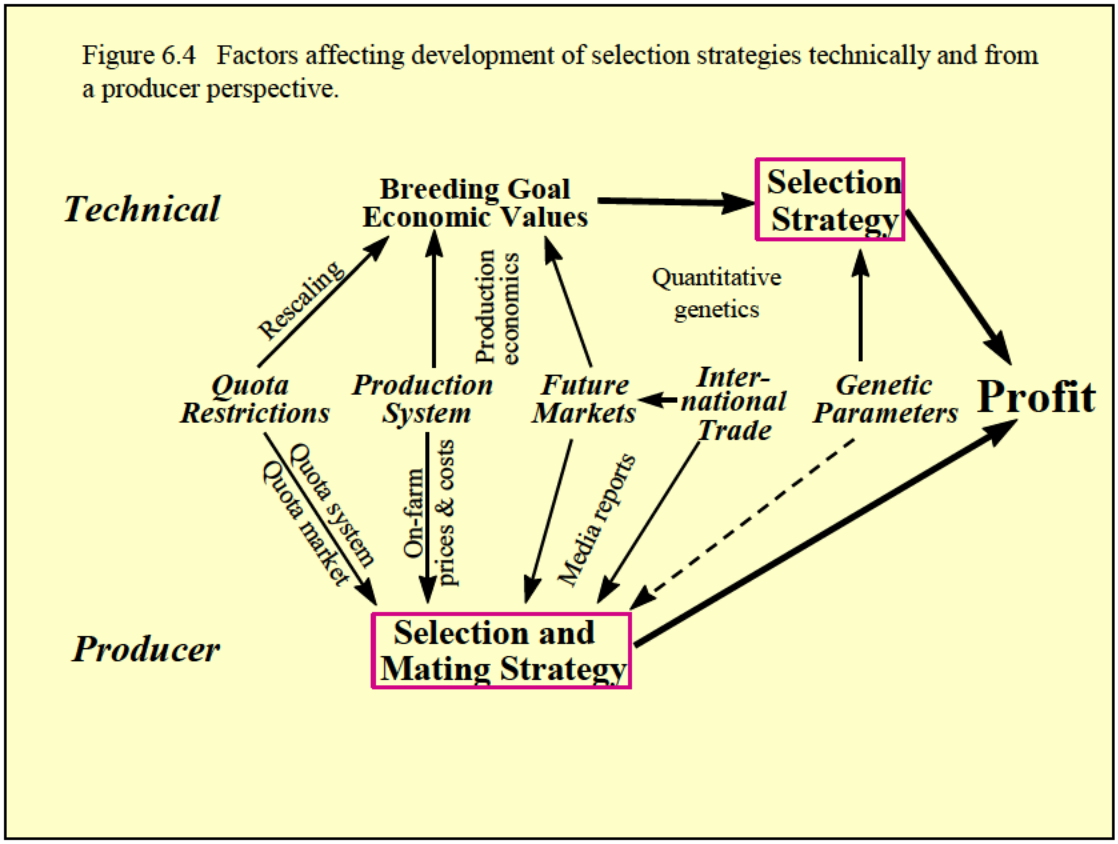
6.9 Some Practical Considerations (Taken from Dekkers and Gibson, 1997, J. Dairy Sci.)

The objective of development of total merit selection indexes is to provide a guide to selection that maximizes selection for the overall breeding objective and to promote selection strategies that minimize misuse of information. Although development of total merit selection criteria can be based entirely on the scientific principles that were summarised above and that will be described further in Chapter 7 with regard to estimation of economic values, implementation of selection indexes in the industry requires careful consideration of acceptance of the index by its target group (i.e., the breeders or producers). An index that is based on the application of sound scientific

principles but not accepted, has much less impact on selection for an overall breeding goal than an index that may not be optimal technically but that receives acceptance by the targeted users.

The main issues involved in the development of selection strategies from a technical versus a practical perspective are summarized and contrasted in Figure 6.4. To overcome these differences in perspective, development of selection indexes for use in the field must take place in close collaboration with the industry. Although this interaction may compromise scientific principles and objectivity, there are several aspects that can be pursued from a technical basis that can aid in development of an index that is technically sound while simultaneously maximizing acceptance of the index by its intended users. Several of these approaches, as well as their theoretical background, are discussed here (taken from Dekkers and Gibson, 1997), while others have been discussed earlier in this Chapter and in Chapter 7.

Figure 6.4 Factors affecting development of selection strategies technically and from a producer perspective.



6.9.1 Development of Selection Criteria

A total merit selection index combines EBV (or ETA or EPD's) for individual traits into a single number that can be used for selection. Although this index facilitates simultaneous selection for multiple traits, most breeders require insight into the index and, in particular, into the index weights. When derived based on selection index principles, weights in a selection index are equivalent to partial regression coefficients in a multiple regression equation. Similar to the difficulty of interpreting partial regression coefficients in a multiple regression analysis when explanatory variables are confounded, selection index weights can be difficult to interpret when

traits are genetically correlated. This complicates extension of the index to producers. There are several avenues that can be pursued in a) derivation of the index, b) expression of the index, or c) extension, to facilitate the understanding and, thereby, implementation of the index by its target audience.

Index derivation. Four approaches, among other, that can be used to develop total merit indexes that facilitates their understanding, acceptance, and implementation of the index are 1) alignment of index weights with economic values, 2) evaluation of the accuracy of alternative index weights, 3) use of customized indexes, and 4) development of indexes based on subindexes. In principle, all four approaches focus on consideration of alternative indexes that are more acceptable or easier to interpret and on exploring the accuracy of these indexes relative to the optimum index. Flexibility in consideration of alternative indexes without compromising accuracy stems from the robustness of genetic change to economic values, which was first explored by Vandepitte and Hazel (1977) and Smith (1983).

Aligning index weights with economic values.
Economic values of traits in the breeding goal are frequently easier to interpret than are selection index weights. Therefore, aligning index weights closely to economic weights in the breeding goal will facilitate understanding of the index. The ability to accomplish this depends on the degree to which traits in the index differ from traits in the breeding goal, whether single- or multiple-trait procedures are used for genetic evaluation of traits in the index, and, in the case of single-trait evaluation procedures, on the accuracy of EBV.

For index traits that are indicators of traits in the breeding goal, index weights can be related to their indirect economic importance based on the relationship between the indicator trait and the economic trait in the breeding goal. For example, for SCC as an indicator of susceptibility to mastitis, an indirect economic value can be derived as the genetic regression coefficient of SCC on mastitis, multiplied by the economic value of mastitis. This derivation was implicit to the economic value of SCC derived by Schutz (1994).

When EBV for traits in the index are based on a joint procedure for multiple-trait evaluation and traits in the index are the same as the traits in the breeding goal, economic values can be used directly as selection index weights, as described previously. When traits in the index are different from traits in the breeding goal, procedures described in previously can be used, and resulting index weights amount to the indirect economic values described.

When EBV are from single-trait genetic evaluation models, index weights can still be approximated based on economic weights if the accuracy of individual EBV is high. The efficiency of indexes derived on this basis relative to the optimum index must be considered in these cases.

Exploring the accuracy of alternative index weights.
Optimum index weights for a given set of economic values can be difficult to accept by producers if those weights do not correspond to the perceived incentives for genetic change, as discussed previously. For example, Gibson et al. (1992) found optimum index weights for Ontario to result in a negative weighting for milk yield, although producers were paid in part for milk volume, in addition to payments for fat and protein yield. The negative weighting on milk yield was a result of

the nature of the dual quota system that was in operation in Ontario at the time. In this sytem, one quota was based on volume for fluid milk sales, with a substantial premium for milk volume, and the second quota was based on fat sales, with no premium for milk volume. Because of the partial payment for volume, an index with a negative weighting on milk was difficult to accept by producers. An index with a zero weighting on milk volume was explored as an alternative. This index was found to be over 98.5% as efficient as the optimal index and was subsequently implemented by the industry as part of the LPI. Another aspect that helped implementation of this index was the demonstration that selecting on this index was expected to result in substantial improvement in milk yield, despite a zero emphasis on milk in the index.

Customised Indexes.
Customized indexes allow producers to develop a selection index based on economic circumstances that are specific to their herd Bowman et al. (1996). The use of customized indexes is justified technically if economic circumstances differ between herds or if traits are genetically different across herds Visscher and Amer (1996), which would make use of a single index derived based on economic circumstances of an average farm inappropriate. Visscher and Amer (1996) found that, for economic and genetic parameters typical for dairy cattle, customized indexes did not result in substantially greater improvements in profit at the population level than a single index that was based on average parameters. This result was caused in particular by the dominating economic importance of production traits and the high genetic correlations among milk, fat, and protein production. Visscher and Amer (1996) concluded that the main reason for use of customized indexes in dairy cattle is promoting the use of a selection index approach to multiple-trait selection instead of selection on independent culling levels. In development of customized indexes care must be taken, however, that objective economic information is used in their derivation; Visscher and Amer (1996) assumed that customized indexes were based on accurate economic data for individual farms.

Although customized indexes can play an important role in implementation of selection index principles in the field, we do advocate availability and promotion of an overall index that applies to average circumstances in the population. Such an index can provide default values for weights in customized indexes and provide a unified and global focus for selection and marketing decisions for all sectors of the breeding industry.

Development of indexes based on subindexes.
This was described previously.

6.9.1 Expression of the index

Interpretation and implementation of an overall selection index can be facilitated by its expression. Expression of this index includes the name given to the index which must convey the purpose and meaning of the index, expression of the index formula, and the scale on which the index is expressed. The latter two issues are discussed further.

Expression of the index formula.
The magnitude of economic values and index weights is dependent on the scale of expression of EBV and may not reflect the relative emphasis on traits in relation to genetic selection decisions. Multiplying index weights by the genetic standard deviation for the trait provides standardized weights that reflect the emphasis put on each trait in relation to the genetic variability that is present in the population.

Index weights can also be standardized by the standard deviation of EBV. Although the type of standardization (i.e., based on the standard deviation of true versus estimated breeding values) does not affect the eventual index values, there are some significant differences with regard to interpretation of the resulting index weights. These are summarised in the following: 1) index weights standardized by the genetic standard deviation are independent of the accuracy of genetic evaluations and are, therefore, more closely related to the economic importance of traits in the breeding goal; 2) standardization on the basis of the standard deviation of EBV reduces the emphasis that is perceived to be put on traits with low heritability (i.e., low accuracy of EBV). Standardization on the basis of the genetic standard deviation maintains the distinction between the ability to identify genetic differences in the population and the relative emphasis that is put on a given difference in EBV between animals; 3) the relative magnitude of the standard deviation of EBV among traits may differ between, for example, sires and cows. This would result in different indexes if standardization is on the basis of the standard deviation of EBV.

Scale of expression.
A total merit index ranks sires and cows based on genetic merit for the overall breeding goal. Apart from use of the index as a ranking tool, understanding, implementation and use of the index can be enhanced if the index is expressed in meaningful units. For total merit indexes that are intended to rank animals based on genetic merit for profitability, expressing the index in monetary units facilitates and promotes use of the index. The ultimate goal is to express the index in a way that enables its use for investment decisions (e.g., semen purchase). Expression of total merit indexes for sires as a net present value of a dose of semen was explored by McGilliard (1978), Bakker et al. (1980), and others. This requires consideration of the time and frequency of expression of genetic superiority in resulting daughters, discount rate, and conception and survival rates. Based on such indexes, differences between bulls in expected returns from a dose of semen can be compared directly to differences in semen price.

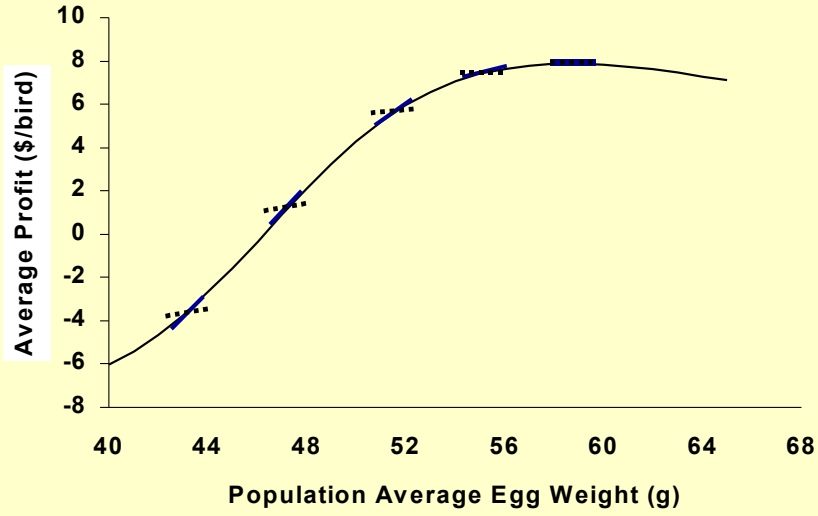
6.9.2 Extension of the index

Effective implementation of any index relies on promotion of the index through extension activities. Extension efforts should focus on the consequences of selection on an index rather than on the composition of the index. Index weights or economic values may not give a clear indication of what can be expected from use of the index. This is because responses to selection on an index are affected by genetic constraints on improvement of individual traits, which are quantified by the genetic parameters, as well as by the emphasis put on each trait in the index through the index weights.

Responses to selection can also be used to illustrate that, although for dairy cattle an increased protein to fat ratio may seem a reasonable selection objective, and selection indexes can be derived that maximise that objective, selection on such an index may have detrimental effects on responses for yield traits and result in reduced fat and protein yields.

Another example of the use of predicted responses to index selection in extension is to alleviate concerns regarding indexes with a negative weight on SCC as indicators of susceptibility to mastitis. Such indexes are often perceived to reduce SCC, which raises concerns about reaching levels of SCC that are too low to manifest an effective resistance to infection. Consideration of responses to selection on an index that includes production and SCC illustrates that such an index does not result in dramatic reductions in SCC but may instead reduce the rise in SCC that is the result of the positive genetic correlation between SCC and production.

Figure 7.5 Economic values for multiple generation selection



The example chosen by Dekkers et al. (1995) was for a rather extreme practical situation, with a highly nonlinear profit function involving an optimum (for egg weight). Thus, it seems likely that multiple generation optimization of index weights would rarely, if ever, be necessary in practice. The simpler recursive procedure to derive economic weights based on progeny generation performance might give one or two percent extra gains in some practical situations. But in most situations, the classical approach of using partial derivatives of profit based on current generation means, will be sufficiently accurate, rarely giving more than one or two percent less gain in profit than an optimized procedure.

7.3.6 Non-Linear Indexes for Non-Linear Profit Functions

Although most profit functions are non-linear, in the previous we only considered linear indexes. At first glance, it seems reasonable to assume that a non-linear index would be better than a non-linear index. In this section, non-linear indexes will be introduced, followed by a discussion of using linear versus non-linear indexes.

Wilton et al. (1968) showed that with quadratic profit functions, quadratic indexes of quadratic aggregate genotypes can be defined that are maximum likelihood solutions to maximizing economic genetic progress.

For example, if the profit function took the form $P = \pi_1 y + \pi_2 y^2$
then the aggregate genotype could be defined as $H = v_1(\mu_g + g) + v_2(\mu_g + g)^2$

and the selection index as $I = bx_1 + b_2 x_2$

where μ_g is the mean genotypic value of the trait, x_1 is an observation (e.g. phenotype, full-sib mean, etc.) based on y, and x_2 is the equivalent observation based on y^2 (e.g. mean phenotype squared, full-sib mean of squared phenotypes, etc.). The term μ_g is introduced into H because the economic merit of an animal relative to other animals is now, because of the quadratic profit relationship, dependent on the population mean. With a linear aggregate genotype, relative economic values are independent of the mean genotypic value. Optimal weights for the quadratic index were derived by minimizing the sum of squared differences between the index and genetic merit of selection candidates for the profit function. Solutions to the problem were given by Wilton et al. (1968) and are not dealt with here.

Given that we have phenotypic and genetic variances and covariances for y, it is straightforward to derive them for y^2 and to derive covariances between y and y^2 (see Appendix B). And hence it is straightforward, though obviously a little more tedious, to derive quadratic indexes. However, if the initial variance/covariance matrixes are close to being non-positive definite, adding the additional terms for y^2 will often cause the matrixes to become non-positive definite. This may sometimes happen when the initial matrixes are not close to being non-positive definite.

Ronning (1971) extended the method of Wilton et al. (1968) to derive a cubic index for a cubic profit function.

In several practical applications, a non-linear index is developed by substituting multiple trait EBV directly into the profit function.

E.g., for a profit function $P = f(\boldsymbol{\mu}+\mathbf{g})$

the selection index is $I = f(\boldsymbol{\mu}+\hat{\mathbf{g}})$.

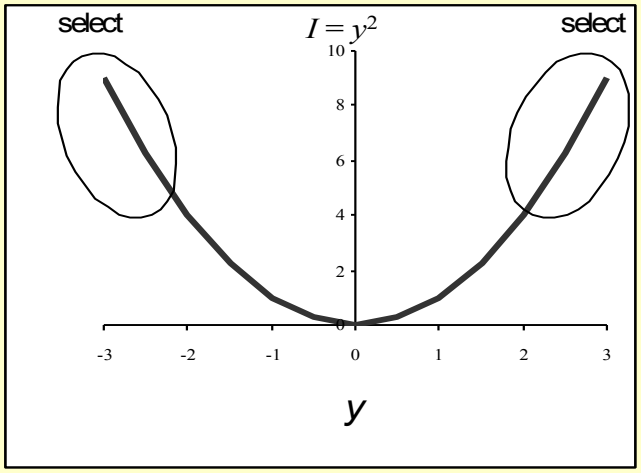
7.3.7 Non-Linear vs. Linear Indexes for Non-Linear Profit Functions

Goddard (1983) argued that the quadratic index of Wilton et al. (1968) does not maximize profit of progeny. The reason is that the index is derived to maximize the correlation of the index with genetic merit for profit of the *selection candidates*. This does not necessarily maximize the correlation of the index of selection candidates and genetic merit for profit of their *progeny*.

To illustrate the inadequacy of a quadratic index, Goddard (1983) used an example of a profit function $P = y^2$, where y is an additive trait with heritability 1.0 and $\bar{y}=0$. For this case, the index is $I = y^2$. As illustrated in Figure 7.6, selection on this index would select individuals with high y, as well as individuals with low y. Mating the selected parents (at random), would result in a progeny generation for which the genetic mean for the trait still is zero. Thus, there is no response to selection.

This can be referred to as profit heterosis

Figure 7.6: Illustration of selection based on a quadratic index



The reason for the lack of response to selection on the quadratic index is that, although all variation in profit is genetic, additive genetic variance for profit is zero. In fact, all genetic variance for profit is epistatic (additive x additive), which is not inherited from parents to progeny. This despite the fact that all genetic variance for the trait y is additive. The presence of non-genetic variance is a general property of non-linear profit functions and forms the basis of the concept of profit heterosis of Moav (1966). Presence of non-additive variance for profit also implies that mean profit can be increased by assortative mating. And indeed, when selecting individuals based on the quadratic index, mean profit of the progeny will be increased if individuals with high negative trait values for y are mated to each other and, similarly, individuals with high positive trait values are mated. Ultimately, this would result in the development of two separate lines, one with high y and one with low y.

For the example of $P = y^2$, the first derivative of the profit function evaluated at the current population mean (=0) is equal to zero. Thus a linear index $I=by$ that is derived based on the current population mean would have b=0 (b can be set equal to v = 0 because heritability = 1) and would therefore be equal to zero for all individuals and result in no response to selection. However, when the linear index is derived based on the first derivative of the profit function at the mean of the progeny, as in section 7.3.5.1, the index weight b will be non-zero. In this case, the linear index will either select individuals with high negative values for y, if b<0, or select individuals with high positive values, if b>0, but not both. This will result in a change in the population mean and an increase in mean profit.

More generally, Meuwissen and Goddard (1997) explained the sub-optimality of a non-linear index that is developed by substituting multiple trait EBV in the profit function, i.e. $f(\boldsymbol{\mu}+\hat{\mathbf{g}})$, as follows: if the genetic traits are additive and distributed multivariate normal, $\hat{\mathbf{g}}$ are maximum likelihood estimates of the genetic value of the individual for component traits and, therefore, $f(\boldsymbol{\mu}+\hat{\mathbf{g}})$ provides a maximum likelihood estimate of the genetic value of the animal for profit . If profit is a non-linear function and, therefore, includes non-additive genetic variation, $f(\boldsymbol{\mu}+\hat{\mathbf{g}})$ is not guaranteed to provide a maximum likelihood estimate of the genetic value of the progeny.

Although the previous illustrates that the quadratic index of Wilton et al. (1968), and non-linear indexes in general, do not maximize response in profit, it does not imply that the best index is a linear index. To pursue this, it is important to distinguish between the following two selection objectives:

a) maximize profit of an average progeny, i.e. Max[$f(\bar{\boldsymbol{y}}_{t+1})$]

b) maximize the average profit of progeny, i.e. Max[$\overline{f(\boldsymbol{y}_{t+1})}$]

Note that in section 7.3.5.1, the first objective was used to derive the optimal linear index.

For a linear profit function, it is clear that $f(\bar{\boldsymbol{y}}_{t+1}) = \overline{f(\boldsymbol{y}_{t+1})}$ and the same index maximizes both objectives. The same holds for a quadratic profit function if genetic traits are distributed multivariate normal. To see this, consider the following general quadratic profit function:

$$P = f(\boldsymbol{y}) = \mathbf{a}'\boldsymbol{y} + \boldsymbol{y}'\mathbf{A}\boldsymbol{y}$$

where **a** and **A** are a vector and matrix of constants.

Then $$E[f(\boldsymbol{y})] = E[\mathbf{a}'\boldsymbol{y} + \boldsymbol{y}'\mathbf{A}\boldsymbol{y}] = \mathbf{a}'E[\boldsymbol{y}] + E[\boldsymbol{y}']\mathbf{A}E[\boldsymbol{y}] + tr(\mathbf{AV})$$

where *tr* is the trace operator, and **V** is the variance-covariance matrix of $\boldsymbol{y}$.

Now, since $tr(\mathbf{AV})$ is equal to a constant, $E[f(\boldsymbol{y})] = f(E[\mathbf{y}]) + \text{constant}$

Thus, for a quadratic profit function with multi-variate normal variables, $E[f(\boldsymbol{y})]$ and $f(E[\mathbf{y}])$ only differ by a constant and, thus, maximizing $f(\bar{\boldsymbol{y}}_{t+1})$ is equivalent to maximizing $\overline{f(\boldsymbol{y}_{t+1})}$ and the two objectives are equivalent. For profit functions with higher degrees of non-linearity, however, the two objectives are not equivalent.

We will first consider the first objective, i.e. maximizing profit of an average progeny, for which the optimal linear index was derived in section 7.3.5.1. Goddard (1983) provided an intuitive proof that for the index that maximizes profit of an average progeny is a linear index regardless of the degree of non-linearity of the profit function if the traits that contribute to profit are additive. Itoh and Yamada (1988) provided a formal proof. The intuitive proof is based on the fact that the greatest genetic change in any direction in the multi-dimensional space of population trait means will be achieved by a linear index when the traits are additive. Referring to Figure 7.4, the origin is the current combination of population means for the two traits and the circles represent the response circle for all possible linear indexes for a given selection intensity. Consider the second response circle. As suggested, a linear index will result in the greatest change in the combination of population means in any direction and thus in the greatest distance of any point on the circle from the current population mean A. Any non-linear index will result in a combination of population means that lies within the response circle. Then, when the objective is to maximize $f(\bar{\boldsymbol{y}}_{t+1})$, two possibilities exist:

1) The profit function has no maximum within the response circle. In this case there will be a point on the response circle that will result in greater $f(\bar{\boldsymbol{y}}_{t+1})$ than any other point on or within the response circle. In Figure 7.4 this is represented by point *b* on the second response circle, for which the response circle is tangent with the furthest profit contour.

2) The profit function has a maximum within the response circle. In this case we can reduce the selection intensity such that the maximum falls on the response circle for linear indexes. Now situation 1) applies and, thus, a linear index results in at least as much gain as a non-linear index.

Thus, if the objective is to maximize profit of an average progeny, a linear index derived as described in section 7.3.5.1 is optimal.

When the objective is to maximize average profit of the progeny, rather than profit of an average progeny, the linear indexes derived above will also maximize the objective for linear and quadratic profit functions because, as shown above, $E[f(\boldsymbol{y})]$ and $f(E[\mathbf{y}])$ only differ by a constant.

Itoh and Yamada (1988) argued that most higher-order profit functions can be approximated reasonably well by a quadratic profit function. Although this may not hold for the entire range of a profit function, this will indeed most often hold for the range of EBV present in any given generation. If the quadratic approximation is accurate, we are again back to the situation of a quadratic profit function, for which a linear index derived based on maximizing profit of an average progeny is optimal.

One can also argue whether the objective should indeed be to maximize the average profit of progeny instead of maximizing the profit of an average progeny. Profit of an average progeny, $f(\bar{\boldsymbol{y}}_{t+1})$, is determined by progeny means for the genetic traits in $\boldsymbol{y}$. In general, the difference between $f(\bar{\boldsymbol{y}}_{t+1})$ and $\overline{f(\boldsymbol{y}_{t+1})}$ depends on the nature of the profit function and on the distribution of traits among the progeny. Selection has a direct effect on population means for the genetic traits, i.e. on $\bar{\boldsymbol{y}}_{t+1}$. Apart from the effects of selection on genetic variance, as described in Chapter 3, selection does not affect the distribution of traits among the progeny. The distribution of traits among progeny can, however, be affected to some degree by mating, independent of selection, by capitalizing on profit heterosis (Moav and Hill, 1966). This means that for selection purposes, the main objective should be to improve population means for biological traits and, as a result, prime emphasis should be on improving the profit of an average progeny, while the mating strategy can focus on improving distribution aspects. This is in particular true if longer-term objectives are considered, e.g. maximizing cumulative discounted profit over a planning horizon, because mating will only have a temporary effect, whereas changes in population means are permanent and passed on from generation to generation.

The previous assumes selection and mating strategies are independent. Ideally, however, selection and mating should be combined into mate-selection strategies (Kinghorn, 1997). These strategies are, however, beyond the scope of this course.

Meuwissen and Goddard (1997) compared by simulation the performance of alternative selection indexes with regard to profit obtained after 10 generations of selection. Their profit function included non-linear economic, as well as non-linear genetic relationships between traits. The latter caused genetic parameters to change as population means changed. In this case, derivation of the optimal linear index also involved updating genetic parameters. They showed that, in general, the difference between the optimal linear index and a non-linear index created by

substituting multiple-trait EBV in the profit function, i.e. $f(\boldsymbol{\mu}+\hat{\mathbf{g}})$, was small. The optimal linear index was, however, more difficult to derive and required updating of genetic parameters, whereas genetic parameters were not updated for the non-linear index. If parameters were not updated for the linear index, the non-linear index was slightly better. Thus, the non-linear index appeared quite robust to non-linearities in genetic parameters. Thus, a 'simple' non-linear index may be slightly better than a 'simple' and, therefore, sub-optimal linear index. The same held true for selection on direct EBV for profit (Meuwissen and Goddard, 1997).

7.3.8 Some Practical Considerations Regarding Non-linear Profit Functions

(Taken from Dekkers and Gibson, 1997)

In practice, nonlinear relationships are frequently implied in the selection and mating decisions that are made and promoted in the industry. In particular, computerized mating programs, in which mating pairs are arranged or a specific sire is sought for a particular cow, often rely on the concept of corrective mating, which implies existence of nonlinear relationships. Mating strategies often pay particular attention to traits that have or are perceived to have an intermediate optimum. Several conformation traits present examples of such traits, such as set of rear legs when viewed from the side (rear legs that are too curved or too straight are deemed undesirable), teat length, stature, and udder depth (shallow udders are associated with less production, but udders that are too deep are associated with more mastitis). Examples of non-conformation traits that are perceived to have an intermediate optimum are milking speed (slow milkers are associated with increased labor but fast milkers are associated with increased susceptibility to mastitis) and to some extent somatic cell count (SCC) [high SCC is associated with increased susceptibility, but SCC that is too low may also be associated with reduced resistance]. Intermediate optima can relate to traits in the breeding goal or to traits that appear exclusively in the selection index (e.g., conformation traits).

Emphasis by producers on traits with an intermediate optimum or nonlinear relationships, and emphasis on mating in breeding strategies takes away from the use and implementation of the total merit selection strategies that are promoted for genetic improvement of an overall breeding goal. Dealing with these antagonistic perspectives requires a better understanding of the nature of the nonlinear relationships considered by producers and of the role of selection versus mating in genetic improvement strategies.

7.3.8.1 Intermediate optimum traits and non-linear relationships.

A trait can be perceived to have an intermediate optimum because of simultaneous consideration of antagonistic pleiotropic effects of the trait. For traits in the breeding goal, an obvious example is milking speed, which has a positive effect on milking labour but a negative effect on susceptibility to mastitis. Udder depth is a selection index trait that is often perceived to have an intermediate optimum. This is caused by antagonistic relationships that udder depth has with two traits in the breeding goal: udder depth has an undesirable relationship with susceptibility to mastitis but a desirable relationship with production.

For traits with a real, rather than perceived, intermediate optimum, a distinction must also be made between traits in the breeding goal and selection index traits. For traits in the breeding goal, the intermediate optimum relationship between a trait and the overall goal (e.g., profit) can be formulated in terms of a non-linear profit function. Other, less extreme non-linear relationships also fall within this category. Examples of such traits in dairy cattle are conception rate, persistency, and SCC in relation to payment or penalty schemes. Non-linear relationships can also pertain to the relationship between a selection index trait and one or more traits in the breeding goal. An example is set of rear legs, which has an intermediate optimum relationship with longevity. Another example is the assertion that the need for good conformation is more important for cows at high production because better conformation enables the cow to better withstand the stresses of high production (balanced breeding).

Solkner and Furst-Waltl (1996) discussed the potential for non-linear heritabilities and of non-linear genetic correlations for functional traits in dairy cattle. Non-linear genetic relationships can be due to segregation of genes of large effect at low frequency, physiological limits, and others.

7.3.8.2 Nonlinear effects in the formulation of selection and mating strategies.

Three situations can be distinguished with regard to presence or perception of nonlinearity in relation to formulation of selection strategies: antagonistic pleitropic effects, non-linear breeding goal (non-linear profit function), and non-linear genetic parameters. Although, strategies to deal with non-linear effects in genetic improvement programs cannot ignore strategies for mating, in what follows the impact of non-linear relationships on the development and implementation of selection indexes is discussed separately.

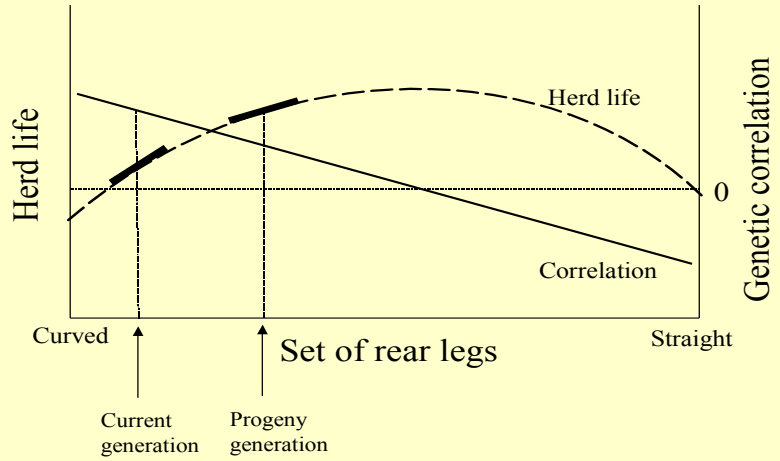
Antagonistic pleitropic effects. Antagonistic pleiotropic effects of a trait and the resulting perceived intermediate optimum, can often be resolved through proper formulation of the breeding goal or selection index and through consideration of the role of the trait in the breeding goal or index in relation to other traits that are included. For example, if resistance to mastitis is a trait in the breeding goal, along with milking speed, the economic value of milking speed should only consider the effects of milking speed on milking labour. The negative effect of milking speed on susceptibility to mastitis is accounted for in derivation of the index through the genetic correlation between milking speed and mastitis. With regard to traits in the index, pleiotropic effects of, for example, udder depth are accounted for in formulation of an index for a breeding goal that includes production and mastitis through the genetic correlations among udder depth, production, and mastitis.

Non-linear breeding goal This has been discussed previously.

Non-linear genetic parameters.
For selection index traits, non-linear relationships with traits in the breeding goal can be caused by or modelled as non-linear genetic correlations (Solker and Furst 1996). For example, the intermediate optimum relationship between set of rear legs and longevity can be modelled as a non-linear genetic correlation with a positive genetic correlation at low values of the trait (curved), a zero correlation at the intermediate optimum, and a negative genetic correlation at high values of the trait (straight) (Figure 7.7). Similarly, the assertion of the increased importance of conformation

at high production implies that the strength of the relationship between conformation and longevity increases as level of production increases. This relationship can also be modelled as a non-linear genetic correlation, which increases as level of production increases.

Figure 7.7 Relationship of set of rear legs with herd life and of the genetic correlation between set of rear legs and herd life with the population mean for set of rear legs.



Limited research has been conducted on methods for detecting and estimating non-linear genetic relationships. In addition little research has been done on the impact of non-linear relationships on selection strategies. Gowe (1983) suggested that for a trait with a non-linear heritability that is caused by presence of a major gene, selection based on an independent culling level is preferred over inclusion of the trait in a selection index. This advantage, however, was not confirmed in simulation studies by Meuwissen et al. (1995), which suggested the use of an empirical restricted selection index to deal with such traits. Strategies for dealing with this and other types of non-linear genetic parameters require further investigation. Extrapolating from results for non-linear profit functions, linear selection indexes that are derived based on linear genetic parameters evaluated at future rather than current trait means would be expected to be close to optimum for most situations. This situation is illustrated in Figure 7.7 for the non-linear relationship between set of rear legs and profit, for which the genetic correlation evaluated at the mean of progeny would be used to derive an index that maximises profit in the next generation.

For dairy cattle, traits that potentially involve nonlinear genetic relationships have limited economic importance relative to production traits. In addition, although significant nonlinear relationships (e.g., conformation traits and herd life) may be observed at the phenotypic level, as is perceived by breeders, the extent of nonlinearity may be limited at the genetic level. This limitation occurs because the range of breeding values and, especially, the range of estimated breeding values, is much smaller than the range of phenotypic values, in particular for traits with low heritability. Therefore, use of selection indexes derived based on linear genetic parameters estimated at current population means will likely be close to optimum for most applications in dairy cattle.

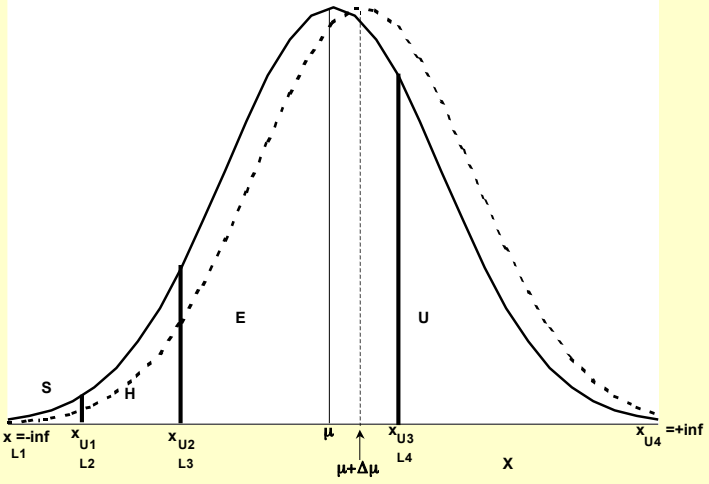
7.4 Economic Weights for Categorical Traits

7.4.1 A Graphical Illustration

Many traits are either measured on a categorical scale or, although expressed on a linear scale, incur financial rewards or penalties in some stepwise manner. In either case, the relationship between enterprise profit and individual expression of the trait is discontinuous. This does not, however, mean that profit is a discontinuous function of population mean expression for the trait. This is illustrated for the case of calving ease in dairy cattle, which is recorded on a four point scale of decreasing difficulty of calving: S (surgical intervention), H (hard pull), E (easy pull) and U (unassisted). The trait is assumed to operate as a threshold model, which assumes that there is an underlying normal distribution of susceptibility to calving ease. This distribution has a mean of zero and variance of one. Incidences of the categories define the threshold values for susceptibility, as illustrated in Figure 7.8.

An increase in the mean of the population on the underlying scale is illustrated by the broken line in Figure 7.8. The thresholds retain their absolute values but now occur further to the right relative to the mean of the new population. This leads to a decreased incidence of the deleterious calving ease categories.

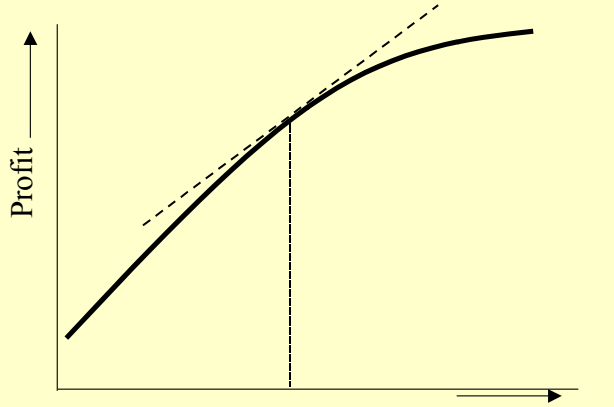
Figure 7.8: Normal distribution on underlying scale, showing thresholds corresponding to the five categories of calving ease, and reduced incidence of deleterious categories with an increase in population mean (broken line).



If the incidence of each category for a given population is p_i, i = 1, 4, and their respective economic values (i.e. profit) are w_i, then the overall profit is $P = \Sigma p_i w_i$. The change in profit with a change in the population mean, μ, can be found by changing the mean in successive small increments, recalculating incidences p_i and profit P, and then plotting P against μ. For an example situation, the resulting profit function for the population mean of calving ease is shown

in Figure 7.9. The economic weight for calving ease (expressed on the underlying scale) is the tangent to this profit curve at the current population mean. Thus at the current mean of zero, the economic weight is clearly close to zero, and is not much affected by substantial changes on either side of the current mean. But increasing calving difficulties (a decrease in calving ease on the underlying scale) leads to an accelerating increase in the economic weight.

Figure 7.9: Change in average profit with change in genetic mean for calving ease expressed on the underlying (normal) scale



The derivation of profit curves in Figure 7.9 assumes that: 1) variance on the underlying scale is not affected by the mean, and 2) profit per animal associated with each category is not dependent on its incidence. The first assumption can easily be relaxed if there is good reason for doing so, which in most cases there likely would not be. The second assumption may be false in a competitive market between breeding companies where there are acceptable limits for incidence beyond which an increasing proportion of customers would refuse to purchase a particular strain. This situation can be dealt with following the approach of de Vries (1992), after correcting for the errors in his derivations. There may also be direct associations between profit per animal in each category and overall incidence, if increasing incidences make the enterprise increasingly difficult to operate efficiently. If so, the functional relationship between incidence and v_i can be included directly in the process of constructing the profit curve.

7.4.2 An Algebraic Solution

While the graphical approach illustrated above is useful and recommended for exploring the relationship between profit and expression of a categorical trait, economic weights can also be derived directly as follows.

Consider an underlying normally distributed variable with mean μ and variance σ^2. Assume the trait is observed in n categories. Denote the lower and upper thresholds for the normally distributed underlying variable for category i by x_{Ui} and x_{Li}, the proportion in category i by p_i,

and its value by w_i. Note that $x_{L1} = -\infty$ and $x_{Un} = +\infty$. Given the mean and SD of the normal distribution and the thresholds, the proportion of the population that is in category i can be derived as: $p_i = \int_{x=x_{L_i}}^{x_{U_i}} N(x|\mu)\partial x$, which can be derived using the cumulative distribution function of the Normal distribution as $p_i = \phi(x_{Ui}) - \phi(x_{Li})$, where $\phi(x_{Ui}) = \int_{x=-\infty}^{x_{U_i}} N(x|\mu)\partial x$, which can be obtained using function NORMDIST in Excel.

Profit, P, of an average individual is then equal to: $P(\mu) = \sum_{i=1}^{n} w_i p_i$

The economic value of the trait can then be derived by evaluating the increase in profit from increasing the population mean from μ to $\mu+\Delta$ as: $v = [P(\mu+\Delta)-P(\mu)]/\Delta$.

Alternatively, assuming small genetic change, the economic weight for this trait expressed on the underlying scale can be derived as the partial derivative of the profit function evaluated at the current population mean: $v = \frac{\partial P}{\partial \mu}[\mu] = \sum_{i=1}^{n} w_i \frac{\partial p_i}{\partial \mu}[\mu]$

First partial derivatives $\frac{\partial p_i}{\partial \mu}$ can be derived as follows:

Using properties of the normal distribution,

where $N(x|\mu)$ is the normal probability density function, i.e. $N(x|\mu) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

Then, $\frac{\partial p_i}{\partial \mu}[\mu] = \frac{\partial}{\partial \mu} \int_{x=x_{L_i}}^{x_{U_i}} N(x|\mu)\partial x$

Using the rule of Leibnitz, which allows exchange of derivatives and integrals,

$$\frac{\partial p_i}{\partial \mu}[\mu] = \int_{x=x_{L_i}}^{x_{U_i}} \frac{\partial}{\partial \mu} N(x|\mu)\partial x = N(x_{L_i}|\mu) - N(x_{U_i}|\mu)$$

Thus the economic value is equal to $v = \sum_{i=1}^{n} w_i \frac{\partial p_i}{\partial \mu}[\mu] = \sum_{i=1}^{n} w_i \{N(x_{L_i}|\mu) - N(x_{U_i}|\mu)\}$

Thus, the economic value is equal to a weighted sum over categories of the difference between heights of the ordinates of the Normal distribution at the lower ($N(x_{L_i}|\mu)$) and the upper bound ($N(x_{U_i}|\mu)$) for each category.

7.5 Economic Values for Infectious Disease Resistance

Genetic improvement of resistance to infectious disease has a direct impact on the improved animal through increased health and performance and through reduced veterinary costs. Genetic improvement of these traits, however, also has an indirect impact on other animals in the herd or population through reduced infection rates because fewer animals transmit the disease, which

Direct impact on improved animal is a function of increased health and performance, reduced veterinary costs and an indirect impact on herd/population through reduced infection rates.

reduces the probability that non-resistant animals become infected. Both the direct and indirect benefits of increasing resistance must be considered when deriving economic values for these traits. Spread of the disease can be modeled through epidemiological models.

Models that combine genetics and epidemiology were developed by Bishop and Stear (1997 and 1999) to investigate the impact of increasing nematode resistance in sheep. They showed that selection on nematode egg count in faeces results in considerably greater responses in faecal egg count and live-weight gain than expected based on genetic principles when the epidemiological effects through reduced infection pressure were accounted for. These models can also be used as the basis for deriving economic values for infectious disease traits (see 7.6.1).

7.6 Economic Values for Unpriced Traits

There are several categories of traits that are associated with genetic characteristics that have no direct market value (at present). This can include traits associated with product quality (e.g. meat quality, which may at present not have a direct economic value), quality of production (i.e. traits that are valued by the producer but that have no direct economic value, e.g. temperament of dairy cows), animal welfare, environmental quality, and environmental sustainability.

One approach for such traits would be to use a desired gains index. For example, if temperament is a concern and negatively correlated with, e.g., milk yield, one may want to develop an index that keeps temperament constant. Although this appears to be an attractive alternative, it can also be very dangerous, as discussed in section 6.9. Thus, as a minimum, the impact of restricting change in temperament on change in other traits and quantifiable profit should be evaluated.

Several alternative approaches have been used in the literature to derive economic values for such traits and these will be discussed briefly below.

How do we put an economic value on these traits?

Pseudo economic value

7.6.1 Economic values for unpriced production traits

For traits that are of importance to the producer because of their inherent impact on the production process (e.g. temperament in dairy cows), one approach to derive an economic value is to evaluate the impact of the trait in relation to traits whose economic value can be quantified. For example, Wickham (1979) regressed survival on milk yield and temperament and used the ratio of the resulting estimates of the regression coefficients (b_M and b_T) to quantify the economic value of temperament: when culling cows, the producer values one unit of temperament score as high as b_T/b_M kg of milk. Thus, the economic value for temperament is b_T/b_M times the economic value for milk: $v_T = \frac{b_T}{b_M} v_M$

How producers weigh temperament vs milk yield when considering culling cow in ratio form

It is clear that survival is not the only aspect of a cow's 'socio-economic' life on the farm that temperament affects and, for that matter, neither is this the case for milk yield. The method proposed by Wickham (1979) does, however not exclude the potential impact of temperament on aspects beyond culling. However, it does assume that the culling decision provides a good

assessment of the relative importance of temperament versus milk yield. To the extent that culling provides a good assessment of a cow's value to the farmer, this assumption holds. Culling decisions, however, are or should be based on future profit that is expected from a cow (relative to a replacement), rather than profit over the entire lifetime; although past profit is a good indicator of expected future profit, other factors, such as health or fertility status are important determinants of expected future profitability. Thus, the method of Wickham (1979) will overvalue traits that affect profitability later in life, because those are important determinants of expected future profitability. In addition, because results are based on field data, the method assumes that farmer culling decisions are based on sound economic decision making.

Bishop and Nagel (unpublished, as presented in Bishop 2003), estimated a lower bound to the economic value of nematode egg count in sheep by quantifying the impact of reducing egg count on live weight, which is one of a number benefits of reducing egg counts – others are enhancing animal health and reducing anthelmintic costs. The combined genetic – epidemiological model of Bishop and Stear (1999) was used to quantify the impact of reducing egg count on growth rate in the flock.

7.6.2 Market surveys

Market surveys can be used to derive economic values for traits that are important to consumers but that (at present) do not have a direct economic value. Meat quality traits are good examples of such traits. Such surveys can be conducted at the level of the consumer, processors, or producers.

Von Rohr et al. (1996) presented a contingent valuation method (Mitchell and Carson 1993) to derive economic values for meat quality traits in pigs. This method is used to obtain estimates of costs and benefits for goods and services that are not traded on ordinary markets by presenting respondents in the survey a market in which the goods under study are treated as if they were tradable. This obviously requires respondents to be familiar with the goods and services being evaluated.

In the study of Von Rohr et al. (1996), meat technology experts of several large meat processing companies were asked to assign price changes from the base market price to a set of hypothetical carcasses with different quality characteristics (color, drip loss, intramuscular fat, iodine value, pH, and proportion of premium cuts). Six classes were set up for each quality trait and hypothetical carcasses consisted in a change from a standard carcass in only one attribute.. Economic values were then estimated using the categorical trait approach described in section 7.4.2.

Melton (1995) and Melton et al. (1996) used the experimental auction method (Shogren 1993) and the trial/repeat purchasing model to estimate economic values for pork quality traits based on consumer preferences. In this approach, consumers tasted pork with different quality attributes and compared it to chicken breast, which they tasted simultaneously. The pork was assigned one of 5 price levels, in relation to a standard price for chicken, and consumers were then asked whether they would be more likely to buy this pork, with its specific quality and price attributes,

or the chicken breast. Logistic regression was then used to analyze the probability of purchase (non-purchase) as a function of pork quality attributes and price:

$$\text{Prob(pork purchase)} = P_i = 1/(1+e^{\boldsymbol{\beta}_i'\boldsymbol{X}_i})$$

with $$\boldsymbol{\beta}_i'\boldsymbol{X}_i = \text{ fixed effects} + \Sigma b_{ij}x_{ij} + b_{price}\frac{pork\,price}{chicken\,price}$$

where x_{ij} is the value for pork quality attribute j. This model estimates the effect of a change in attribute or price on the probability of purchase. The economic value of a quality attribute was then derived by evaluating the change in price that is needed to keep the probability of purchase constant, when quality is changed by one unit. Algebraically, this can be solved from the above model as follows:

$$\boldsymbol{\beta}_i'\boldsymbol{X}_i = \ln\left[\left(\frac{1}{Prob(pork\,purchase)}\right)-1\right]$$

Thus:

$$pork\,price = \frac{chicken\,price}{b_{price}}\left\{\ln\left[\left(\frac{1}{Prob(pork\ purchase)}\right)-1\right] - fixed\ effects - \sum b_{ij}x_{ij}\right\}$$

Using a given probability of pork purchase, which was based on current consumption patterns (=0.15), and using parameters estimated from the logistic regression model (i.e. for fixed effects and b_{ij}, this equation represents a functional relationship between quality characteristics x_j and price that consumers are will to pay per kg pork. The first derivative of this equation for a given quality trait j, then gives the economic value of that trait on a per kg basis. Multiplying be the kg product per animal, and assuming no costs associated with this change in quality, gives the economic value on a per slaughter pig basis.

7.7 Incorporating competitive position in economic values

For breeders that are operating in a competitive market, market share is the driving force behind breeding objectives. De Vries (1989a) argued that in that case, economic values must take into account the competitive position of the company for individual traits. I.e. economic values should be increased for traits for which the company lags behind its competition, and economic values should be reduced for traits for which the company is ahead of the competition. De Vries (1989) used a market model to incorporate the impact of competitive position on profit for the breeder and, consequently, on economic values.

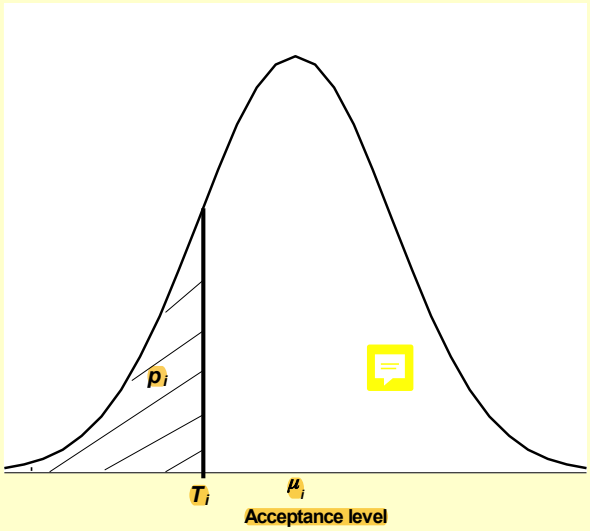
For each trait i in the breeding goal, an acceptance level T_i is defined as the minimum level for the breeding stock to be acceptable to the potential buyer. Assuming a normal distribution of acceptance level over all customers, the trait level of a given stock then determines the percentage of the customers that will find the stock unacceptable. The proportion of customers that accepts the trait level for trait i is equal to the proportion of customers whose acceptance level is below the performance level of the stock. With n traits, each with acceptance proportions p_i, market share for the stock is:

$$ms = c\,(p_1\,p_2\,p_3\,\ldots\,p_n)$$

where c is a constant that depends on the number of competitors.

For some traits such as animal welfare, and manure/methane production, it is difficult to put an objective number on; however, we should not ignore them because that would have detrimental effects as well.

Figure 7.10: Distribution of clients according to their acceptance level for a trait *i*.



For stock that is above the acceptance level for trait *i*, it is assumed that price is proportional to the regular economic value of the trait, derived at the producer level, v_i.

Then, the economic weight of the trait when taking into account saleability is equal to the regular economic value of the trait for the producer, v_i, multiplied by a factor that depends on the acceptance level for trait *i*:

$$v_i^* = \frac{z_i}{p_i}\sqrt{\frac{\pi}{2}}\ v_i = i_i\sqrt{\frac{\pi}{2}}\ v_i$$

Where p_i is the current acceptance level of the stock for trait *i*, z_i the ordinate of the standard normal distribution associated with that proportion, and i_i the selection intensity associated with proportion p_i. Note that the economic value decreases with an increase in the acceptance level for the trait.

Acceptance level for a given trait is related to the level of that trait relative to competitors. Thus, the economic value of trait *i* for a given stock depends on the genetic level of that trait in that stock relative to competitors; if the genetic level is below that of competitors, the acceptance level with be low and the economic value high; if it is above that of competitors, the acceptance level will by high and the economic value low.

Problems associated with applying this approach were discussed by De Vries (1989b) and include lack of knowledge of the buying behaviour of customers and sub-optimality of resulting indexes for longer-term responses to selection.

7.8 Some practical Considerations

(Taken from Dekkers and Gibson, 1997)

7.8.1 Identification of the Target Group and Intended Use of the Selection Index

The first step in development of breeding goals and selection indexes for practical implementation involves specifying the purpose for which the selection index will be designed. This process includes identification of its target audience, identification of financial and other incentives to which the target audience is exposed which may impact on the perceived importance of traits (Figure 6.2), and consideration of the manner in which the index is to be used. For example, although selection indexes are intended as an initial guide to selection, few breeders would base their entire selection decision on a single overall index. Prior or subsequent selection may be on individual traits, in particular, on conformation traits or individual milk component traits. Consequences of secondary selection decisions on emphasis on traits in the overall selection strategy must be monitored and perhaps incorporated when developing selection indexes.

If selection indexes are made available for cows as well as sires, it must be recognized that selection of dams of cows, which with current female reproductive rates is closely related to culling of cows, should be based on expected profit from the cow herself rather than on a genetic selection index, which is based on expected profit of the descendants. To investigate the potential consequences of implementation of the lifetime profit index (**LPI**) as criterion for genetic selection of cows in Canada, a study was undertaken (Dekkers and Gibson, 1992, unpublished) to investigate the relative efficiency of culling cows on the index of EBV for milk, fat, and protein that was incorporated in the LPI versus culling cows on an index based on estimated producing abilities for the production traits. Estimated producing ability predicts production of the cow in future lactations and is more appropriate for culling decisions than is EBV. Concern regarding misuse of the LPI for cows was exacerbated by the lack of specific guidelines for culling cows in Canada and implications of the name chosen for the genetic index (Lifetime Profit Index). Results from this study showed that culling on an index of EBV was only 4 to 7% less efficient in improving future production of current cows in the herd than culling on an index of estimated producing abilities. These results alleviated concerns regarding the consequences of potential use of the index for culling rather than genetic selection.

Consideration of the intended use of the index is also important when developing criteria for selection of sires of sons versus for selection of sires of cows. Selection of sires of sons and dams of sons requires a longer planning horizon than selection of sires of cows and selection of dams of cows. In changing markets, the same index may not be appropriate for alternative paths of selection (see subsequent discussion).

In development of breeding goals and selection indexes, a clear distinction must be made between economic traits that are included in the breeding goal and indicator traits that are included in the selection index. With regard to interpretation of the selection index, this involves clarification of the role of indicator traits in relation to the economic traits in the breeding goal. For example, a frequent assertion of breeders is the need to include conformation traits in the breeding goal. Although conformation traits can have a direct economic value for breeders who sell breeding stock, conformation only has an indirect economic value in a commercial milk production

environment through its relationship with herd life and functionality. In this case, conformation traits should not be in the breeding goal but belong in the selection index as indicator traits for components of the breeding goal.

7.8.2 Consideration of Current vs. Future Economic Circumstances and Market Demands

Development of breeding objectives and derivation of economic values must consider future conditions rather than current economic and market conditions because of the delay in the expression of selected genes. The length of the planning horizon depends on the path of selection and is different, for example, for selection of sires of sons than for selection of sires of cows.

Breeders, however, tend to judge the suitability of indexes and economic values primarily in relation to present economic circumstances, perhaps modified by their perception of future trends in consumer demands (Figure 6.2). The latter may be influenced by, for example, media reports on the need for low fat diets. The manner in which producers are paid for milk and its components provides particularly strong economic incentives. In many countries, incentives provided by the pricing system are complicated by the presence of a quota system, which is frequently based on production of one of the components (e.g. fat). This can eliminate the perceived benefits of selection pressure on that trait.

Although payment systems for milk across the world currently tend to converge toward multiple-component pricing, with payments per kg of milk, fat, and protein that are increasingly reflect world market prices, substantial differences remain. Some payment systems lack payment for protein and others base payments for fat and protein on a differential. Pricing systems for milk are typically based on past or current market considerations rather than on anticipated future market conditions. Differences in pricing systems are partly a reflection of regional differences in milk markets and partly a reflection of traditional payment schemes and their inflexibility to change. Economic incentives that are provided to producers through existing pricing systems may, therefore, not promote optimum genetic decisions. Although anticipated future market trends provide indirect incentives that can modify the impact of direct economic incentives on selection decisions (Figure 1), they are often incorporated subjectively. An index that is developed based on economic values that incorporate future market trends (e.g., J. P. Gibson, M. Greimel, and J. C. M. Dekkers, 1996, unpublished) may, therefore, not reflect producer perceptions. Such an index may be difficult to implement. For example, it may be difficult to convince a producer to select for protein yield if the pricing system reflects no payment for protein or if it reflects a protein differential rather than a payment for protein yield.

Given the impact of price incentives on breeding and management decisions, pricing and quota systems must be proactive and aimed toward the future. Ideally, pricing systems are developed in an interactive manner in close relation to anticipated changes in management and genetics that would result from the incentives they provide. This type of development may, however, be unrealistic and would be further complicated by the different planning horizons for management versus genetic decisions. More realistic is the development of pricing systems that reflect the true value of products in the market and that are flexible to accommodate changes in market values.

Breeding goals for production traits should be developed on the basis of fat and protein rather than on their percentages, because component quantities rather than their concentrations in raw milk are the marketable commodities at the level of the processing industry. Many pricing systems at the farm level, however, have traditionally been based on a price per kg of milk and a percentage differential premium based on the fat and protein content of milk. Pricing systems based on fat and protein differentials can be converted to multiple-component pricing systems based on kg of milk, fat, and protein that have an identical payout to the producer. Similarly, for fat and protein differential pricing systems, selection can be based with equal accuracy on an index of milk, fat, and protein yield as on an index that is based on milk yield and fat and protein percentage. An example is given in Table 7.7, in which the 1997 multiple components pricing system in Ontario is converted to an equivalent payment system based on differentials for fat and protein.

The main difference between the two pricing system of Table 7.7 is that the perceived value of milk yield is much higher under the differential pricing system. This is reflected in the economic values and in the resulting index weights. In fact, under the multiple-component pricing system, milk yield has a negative economic value and index weight (Table 7.7). Such an index would be difficult to implement when producers are paid based on a differential pricing system. Ideally the pricing system should be changed to reflect more closely the real economic value of milk and its components. Given the complexities of making such changes, however, the breeding goal and selection index based on milk volume and fat and protein would facilitate their implementation in such situations.

Table 7.7 Impact of alternative pricing systems, which result in identical payments to producers, on formulation of the breeding goal.

	Multiple-component pricing[1]			Percentage differential pricing		
	Milk (kg)	Fat (kg)	Protein (kg)	Milk (kg)	Fat (%)	Protein (%)
Price, $	0.071	5.31	8.44	0.53755	0.0531	0.0844
Marginal cost, $	0.152	3.11[2]	1.70	0.20928	0.0311	0.0179
Economic value, $	- 0.081	2.20	6.65	0.32827	0.0220	0.0665
Index weights[3]	- 0.067	2.19	6.19			
Standardized[4]	- 4.0	+ 4.9	+10.0			
Response[5]	+351.07	+14.10	+10.54	+351.07[6]	+0.0277[6]	+.0056[6]

[1] Based on 1996 Ontario prices and costs.
[2] Includes interest cost on fat quota.
[3] Index weights on a per kg basis for sires with 50 daughters.
[4] Index weights on a per genetic standard deviation basis.
[5] Response in daughter performance to one standard deviation selection in sires on the index.
[6] Based on assumed linear relationships between yield traits and % traits.

Concern when developing this was that producers would use it to cull cows. Culling should be much more based on production versus genetics.

7.10 Different Perspectives in Animal Breeding

In discussing profit equations, so far we have given little attention to the following four related issues that refer to definition of the breeding objective:

1) *From what perspective* should the benefits of genetic improvement be viewed. As noted in the introduction to this Chapter, we might view genetic improvement, and hence profit, from the view of the breeder/breeding company, the producer, the processor, the consumer, the whole industry, or some other view.

2) Should profit be expressed *per farm, per animal, or per unit of product*?

3) Should the breeding objective be to *maximize profit* (i.e. R-C) or to *maximize economic efficiency* (i.e. R/C), where R = total returns and C = total costs. Note that maximizing R/C is equivalent to minimizing C/R.

4) Should the breeding objective be defined *per farm, per animal, per unit of product, per unit of an input factor, or subject to any other constraint*?

It was Moav (1973) who first noted that different perspectives can yield different profit functions and different absolute and relative economic weights in the aggregate genotype. Subsequent authors have discussed this problem, and we illustrate it here with the example provided by Brascamp, Smith and Guy (1985).

Imagine a meat production enterprise consisting of N breeding females, and producing n offspring for slaughter each year. A simple profit function for the production enterprise could take the form,

$$P = N(nwr - nc_1d - c_2)$$

where w is the weight of meat produced per offspring, r is the returns per unit product, d is the number of days to slaughter, c_1 the cost per day, and c_2 the cost of maintaining each female for one year. There are three traits under genetic control, n, d and w.

Consider four different perspectives:

1) *profit per enterprise:* the viewpoint of the producer with potentially unlimited space for breeding females;
2) *profit per breeding female*: the viewpoint of the producer with a fixed number of breeding females;
3) *profit per slaughter progeny*: the viewpoint of the processor buying slaughter animals and interested in minimizing the cost per head;
4) *profit per unit of product*: the viewpoint of the consumer interested in minimum price per unit product.

The argument behind perspectives 3) and 4) is that an initial increase in profits due to genetic improvement ultimately results in reduced prices as competition forces prices down, so that something close to the original profit margin prior to genetic improvement is attained.

The appropriate profit equations are shown in Table 7.1 along with the resulting economic weights obtained as the partial derivatives of the profit equation: $v_{ij} = \frac{\partial P_j}{\partial y_i}$

where *j* indicates the perspective taken (1 to 4) and *i* indicates trait *i*.

Table 7.1 Profit equations and economic weights for four profit perspectives*.

Perspective	Profit equation	Economic Weight, v_i		
		v_n	v_d	v_w
Per enterprise	$P_1 = N(nwr - nc_1d - c_2)$	$N(\bar{w}r - c_1\bar{d})$	$-N\bar{n}c_1$	$N\bar{n}r$
Per female	$P_2 = nwr - nc_1d - c_2$	$\bar{w}r - c_1\bar{d}$	$-\bar{n}c_1$	$\bar{n}r$
Per individual	$P_3 = wr - c_1d - \frac{c_2}{n}$	$\frac{c_2}{\bar{n}^2}$	$-c_1$	R
Per unit product	$P_4 = r - \frac{c_1d}{w} - \frac{c_2}{wn}$	$\frac{c_2}{\bar{n}^2\bar{w}}$	$\frac{-c_1}{\bar{w}}$	$\frac{1}{\bar{w}^2}\left(c_1\bar{d} + \frac{c_2}{\bar{n}}\right)$

* n = number of progeny per breeding female per year; w = weight of product per slaughter animal; r = returns per unit product (price); c_1 = cost per slaughter animal per day; c_2 = cost per breeding female per year; d = days to market for slaughter animals.

It is clear from Table 7.1 that the relative economic weights for *n*, *d* and *w* are the same for perspectives 1 and 2, the absolute values differing only by a scaling factor, *N*. Thus, these two perspectives would result in equivalent genetic progress. Relative economic weights for *n*, *d*, and *w* do differ for other perspectives. This is disturbing, since it implies that different perspectives in the industry would have different indexes (and hence different directions of genetic change). But, the same animals must serve all levels of the industry.

The question then is whether it is possible to develop a consistent selection strategy (i.e. a consistent set of economic values) that meets the objective from every perspective. The answer to this question is yes, provided some important assumptions are made. Five related approaches have been suggested to obtain consistent economic values (after Goddard 1998):

1) *Zero or normal profit* (Brascamp et al. 1985): change the economic model by including normal return on investment as a cost, such that current profit equals zero.
2) *Rescaling* (Smith et al. 1986): subtract from the change in profit that results from genetic change the increase in profit that is due to a change in scale of the enterprise.
3) *Fixed base of comparison*: restrict total returns, total costs, or total profit to be constant.
4) *Define the objective as economic efficiency* (R/C) (Dickerson (1978).
5) *Scale optimization* (Amer and Fox, 1992): increase the scale of the production system until an optimum is reached for the new genetic level.

7.10.1 Zero or Normal Profit

One branch of economic theory predicts that in stable but competitive markets, profit obtained at each level of an industry settles down to the "normal profit". ***Normal profit*** is the profit necessary for persons operating a given level of the industry to make a reasonable return on their investment in time and money. In this case, normal profit can be viewed as a necessary operating cost, and would appear on the right hand side of the profit equation as a cost of production, so that profit now equals zero.

If the four profit equations in Table 7.1 are rewritten as zero profit equations, i.e. P_i is set equal to zero, a new set of economic weights can be derived and these are shown in Table 7.2. To illustrate how these economic weights are derived, consider the profit equation expressed per slaughter animal, $P_3 = wr - c_1 d - \frac{c_2}{n}$ which, from Table 7.1, gives an economic weight for n of $v_n = \frac{c_2}{\bar{n}^2}$

With zero profit: $P_3 = 0 = wr - c_1 d - \frac{c_2}{n}$

so that $c_2 = n(wr - c_1 d)$ and substituting for c_2 in the expression for v_n we get: $v_n = \frac{\bar{w}r - c_1\bar{d}}{\bar{n}}$ which is the value in Table 7.2.

Table 7.2 Profit equations and economic weights for the four profit perspectives shown in Table 7.1 when net profit is zero.

Perspective	Profit equation	Economic Weight, v_i		
		v_n	v_d	v_w
Per enterprise	$P_1 = N(nwr - nc_1 d - c_2) = 0$	$N(\bar{w}r - c_1\bar{d})$	$-N\bar{n}c_1$	$N\bar{n}r$
Per female	$P_2 = nwr - nc_1 d - c_2 = 0$	$\bar{w}r - c_1\bar{d}$	$-\bar{n}c_1$	$\bar{n}r$
Per individual	$P_3 = wr - c_1 d - \frac{c_2}{n} = 0$	$\frac{\bar{w}r - c_1\bar{d}}{\bar{n}}$	$-c_1$	r
Per unit product	$P_4 = r - \frac{c_1 d}{w} - \frac{c_2}{wn} = 0$	$\frac{\bar{w}r - c_1\bar{d}}{\bar{n}\bar{w}}$	$\frac{-c_1}{\bar{w}}$	$\frac{r}{\bar{w}}$

In the case of zero profit, relative economic weights given in Table 7.1 for n, d, and w are the same for all perspectives, with absolute values for P_1, P_3 and P_4 differing from those for P_2 by factors of N, $\frac{1}{n}$, and $\frac{1}{nw}$. Thus all perspectives would result in the same relative index weights and the same direction of genetic change.

Brascamp et al. (1985) went on to give a general proof that this was true of zero (or normal) profit equations whenever the profit function could be written in the form: $P = f(y,k_1) / g(y,k_2)$ where f is any function of genetically controlled traits, y, and economic traits, k_1, and g is any function of y and a vector of constants, k_2.

The concept of zero profit should not be interpreted as meaning that there is no incentive for genetic improvement. Under normal profit, if all producers were able to form a unified cartel, they could agree that no one should practice genetic improvement and all would retain their current profit without the expense of genetic improvement. But, in a competitive market, those producers who practice genetic improvement will increase their profits above those who do not or who do so less effectively. The incentive for change can then be viewed as either the economic advantage of practicing improvement when others do not, or the economic opportunity cost of not practicing improvement when others are (and are hence causing reduced prices as their normal profit returns to pre-improvement levels).

One problem with the zero profit approach is that incentives for appropriate change may not occur in rigidly structured industries with different sectors pursuing their own sectional interests. For example, consumers may desire lean beef, producers of slaughter calves may get paid premiums for conformation or carcass quality, but sellers of weaner calves seldom get premiums on the genetic quality of calves to yield lean beef. While this situation is maintained, breeders of weaner calves may pursue economic goals quite different from the interests of the consumer.

7.10.2 Rescaling

In this section a method of deriving economic weights proposed by Smith et al. (1986) is outlined, along with some related extensions of their proposal. The original intention of the authors was to show how, given certain assumptions, a variety of different methods and perspectives that had hitherto been seen as conflicting, were actually equivalent. (As always, it is up to you to decide whether or not you feel that the methods are valid and under what situations you might feel happy applying them.)

We have just seen that when normal profit applies, economic weights are the same from all perspectives. What is introduced here are arguments for treating all costs as variable costs and the need for enterprise rescaling due to genetic changes in production.

7.10.2.1 Fixed Versus Variable Costs

The definition of fixed and variable costs is important when deriving economic weights. If fixed costs exist, genetic increases in output can cause extra returns at the same fixed cost. Do fixed costs exist? When an enterprise is started up the answer is certainly no. The investment is geared to the level of production anticipated. Equally, when re-investment in an enterprise to change its scale occurs, that investment is geared to the level of production anticipated. Thus over long time periods, so called fixed costs are related to production (output). Similarly when summed

over many production units (the national perspective) or when investment is continuous, fixed costs are variable in relation to the level of production and size.

Another line of reasoning for considering all costs as being variable argues that, if genetic increases in output can be accommodated without a change in fixed cost, the original enterprise must not have been at maximum efficiency. Any selection, which is made to fill existing inefficiencies in production, will be of short-term value. Such selection will therefore be made at the opportunity cost of selecting for those traits that reduce costs per unit output.

7.10.2.2 Rescaling Concept

The second argument is that any profit of genetic change, which could have been made by rescaling (changing the size) of the enterprise should not be attributed to that genetic change. For example, consider a genetic change, which increases the output of lean meat from a swine enterprise. The producer might have achieved the same increase in output by increasing the size of his enterprise, probably by changing the number of swine, by 10%. The true net value of the genetic change is therefore the difference in profit due to a 10% increase in output per pig versus a 10% increase in enterprise size. The difference is the economic improvement due to reduced costs per unit output.

7.10.2.3 Derivations of Economic Weights with Rescaling to Equal Output Value

If you are unfamiliar with the relationship between partial differentials and small changes, refer to Appendix B2 before reading this section.

Consider a profit equation of the general form $P = R - C$

where P = profit, R = returns and C = costs. R and C may be any function of any number of trait values. Assume that the enterprise has a scaling factor, N, such that rescaling produces equal proportional changes in R and C. Note that this rescaling factor means that there are no fixed costs. N could be interpreted as the number of animals but does not have to be.

Given the definition of the scaling factor N: $$\frac{1}{R}\frac{\partial R}{\partial N} = \frac{1}{C}\frac{\partial C}{\partial N}$$

Consider a trait, y. Genetic change in y will lead to a change in profit of

$$\Delta P_1 = \Delta R - \Delta C$$

which, for a small change in y, Δy, gives $$\Delta P_1 = \left[\frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}\right]\Delta y$$

The change in profit from genetic improvement is the result of both a change in output (returns), $\left[\frac{\partial R}{\partial y}\right]\Delta y$, and a change in costs, $\left[-\frac{\partial C}{\partial y}\right]\Delta y$. As argued earlier, the increase in output could have been achieved without genetic improvement, by rescaling the enterprise. Let the enterprise be

rescaled by a small change in N, ΔN, to match the change in output (returns) caused by genetic change. Change in profit in this situation would be $\Delta P_2 = \left[\frac{\partial R}{\partial N} - \frac{\partial C}{\partial N} \right] \Delta N$

and the net value of genetic change is $\Delta P_3 = \Delta P_1 - \Delta P_2$

Equating the change in output from rescaling the enterprise to the change in output from genetic improvement, note that $\frac{\partial R}{\partial n} \Delta N = \frac{\partial R}{\partial y} \Delta y$

and from $\frac{1}{R} \frac{\partial R}{\partial N} = \frac{1}{C} \frac{\partial C}{\partial N}$ ➔ $\frac{\partial C}{\partial N} = \frac{C}{R} \cdot \frac{\partial R}{\partial N}$

Hence, $\Delta P_2 = \left[\frac{\partial R}{\partial y} - \frac{C}{R} \frac{\partial R}{\partial y} \right] \Delta y$

Substituting the previous equations we get, $\Delta P_3 = \left[\frac{C}{R} \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y} \right] \Delta y$

Dividing by Δy to get the economic value of unit change in y: $v_y = \frac{C}{R} \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$

Example
As an example, consider again the profit function: $P = N(nwr - nc_1d - c_2)$

In this case $R = Nnwr$

and $C = N(nc_1d + c_2)$

Note that for this profit function, all costs are variable in relation to the scaling factor N, the number of animals. Thus, at the level of number of animals, there are no fixed costs and rescaling by changing N produces equal proportional changes in returns.

The traditional economic value for n is: $\frac{\partial P}{\partial n} = N(\overline{w}r - c_1\overline{d})$

Thus, a change in n by Δn results changes profit by: $\Delta P_1 = N(\overline{w}r - c_1\overline{d})\Delta n$

and output value by: $\Delta R_1 = N\overline{w}r\Delta n$

Output value can, however also be increased by changing N; a change by ΔN results in the following change in output value: $\Delta R_2 = \Delta N\overline{n}\,\overline{w}r$
And the following change in profit: $\Delta P_2 = (\overline{n}\overline{w}r - \overline{n}c_1\overline{d} - c_2)\Delta N$

Setting ΔN to match the change in output from Δn: $\Delta R_1 = N\,\overline{w}\,r\Delta n = \Delta R_2 = \Delta N\overline{n}\,\overline{w}\,r$

Thus: $$\Delta N = \frac{N}{\overline{n}}\Delta n$$

Then, subtracting the profit that could be obtained from rescaling the enterprise to match the change in output, the net value of the genetic change by Δn is:

$$\begin{aligned}\Delta P_3 &= \Delta P_1 - \Delta P_2 \\ &= N(\overline{w}r - c_1\overline{d})\Delta n - (\overline{n}\,\overline{w}r - \overline{n}c_1\overline{d} - c_2)\Delta N\end{aligned}$$

Substituting $\Delta N = \frac{N}{\overline{n}}\Delta n$ gives: $$\begin{aligned}\Delta P_3 &= N(\overline{w}r - c_1\overline{d})\Delta n - N(\overline{w}r - c_1\overline{d} - \frac{c_2}{\overline{n}})\Delta n \\ &= N\frac{c_2}{\overline{n}}\Delta n\end{aligned}$$

Thus the economic value is: $$v_n = N\frac{c_2}{\overline{n}}$$

Rescaling against alternative methods of increased output is not the only form of rescaling that can be achieved. Other, equally plausible, possibilities are to rescale against increased input value or increased profits.

Economic values for *w* and *d* can be derived similarly, resulting in: $v_w = N\frac{\overline{n}c_1\overline{d} + c_2}{\overline{w}}$

Since a change in *d* only affects costs, the economic value for *d* is not affected by rescaling:

$$v_d = -N\overline{n}c_1$$

7.10.3 Fixed Base of Comparison (Dekkers)

As an alternative to rescaling to match the increase in output value (or input value, or profit), the enterprise might be rescaled so that total returns (output value), costs (inputs value), or profits remained constant. Economic weights can be derived for these situations by following a similar approach as outlined above. Using the example, as shown previously, for trait *n*, a change in *n* by Δn results in a change in output value by:

$$\Delta R_1 = N\,\overline{w}\,r\Delta n$$

whereas a change in *N* by ΔN results in a change in output value by:

$$\Delta R_2 = \Delta N\overline{n}\,\overline{w}\,r$$

Forcing a change in *N* such that output value remains unchanged when *n* is changed by Δn, ΔN can be solved by setting $\Delta R_2 = -\Delta R_1$:

$$\Delta N\overline{n}\,\overline{w}\,r = -N\,\overline{w}\,r\Delta n$$

Thus: $$\Delta N = -\frac{N}{\overline{n}}\Delta n$$

Changing *N* by $\Delta N = -\frac{N}{\overline{n}}\Delta n$ results in a change in profit equal to:

$$\Delta P_2 = (\overline{n}\,\overline{w}r - \overline{n}c_1\overline{d} - c_2)\Delta N$$

$$= -N(\bar{w}r - c_1\bar{d} - \frac{c_2}{\bar{n}})\Delta n$$

Thus the net economic value if n is changed by Δn is:

$$\Delta P_3 = \Delta P_1 + \Delta P_2$$

$$= N(\bar{w}r - c_1\bar{d})\Delta n - N(\bar{w}r - c_1\bar{d} - \frac{c_2}{\bar{n}})\Delta n$$

$$= N\frac{c_2}{\bar{n}}\Delta n$$

and the economic value is: $v_n = N\frac{c_2}{\bar{n}}$

Note that this is equivalent to the economic value derived previously with rescaling to match changes in output. Thus, rescaling to fixed output value is equivalent to rescaling to match changes in output value. The same holds for rescaling to fixed input value or profit.

7.10.4 Economic Efficiency

Starting in the early 70's, Dickerson argued that the only reasonable way of evaluating genetic change is by examining the effect of genetic change on the economic efficiency ratio, $\phi = R/C$, rather than on profit.

Using the previous example, economic values based on economic efficiency can be derived as follows:

$$\phi = R/C = \frac{nwr}{nc_1d + c_2}$$

$$v_n = \frac{\partial\phi}{\partial n} = \frac{\bar{w}r}{\bar{n}c_1\bar{d} + c_2} - \frac{\bar{n}\bar{w}rc_1\bar{d}}{(\bar{n}c_1\bar{d} + c_2)^2} = \frac{c_2\bar{w}r}{(\bar{n}c_1\bar{d} + c_2)^2}$$

Similarly, economic values for the other two traits can be derived to be equal to:

$$v_d = -\frac{c_1\bar{n}^2\bar{w}r}{(\bar{n}c_1\bar{d} + c_2)^2}$$

$$v_w = \frac{\bar{n}r}{\bar{n}c_1\bar{d} + c_2}$$

7.8.5 Comparision of Zero Profit, Rescaling, and Economic Efficiency

(Modified by Dekkers)

Table 7.3 summarizes economic values for the example when derived using the zero profit approach, rescaling to output value, or based on economic efficiency.

Table 7.3. Economic values for the example profit function using three alternative approaches

Perspective	Economic Weight, v_i		
	v_n	v_d	v_w
Zero profit	$N(\bar{w}\,r - c_1\bar{d}) = N\frac{c_2}{\bar{n}}$	$-N\bar{n}c_1$	$N\bar{n}r = N\frac{\bar{n}c_1\bar{d}+c_2}{\bar{w}}$
Rescaling to output value	$N\frac{c_2}{\bar{n}}$	$-N\bar{n}c_1$	$N\frac{\bar{n}c_1\bar{d}+c_2}{\bar{w}}$
Economic efficiency	$\frac{c_2\bar{w}r}{(\bar{n}c_1\bar{d}+c_2)^2}$	$-\frac{c_1\bar{n}^2\bar{w}r}{(\bar{n}c_1\bar{d}+c_2)^2}$	$\frac{\bar{n}r}{\bar{n}c_1\bar{d}+c_2}$

Although the economic values derived using the three approaches appear quite different, the relative economic values are actually equal. To see this, note that with zero profit,

$$N(\bar{n}\,\bar{w}\,r - \bar{n}c_1\bar{d} - c_2) = 0$$

and thus

$$\bar{w}\,r - c_1\bar{d} = \frac{c_2}{\bar{n}}$$

and

$$\bar{n}r = N\frac{\bar{n}c_1\bar{d}+c_2}{\bar{w}}$$

This makes the economic values for the zero profit approach equivalent to those for rescaling to output value. Also, note that economic values for rescaling to output and those based on economic efficiency differ by a factor $\frac{C}{\phi} = \frac{N^2(\bar{n}c_1\bar{d}+c_2)^2}{N\bar{n}\bar{w}r}$

Thus, economic values based on zero profit, rescaling, and economic efficiency are equivalent.

Equivalencies of economic values based on rescaling to output value with those based on economic efficiency can also be shown in more general terms as follows. Recalling from section 7.10.2.3 that economic values with rescaling to output value can be derived as:

$$v_y = \frac{C}{R}\frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$$

this can be rewritten as:

$$v_y = \frac{1}{R}\left[C\frac{\partial R}{\partial y} - R\frac{\partial C}{\partial y}\right]$$

$$= \frac{C^2}{R}\frac{\partial(R/C)}{\partial y}, \left(\text{since } \frac{d(R/C)}{dy} = \frac{1}{C}\frac{dR}{dy} - \frac{R}{C^2}\frac{dC}{dy}\right) = \frac{C}{\phi}\frac{\partial\phi}{\partial y}$$

Table 7.4 shows similar equivalencies for rescaling to match or to fixed output value, input value, profit, or zero profit. We can note that C and ϕ are the same for all traits and all situations. Thus the relative economic weights of each trait are the same for all situations. However the absolute economic weights differ, depending on the scaling factor, $\frac{C}{\phi}$, C or $\frac{C}{\phi - 1}$.

Since $\phi = \frac{R}{C}$, $\frac{\partial \phi}{\partial y}$ is the rate of change in the ratio of R:C with genetic change in y. In other words, the economic weight is always the rate of change in economic efficiency, scaled by a constant that depends on whether enterprise scaling is at the level of outputs, inputs or profit.

Table 7.4 Comparison of Economic weights derived using different approaches.[1]

Approach	Economic Weight (v_y)	
Economic efficiency	$\frac{\partial \phi}{\partial y}$	
Zero profit	$\frac{C}{\phi}\frac{\partial \phi}{\partial y}$	
	Scaled	Fixed
Output value	$\frac{C}{\phi}\frac{\partial \phi}{\partial y}$	$\frac{C}{\phi}\frac{\partial \phi}{\partial y}$
Input value	$C\frac{\partial \phi}{\partial y}$	$C\frac{\partial \phi}{\partial y}$
Profit	$\frac{C}{\phi - 1}\frac{\partial \phi}{dy}$	$\frac{C}{\phi - 1}\frac{\partial \phi}{dy}$

[1] C = cost, $\phi = \frac{R}{C}$, and R =returns.

It is also important to note that rescaling to increased or fixed profit holds for any definition of the profit equation P=R-C. Thus, these economic weights also apply to the zero profit approach of Brascamp et al. (1985). Since these authors showed that all perspectives in the market were equivalent in this situation, the results derived above should apply to all perspectives and all forms of enterprise scaling considered.

Thus, under the initial assumptions of no fixed costs and the need to disallow increased profit that could be achieved by enterprise scaling, all conflicting perspectives and derivations presented in the scientific literature are shown to be equivalent.

7.10.6 Absolute Versus Relative Economic Weights

In the absence of rescaling, economic weights would normally be taken as the partial derivative of profit with respect to the trait in question, i.e. $v_y = \frac{\partial P}{\partial y} = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$

Rescaling to output achieved by other means gives the following economic value:

$$v_y = \frac{C}{R}\frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$$

Since C is generally $<R$ (i.e. the enterprise is profitable), absolute economic weights with rescaling are less than those without. The effect on relative economic weights depends on the variation in $\frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$ between traits. In many cases, relative economic weights may be little affected. Thus, rescaling is often most important in deriving cost-benefits of animal breeding.

7.10.7 Some Problems with Rescaling

One possible criticism of the rescaling approach of Smith et al. (1986) is that when scaling to outputs or inputs, all inputs and outputs are considered only in relation to their contribution to profit at the current population mean. For example, consider the principle of scaling to fixed output in relation to the production of meat and wool from sheep. Assume that there is no limitation on inputs, i.e. that far more sheep could be reared if it were profitable to do so. As used by Smith et al. (1986), scaling to fixed output means scaling to fixed output value. It is assumed that the total output value of sheep from meat and wool is fixed. But this may not be true. It could well be that total production is limited by saturation of the market for meat or wool, but not both. Assume that the market for sheep meat is saturated so that excess production of meat has no market and makes sheep rearing beyond that point unprofitable. In that case, sheep that produced more wool at the same carcass weight would be more profitable; the economic value of wool should include all extra profits from increased output and not be scaled.

In practice it is probably very difficult to decide whether or not all traits are saturating the market. Both production systems and markets accommodate themselves to the type of animal available. Thus the method of Smith et al. (1986) achieves maximization of economic efficiency of the existing production/marketing status quo but does not consider the possibility of creating or expanding markets for some traits. Similar criticisms could be given for scaling to inputs. The current balance between breeding for current production or marketing systems and considering new balances among the traits remains to be explored.

7.10.8 Dealing With Quotas

One situation where scaling of all traits clearly runs into difficulties is when markets operate under legislated quotas on one or more but not all outputs. This could include legislated quota on production designed to manage markets, or quota on manure or mineral emissions from

production systems to limit the environmental impact of animal production (e.g. Gibson and Wilton 1998, Olesen et al. 2000).
Provided that such quota and the pricing systems that go with it have long term stability, a producer should allow for increased output opportunities for traits not under quota. The critical assumption is that the quota and pricing system will be around for a sufficiently long time. There would be little point in breeding for a quota system if the very act of producing genetic change caused modifications to the system, which partially negated those genetic changes.

Using the same notation as earlier, since total yield remains constant, for a small change, Δy, in initial output per animal y, $Ny = (N + \Delta N)(y + \Delta y)$

so that, ignoring second order terms, $\Delta N = -\frac{\Delta y}{y} N$

Enterprise profit before genetic change, is $T = NP = N(R - C)$

After a small genetic change in trait y, Δy, and scaling the enterprise so that quota is not exceeded, the new profit, T_1, is $T_1 = (N + \Delta N)\left[P + \left[\frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}\right]\Delta y\right]$

which, ignoring second order terms, gives $T_1 = T + P\Delta N + \left[\frac{\partial \mathrm{R}}{\partial \mathrm{y}} - \frac{\partial \mathrm{C}}{\partial \mathrm{y}}\right] N\Delta y$

and $v_y = \frac{\Delta T}{N\Delta y} = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y} - \frac{P}{y}$

All other traits, unconstrained by quota, can recoup the full value of increased output so that their economic weights are simply $v_y = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$

Note that we now have a situation quite different from that due to enterprise scaling relating to total inputs or outputs. Economic weights are now given by expressions of different form for traits under quota than for those not under quota. Neither equation for the economic values under quota can be written in the form $v_y = a - \frac{\partial \phi}{\partial y}$

and are thus not simply scaled measures of changes in economic efficiency.

Assuming that the initial enterprise is profitable, then $\frac{P}{y} > 0$ and the effect of scaling to quota is to reduce the economic value of the trait under quota relative to those for unconstrained traits. For highly profitable enterprises, $\frac{P}{y}$, the profit per unit output of y, can be large, so that the economic value of the trait under quota can be severely reduced, in some cases changing signs to become negative. In general, but not always, it appears that rescaling to allow for quota has a larger effect on relative economic weights than rescaling to total enterprise inputs or outputs.

It is interesting to note that, if R is a linear function of the trait under quota, then the equation for the economic value of the trait under quota becomes $v_y = \frac{\partial C}{\partial y} + \frac{C(y)}{y} - \frac{P^*}{y}$, where P^* is profit to returns and costs not dependent on y, and $\frac{C(y)}{y}$ is the average cost of production of trait y. The economic weight of trait y under quota is therefore independent of the returns generated by that trait (i.e. the price).

7.10.8.1 A Working Example

Consider a linear profit equation constructed for comparing dairy cattle genotypes,

P = 0.175 Milk Yield + 5.00 Fat Yield - 7.50 Labour - 0.1 Feed Intake - 1.0 Miscellaneous Costs

where P is expressed in $. As expressed in this problem, each trait is considered separately and has only returns or costs associated with it. Thus, milk and fat generate returns of 0.175 and 5.0 $/kg respectively, and labor, feed, and miscellaneous incur costs of 7.5 $/hr, 0.1 $/kg and 1 $/unit.

Population mean production and input levels are given in Table 7.5, along with the breeding values of two alternative sires, assumed known without error.

Table 7.5 Population means and transmitting abilities of two sires as candidates for selection.

Trait	Population Mean	Transmitting Ability Sire A	Sire B
Milk (kg)	5000	0	+1000
Fat (kg)	190	+10	+20
Labour (hr)	22	-1	+2
Feed (kg)	5800	-300	+ 700
Miscellaneous	100	0	+10

Ignoring constraints, at the population mean:

$$P = 0.175\ (5000) + 5.0\ (190) - 7.5\ (22) - 0.1\ (5800) - 1.0\ (100) = 980$$

$$R = 0.175\ (5000) + 5.0\ (190) = 1825$$

$$C = 7.5\ (22) + 0.1\ (5800) + 1.0\ (100) = 845$$

Consider economic weights scaled to output values: $v_y = \frac{C}{\phi}\frac{\partial \phi}{\partial y} = \frac{\mathrm{C}}{\mathrm{R}}\frac{\partial \mathrm{R}}{\partial \mathrm{y}} - \frac{\partial \mathrm{C}}{\partial \mathrm{y}}$

and $$\frac{C}{R} = \frac{845}{1825} = 0.463$$

Resulting economic values are in Table 7.6.

The relative net economic value of sire *A* would be:

T_A = 0.081 (00.0) + 2.315 (10) - 7.5 (-1) - 0.1 (-300) - 1.0 (0.0) = 60.65 $/daughter lactation

T_B = 0.081 (1000) + 2.315 (20) - 7.5 (-2) - 0.1 (700) - 1 (10) = 32.30 $/daughter lactation.

In this example, sire *A* is more valuable than sire *B*. For the definition of economic weights operating here, it must be assumed that the initial enterprise, geared to the population mean, operates at maximum economic efficiency. There is no quota on any one output, though overall economic output is scaled, and there are no constraints on inputs. The number of animals in the initial enterprise does not enter the equation, it is assumed to be at an optimum level.

If there is a quota on milk volume which is recognized as stable, the appropriate economic weight for milk, v_1, would be $$v_1 = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y} - \frac{P}{y}$$

while for all other traits it would be $$v_y = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$$

Resulting economic values are in Table 7.6 and using these values, net economic values of the two sires are:

T_A = 87.5 $/daughter lactation

T_B = -16.0 $/daughter lactation.

If the quota on volume operates, sire *A* is much superior to sire *B*. Again it must be remembered that for this method to be valid, it is assumed that the original production system is optimized and that inputs are not constrained.

Economic values if the quota were to apply to fat instead of volume are also given in Table 7.6., giving:

T_A = 35.92 $/daughter lactation

T_B = 76.84 $/daughter lactation.

In this case, sire *B* is considerably more valuable than sire *A*. Switching quotas from value to fat has a large effect on relative economic weights and consequent changes in sire selection.

Taking the profit equation at face value, and ignoring any constraints or need to rescale, yields economic weights $v_y = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$ (see Table 7.6)

giving $T_A = 87.5$
$T_B = 180.$

Table 7.6 Estimated economic values of traits and net economic values of sires *A* and *B* for different types of scaling, expressed as *$* per daughter lactation.

	Not constrained	Scaled Output	Scaled to Quota on Volume	Scaled to Quota on Fat
v_1	**0.175**	0.463(0.175)-0.0 = **0.0810**	0.175 - 980/5000 = **-0.021**	**0.175**
v_2	**5.0**	0.463 (5.0) -0.0 = **2.315**	**5.0**	5.0-980/190 = **-0.158**
v_3	**-7.5**	0.0 -7.5 = **-7.5**	**-7.5**	**-7.5**
v_4	**-0.1**	0.0 -0. = **-0.1**	**-0.1**	**-0.1**
v_5	**-1.0**	0.0 -1.0 = **-1.0**	**-1.0**	**-1.0**
T_A	87.50	60.65	87.5	35.92
T_B	180.0	32.3	-16.0	76.84

In this example, rescaling to allow for quotas has a very large effect on relative economic weights and absolute sire values. The effect is large because initial profit per cow, *P*, was large so that the scaling factor, profit per unit yield of the trait under quota $\left(\frac{P}{y}\right)$, is also large. This means that the potential to make improvements in economic efficiency of production of the trait under quota by genetically increasing output per cow is more than offset by losses in profit in other traits when reducing the number of cows to stay within quota. Usually, profit margins would not be so high and rescaling to quota would have a less dramatic effect.

7.10.8.2 An Example of Re-Optimization with Constraints

Consider a simple example of a dairy farm with a fixed quota, *Q*, for production of a single output trait with production level per cow of *y*, and zero payment for over quota production. The profit equation, recognizing the existence of the quota but ignoring the opportunity to optimize the system after genetic change, would be

$$P = (R(y) - C(y)|Y \leq Q) - (C(y)|Y > Q)$$

where $Y = Ny$ is total enterprise production of y and N is the number of cows. The first term in this equation is a combination of returns and costs functions that apply to under quota production, while the second term is the cost of producing y over quota (since returns are zero over quota). Initially the enterprise would be optimized so that total production exactly fills the quota, i.e. $Y = Ny = Q$. If re-optimization is ignored, the economic weight for y is found by differentiating that part of the profit equation that applies to over quota production (since all increases in output will be in excess of quota), i.e. $v_{noopt} = \frac{-\partial C}{\partial y}$, where $\frac{\partial C}{\partial y}$ is the cost of production per unit extra output.

The profit function could however be re-written to allow for optimization of the enterprise after genetic change. In this case the number of cows would be altered to stay within quota. The total enterprise profit, T, would be $T = N(R(y) - C(y))$

and, since total production $Ny = Q$, $N = \frac{Q}{y}$, and, $T = \frac{Q}{y}(R(y) - C(y)) = \frac{N_o y_o}{y}(R(y) - C(y))$

giving an economic weight for y for this optimized profit function of

$$v_{opt} = \frac{\partial f}{\partial y} = \frac{1\partial T}{N_o \partial y} = -\frac{Q}{N_o y^2}(R(y_o) - C(y_o)) + \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$$

$$v_{opt} = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y} - \frac{P}{y}$$

In this dairy cattle case with quota, v_{noopt} would be negative and v_{opt} positive and are clearly very different from each other.

The solution for v_{opt} is identical to that given for the economic weight after allowing for scaling to stay within quota. The economic value, v, without rescaling is given as: $v = \frac{\partial R}{\partial y} - \frac{\partial C}{\partial y}$, and is clearly different from v_{noopt} given here. The reason for this discrepancy is that when rescaling to quota was introduced, the initial profit equation ignored existence of the constraint. In the present example, existence of the constraint (quota) is recognized in the original profit function, but the change in management variables to optimize profit is ignored when deriving v_{noopt}

Obtaining economic weights with rescaling to constant output (or quota on a single trait) involved allowing for the change in the number of animals to stay within the constraint. It should be clear, as done here, that this change could be incorporated directly into the profit function, so that profit is now defined as profit allowing for re-optimization of management to stay within a production constraint; and differentiation of the new profit equation leads directly to v_{opt}

The importance of re-optimization of the management system should be examined on a case by case basis and will depend on the original formulation of the profit function (or model). While a

profit function ignoring re-optimization is often simpler, in general it seems safest to make sure that the profit function always allows for optimization of management to match genetic change.

As described above, quota restrictions can be incorporated into derivation of economic values through the concept of rescaling. With rescaling, economic values of the product under quota are equal to the economic value of the trait with unlimited output apart from subtracting a rescaling term. The rescaling term is equal to the average profit over fixed costs per unit of the product under quota. The same result is obtained when profit is described at the level of the whole enterprise (e.g., herd or country) instead of at the level of the individual animal. This equivalence holds provided dependence of number of cows in the enterprise on output per cow of the trait under quota is included in formulation of the profit function.

When quota is a tradable commodity, which is the case for most quota systems, the two approaches just discussed for dealing with quota may not appear sensible at the farm level because both assume an absolute restriction on output. Another approach to account for quota in the derivation of economic values is to charge interest on the market value of quota as a marginal cost for the product under quota. This more closely reflect market circumstances to which individual producers are exposed. This approach leads to economic values that are identical to economic values that are derived with rescaling when interest cost per unit of quota is equal to the average profit over fixed cost per unit of the product under quota, which is the term that is subtracted in derivation of economic values under rescaling. This condition is expected to hold when quota is traded on a free market that is in equilibrium.

and, hence, from year n_1 to n_2 is given by $\Sigma d_i = \sum_{i=n_1}^{n_2}\left(\frac{1}{1+r}\right)^i = \frac{1-\left(\frac{1}{1+r}\right)^{n_1-1}-\left(\frac{1}{1+r}\right)^{n_2}}{r}$ (8.3)

Where undiscounted returns are the same in every generation in perpetuity, the appropriate cumulate discounting factor is given by, $\sum_{i=1}^{\infty}\left(\frac{1}{1+r}\right)^i = \frac{1}{r}$ (8.4)

and if returns started in generation n and went to perpetuity, the appropriated cumulate discounting factor would be $\sum_{i=n}^{\infty}\left(\frac{1}{1+r}\right)^i = \left(\frac{1}{1+r}\right)^{n-1}\sum_{i=1}^{\infty}\left(\frac{1}{1+r}\right)^i = \frac{1}{r(1+\mathrm{r})^{n-1}}$ (8.5)

In the above dairy cattle example, the appropriate cumulative discount rate is given by (8.3) with $n_1 = 4$, $n_2 = 8$ and $r = 0.05$ to give $\Sigma d_i = 3.74$, so that $R = v\Sigma d_i = 20 * 3.74 = \$74.80$ per replacement heifer, as before.

8.1.4 Choosing Discount Rates in Animal Breeding

The first point to note is that all discount rates should be adjusted to be net of inflation since inflation affects all real values in an equal way. This can be seen by considering a kg of milk at today's price of 0.5 $/kg. If milk prices increase with inflation at 5% per annum, after 10 years of inflation, absolute prices will be $0.5*1.05^{10} = 0.814$ $/kg. If the discount rate is also 5% (equivalent to an interest rate equal to the inflation rate), the discount factor is $1/1.05^{10}$, so that the net present value of milk ten years hence is $0.5*1.05^{10}$ x $1/1.05^{10} = 0.5$ $/kg; i.e. today's price. The interest rate chosen to set the discount rate should, therefore, be based on real rates of interest, over and above the inflation rate.

From the point of view of a company setting up in the animal breeding business, the discount rate is often taken to be the real rate of return if the money were instead invested in an average business. In a review of discount rates in the animal breeding literature, Bird and Mitchell (1980) found that the minimum rate used was around 8% per annum. They argued that this was too high, since real rates of return would probably be lower than this, especially in agricultural businesses, and that it was difficult to justify a rate higher than about 5% per annum.

Bird and Mitchell (1980) also argued that government funded projects (and presumably also co-operatively run projects where members directly use results of genetic improvement in their production enterprises) should set the discount rate at the "social time preference rate". This is a rather vague measure of the extent to which people in general give preference to economic events (consumption) now rather than later. Bird and Mitchell (1980) argued that 3% per annum would be a reasonable estimate of this rate, in which case government or co-operative projects would use a lower discount rate than commercial companies.

It has been quite common in animal breeding studies to estimate cost-benefits over a fixed time horizon, say 10 or 20 years. One argument for doing this is that benefits of genetic improvement become increasingly uncertain over time. This could arise because of uncertain competition from other companies or countries in the future, uncertainty over whether the current direction of

genetic change will be appropriate under future management and market conditions and uncertainty over the amount of genetic change actually achieved. However, truncation at some particular point in time is quite arbitrary and reflects a rather peculiar form of uncertainty where returns (and cost) are obtained up until a certain point and then suddenly cease.

An alternative way of dealing with uncertainty would be to increase the discount rate used to evaluate returns, so that returns become increasingly less valuable the further into the future they occur. Since costs are often estimated with a higher degree of certainty, only the returns from the program would receive the extra discounting factor. A 2% differential would be a standard amount for uncertainty; but this may be just as arbitrary as truncating returns at a fixed point.

If these arguments are accepted, it would seem that many, perhaps most, economic evaluations of animal breeding programs have used too high a discount rate. This would favor programs with high returns early in the program rather than later, and with low initial costs and high later costs. This is probably inappropriate in a business such as animal breeding, where long-term success depends on the principle of small rates of gain building cumulatively on previous gains to give impressive long-term changes. Applying high discount rates focuses attention on traits expressed early in life, and early in the program, and away from potential long-term deleterious correlated changes. This would be a dubious practice, even for businesses where there is the potential to replace their original products with new products over time. But, in animal breeding there will be limited opportunity to replace a defective product (line or stock of animals) with another in the future. Thus, use of high discount rates could be particularly risky for long-term health of an animal-breeding program.

As a general guideline, it would seem appropriate to use discount rates in the range of 3 to 5% and to consider the use of a slightly higher discount rate for returns than for costs. But, there is considerable opportunity for more detailed definition of discount rates appropriate to animal breeding and investigation of the effects of different discount rates.

8.2. Gene Flow Methods

Gene flow methods allow study of the flow of genes through a population, which in turn can be used to define the times at which genes are expressed, and by knowing the value of that expression and the number of animals involved, the economic value of that expression can be calculated. Discounting future profits and costs then allows cost-benefit analysis of a breeding program. This section develops the principles of gene flow following the method developed by Hill (1974).

To allow easy comparison with Hill's (1974) original paper, much of the notation used here is the same as that used by Hill. This means that there is some overlap with notation used in earlier Chapters and some terms have quite different interpretations to those used earlier. Except where indicated, the examples used to illustrate development of the method are the same as those used by Hill.

8.2.1 A Diagramatic Approach to Gene Flow

At any given time, animals in a population can be divided into a number of different age and sex classes. A gene flow diagram can then be constructed to follow the movement of genes through the population over time. Imagine a population of pigs with one farrowing every 6 months. Boars are used once only and their progeny are born when they are 12 months old. Sows farrow twice so that half their progeny are born when they are 12 months old and half when 18 months old. One possible and relatively simple division of this population into age and sex classes is to consider the following five classes: 1) males at 6 months; 2) males at 12 months; 3) females at 6 months; 4) females at 12 months; and 5) females at 18 months (see section 8.2.2 for details on defining age and sex classes). Let the additive genetic merit of each of these classes when we first observe the population (at time 0) be *a*, *b*, *c*, *d* and *e*, respectively. If we examine the population six months later, animals will have aged and changed age classes, as shown by the solid arrows in Figure 8.1

For example, 6-month-old males at time 0 are 12 months old when we look at the population 6 months later. Thus, males in class (1) are now in class (2). Their genes have obviously moved with them so that the additive genetic merit of animals in class (2) changes from *b* to *a*. Similarly, class (3) animals (6-month females) move to class (4) (12 month old females), and class (4) animals move to class (5) (18 month old females). At time 6 months, classes (1) and (3) are not accounted for by aging. These animals enter the population as progeny of animals existing at time 0. The origin of genes through reproduction is shown by broken arrows for class (1) animals (6-month-old males) in Figure 8.2.

Figure 8.1 Flow of genes due to aging of initial animals.

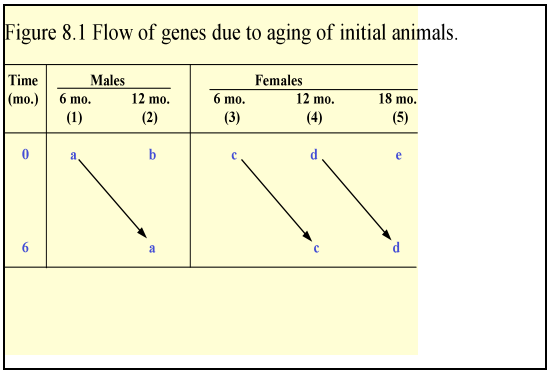


Figure 8.2 Flow of genes through reproduction to males entering population (6 months old)

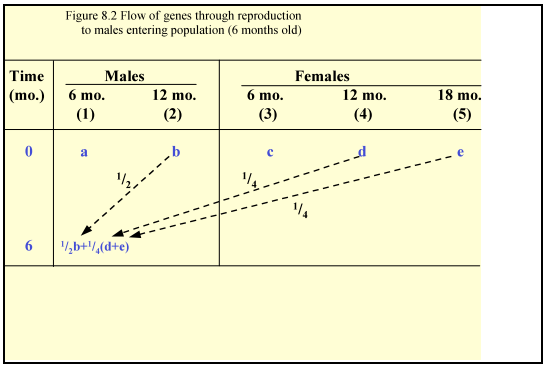
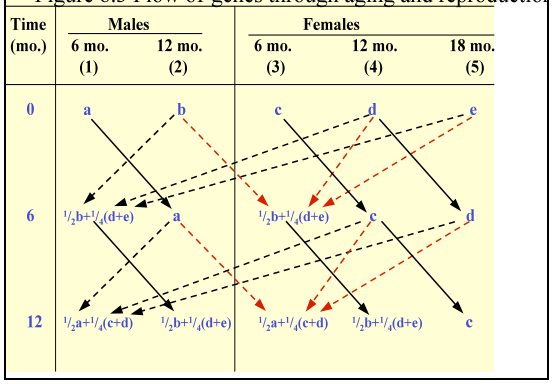


Figure 8.3 Flow of genes through aging and reproduction



Half the genes of class (1) males come from male parents that are 12 months old when their progeny are born. The other half come from female parents, half of which are 12 months and half are 18 months old when their progeny are born. Thus, half the genes of class (1) animals at time 6 months come from class (2), a quarter from class (4) and a quarter from class (5) animals at time 0. The additive genetic merit of class (1) animals is therefore $\frac{1}{2}b + \frac{1}{4}(d + e)$. Obviously the same origin of genes applies to class (3) animals (6-month-old females) whose genetic merit would also be $\frac{1}{2}b + \frac{1}{4}(d + e)$.

Clearly, the origin of genes of all animals at one time period can be accounted for by combining the movement of genes due to aging and reproduction of animals in the previous time period, as illustrated in Figure 8.3 for two time periods.

This diagramatic approach could be extended over any number of time periods but would be tedious to apply and it is easy to make mistakes, especially for the more complex population structures encountered in real life. The problem can be simplified by considering movement of genes of only one group of animals at a time and is more readily solved using the algebraic methods developed below.

8.2.2 An Algebraic Approach to Gene Flow

Typically we wish to examine the spread of genetic improvement through a population coming from a single selected group of animals. To do this, we need to estimate the proportion of genes of the original selected group of animals that are carried by animals in the population as time progresses. We need to specify: 1) the group of animals whose genes we wish to follow; 2) the structure of the population in terms of ages of each sex of animal we are interested in; 3) the frequency with which we wish to examine the population; and 4) the way in which genes are passed on from animals at one time period to animals in the following time period.

Consider the population of pigs described in section 8.2.1. We deal with the questions in the order defined above.

1. The question is how do genes from boars selected at 6 months of age spread through the population over time.

2. The age classes at which animals are examined should reflect all important events in the life of all animals of interest in the population, including as a separate age class their first appearance in the population. These age classes must be a uniform time apart, so that as time progresses in uniform steps the animals in one class can be related to animals in previous classes in the previous time period. In the present case, the population can be adequately described by considering a minimum of 3 age classes; 6, 12 and 18 months, respectively. Males first appear in the population at age 6 months and leave progeny at age 12 months. They then leave the population and do not appear at age 18 months. Females appear at age 6 months and leave progeny at ages 12 and 18 months. With males appearing at two age

classes and females at three age classes, we have a total of five age-sex classes, as defined in section 8.2.1 above.

3. The frequency with which we examine the population will depend on the ages at which events happen to animals in the population. In the present case, the time period would be 6 months since moving 6 months at a time allows animals to move from one event in their lives to the next so that all events are covered. With dairy and beef cattle a one-year interval is often convenient.

4. The object is then to define a vector $\boldsymbol{m}_{(t)}$ whose elements define the proportion of genes in each sex and age class in generation t that come from the original group of animals at time 0. In the present case, $\boldsymbol{m}_{(t)}$ is of length $h + k$, where h is the number of male classes ($h = 2$ in this example) and k the number of female classes ($k = 3$ in this example), and t is measured in 6 month periods.

In generation 0, male age classes (1 to 2) female age classes (1 to 3)

$$\boldsymbol{m}'_{(0)} = \begin{bmatrix} \overbrace{1 \quad 0}^{1 \text{ to } 2} & \overbrace{0 \quad 0 \quad 0}^{1 \text{ to } 3} \end{bmatrix} \tag{8.6}$$

Because we have defined males of age 6 months (class 1) at time $t = 0$ as the group of interest whose genes we wish to follow, that element 1 is 1 and all other elements (males at age 12 months and females at ages 6, 12 and 18 months) are 0.

The elements of $\boldsymbol{m}_{(t)}$ are found by defining the flow of genes from each sex-age class at t-1 to each sex-age class at time t. For example, males in class 2 (= age 12 months) at time t are the same as males in class 1 (= age 6 months) at time t-1 (they have aged by 6 months and clearly possess the same genes). Similar relationships exist for all other sex-age classes, except for age-class 1 of each sex (i.e. elements 1 and $h + 1$ of $\boldsymbol{m}$).

In general, therefore, the j^{th} element of $\boldsymbol{m}_{(t)}$ is given by $\boldsymbol{m}_{(t)_j} = \boldsymbol{m}_{(t-1)_{j-1}}$, for $j \neq 1$ and $j \neq h+1$.

The exceptions for the first age class of each sex are because these animals appear for the first time in the population and are new progeny of previous sex-age classes. For example, the genes of males in class 1 came half from males in class 2, one-quarter from females of class 4, and one-quarter from females of class 5 of the previous time period. Thus

$$\boldsymbol{m}_{(t)_1} = \tfrac{1}{2}\boldsymbol{m}_{(t-1)_2} + \tfrac{1}{4}\boldsymbol{m}_{(t-1)_4} + \tfrac{1}{4}\boldsymbol{m}_{(t-1)_5}$$

Similarly for the first age class of females (= element h+1=3 of $\boldsymbol{m}$), i.e.

$$\boldsymbol{m}_{(t)_3} = \tfrac{1}{2}\boldsymbol{m}_{(t-1)_2} + \tfrac{1}{4}\boldsymbol{m}_{(t-1)_4} + \tfrac{1}{4}\boldsymbol{m}_{(t-1)_5}.$$

Although it would be possible to calculate the elements of any $\boldsymbol{m}_{(t)}$ given $\boldsymbol{m}_{(t-1)}$ using these rules, it is much simpler to think of the problem in terms of a transition probability matrix which relates the proportion of genes in each sex-age class represented in $\boldsymbol{m}_{(t-1)}$ that appear in each age-sex class in $\boldsymbol{m}_{(t)}$, so that

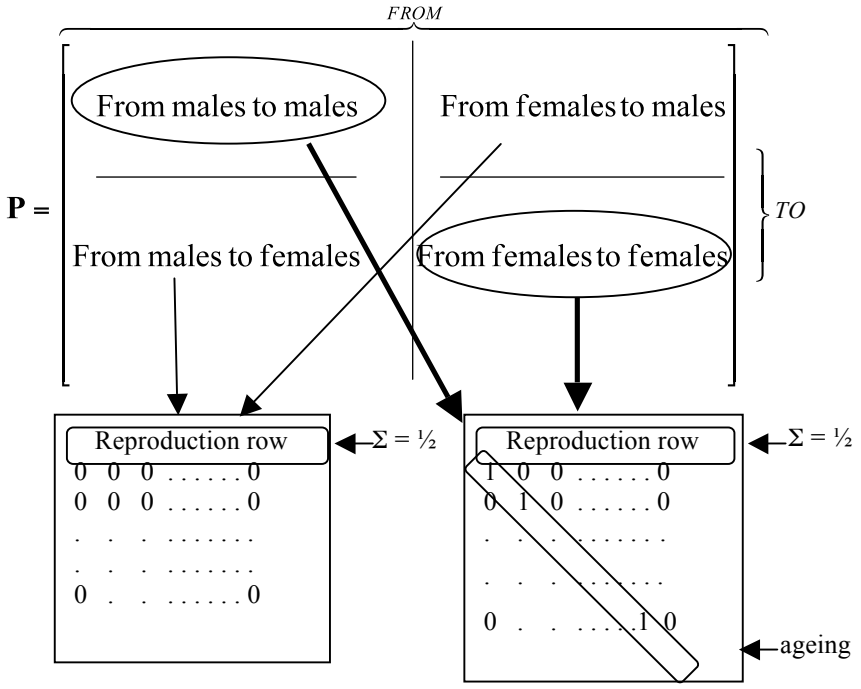
$$\boldsymbol{m}_{(t)} = \mathbf{P}\mathbf{m}_{(t-1)} \tag{8.7}$$

P then has the general form $\mathbf{P} = {}_{genes\ TO}\overbrace{\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1,h+k} \\ \vdots & & & \vdots \\ P_{h+k,1} & \cdots & \cdots & P_{h+k,h+k} \end{bmatrix}}^{\text{genes } FROM}$ (8.8)

where P_{ij} is the proportion of genes in sex-age class i at time t which comes from sex-age class j at time t-1.

In the present case, $\mathbf{P} = \left[\begin{array}{cc|ccc} 0 & 1/2 & 0 & 1/4 & 1/4 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1/2 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right]$ (8.9)

where the horizontal and vertical lines in **P** denote the separate male and female pathways of gene flow



The general form of **P** is that:

- row 1 defines the origin of genes of males entering the population
- row h + 1 defines the origin of genes of females entering the population.

All other rows define the transition by aging to current sex-age class from the previous sex-age classes. Note that the sum of all rows is 1, so that all genes in each current age-sex class are accounted for.

If the elements of $\boldsymbol{m}_{(t)}$ for each time period are required they can be found as,

$$\begin{aligned} \mathbf{m}_{(1)} &= \mathbf{P}\mathbf{m}_{(0)} \\ \mathbf{m}_{(2)} &= \mathbf{P}\mathbf{m}_{(1)} \\ &\vdots \\ \mathbf{m}_{(t)} &= \mathbf{P}\mathbf{m}_{(t\text{-}1)} \end{aligned} \qquad (8.10)$$

Alternatively, if the elements of $\mathbf{m}_{(t)}$ are required for one particular time period they can be found as:

$$\mathbf{m}_{(t)} = \mathbf{P}^{t}\mathbf{m}_{(0)} \qquad (8.11)$$

For the pig example, **P** is given by equation (8.9) and $\mathbf{m}_{(0)}$ by equation (8.6). The resulting proportion of genes in each age-sex class over successive time periods is given in Table 8.1.

Table 8.1 Proportions of genes derived from males of age 6 months in successive time periods*.

	1		2		3		4		5	
Time	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{M}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$
0	**1**	*0*	**0**	*0*	**0**	*0*	**0**	*0*	**0**	*0*
1	**0**	*0*	**1**	*0*	**0**	*0*	**0**	*0*	**0**	*0*
2	**.5**	*.5*	**0**	*0*	**.5**	*.5*	**0**	*0*	**0**	*0*
3	**0**	*0*	**.5**	*.5*	**0**	*0*	**.5**	*.5*	**0**	*0*
4	**.375**	*.375*	**0**	*0*	**.375**	*.375*	**0**	*0*	**.5**	*.5*
5	**.125**	*.125*	**.375**	*.375*	**.125**	*.125*	**.375**	*.375*	**0**	*0*
6	**.281**	*.281*	**.125**	*.125*	**.281**	*.281*	**.125**	*.125*	**.375**	*.375*
10	**.228**	*.228*	**.211**	*.211*	**.228**	*.228*	**.211**	*.211*	**.242**	*.242*
15	**.222**	*.222*	**.223**	*.223*	**.222**	*.222*	**.223**	*.223*	**.221**	*.221*
20	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*

* In bold proportions including ageing, in italic, proportions corrected for ageing

An alternative example is given in Table 8.2 which follows the flow of genes from females of age 6 months (= age class 3), so that in this case,

$$\mathbf{m}_{(0)} = [0\ 0\ 1\ 0\ 0]$$

where $\mathbf{m}_{(0)}$ is treated in exactly the same way as before, but now reflects that the original group of animals of interest were female.

Tables 8.1 and 8.2 clearly illustrate that it takes considerable time for the genes of one group of animals to spread through the population until they affect all individuals in the population equally. In general, equilibrium takes longer to be achieved the greater the number of age classes, the longer the average generation interval, the longer animals are retained for breeding and the greater the difference in breeding ages of males and females. The first factor is mostly dependent on the formulation of the problem whereas the others are inherent to a particular breeding program.

Table 8.2 Proportion of genes derived from females of age 6 months in successive time periods.

	1		2		3		4		5	
Time	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$	$\mathbf{m}_{(t)}$	$\mathbf{r}_{(t)}$
0	**0**	*0*	**0**	*0*	**1**	*0*	**0**	*0*	**0**	*0*
1	**0**	*0*	**0**	*0*	**0**	*0*	**1**	*0*	**0**	*0*
2	**.25**	*.25*	**0**	*0*	**.25**	*.25*	**0**	*0*	**1**	*0*
3	**.25**	*.25*	**.25**	*.25*	**.25**	*.25*	**.25**	*.25*	**0**	*0*
4	**.188**	*.188*	**.25**	*.25*	**.188**	*.188*	**.25**	*.25*	**.25**	*.25*
5	**.25**	*.25*	**.188**	*.188*	**.25**	*.25*	**.188**	*.188*	**.25**	*.25*
6	**.203**	*.203*	**.25**	*.25*	**.203**	*.203*	**.25**	*.25*	**.188**	*.188*
10	**.22**	*.22*	**.227**	*.227*	**.22**	*.22*	**.227**	*.227*	**.215**	*.215*
15	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*	**.223**	*.223*
20	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*	**.222**	*.222*

***** In bold proportions including ageing, in italic, proportions corrected for ageing

In the present example, the generation interval (= average age of parents when progeny are born) is two time periods (= 12 months) in the male parent path and 2.5 time periods (= 15 months) in the female path. The sum of generation intervals over both paths is $\Sigma L = 4.5$ time periods. We can note that the equilibrium genetic contribution from a single group of animals in Tables 8.1 and 8.2 is 0.222, which is equal to $1/\Sigma L$. As discussed more fully in section 8.6, this is a general result, regardless of how many paths of genetic improvement exist in the population.

Usually we are interested only in the expression of genes in their progeny and other descendants. Thus the direct contributions to $\mathbf{m}_{(t)}$ from the original group of individuals must be excluded. This can easily be accomplished by defining a new vector $\mathbf{m}^*_{(t)}$, that refers proportions of genes in each sex-age class at time t that originated from the original group of animals at time 0 *through ageing alone*. Thus,

$$\mathbf{m}^*_{(0)} = \mathbf{m}_{(0)}$$

and

$$\mathbf{m}^*_{(t)} = \mathbf{Q}\mathbf{m}^*_{(t\text{-}1)} = \mathbf{Q}^t\,\mathbf{m}_{(0)} \qquad (8.12)$$

where $\mathbf{Q}$ is a transition matrix that describes ageing alone. Matrix $\mathbf{Q}$ is analogous to $\mathbf{P}$ but describes only transmission from one sex-age class to another due to aging of animals and ignores transmission due to reproduction; i.e. $\mathbf{Q}$ is equal to $\mathbf{P}$ with elements of rows 1 and $h + 1$ set to zero. In the present pig example

$$\mathbf{Q} = \left[\begin{array}{cc|ccc} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \qquad (8.13)$$

Subtracting $\mathbf{m^*}_{(t)}$ from $\mathbf{m}_{(t)}$ results in vector $\mathbf{r}_{(t)}$, which is a vector of proportions of genes in each sex-age class at time t that originated from our group of interest in time 0 through descendants alone:

$$\mathbf{r}_{(t)} = \mathbf{m}_{(t)} - \mathbf{m^*}_{(t)} \tag{8.14}$$

Note that

$$\begin{aligned} \mathbf{r}_t &= \mathbf{P}\mathbf{m}_{(t\text{-}1)} - \mathbf{Q}\mathbf{m^*}_{(t\text{-}1)} \\ &= \mathbf{P}^t\mathbf{m}_{(0)} - \mathbf{Q}^t\mathbf{m}_{(0)} = (\mathbf{P}^t - \mathbf{Q}^t)\,\mathbf{m}_{(0)} \end{aligned} \tag{8.15}$$

Note that after a few time periods, $\mathbf{Q}^t \rightarrow 0$ so that (8.15) simplifies to $\mathbf{r}_t = \mathbf{P}^t\mathbf{m}_{(0)}$

In addition to the ageing matrix **Q**, it is also useful to define the reproduction matrix **R**:

$$\mathbf{R} = \mathbf{P} - \mathbf{Q} \tag{8.16}$$

In our example, **R** is equal to: $\mathbf{R} = \left[\begin{array}{cc|ccc} 0 & 1/2 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1/2 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

Thus, matrix **R** consists of the reproduction rows only.

Then, from equation 8.15, vector $\mathbf{r}_t$ can be rewritten as:

$$\begin{aligned} \mathbf{r}_t &= \mathbf{P}\mathbf{m}_{(t\text{-}1)} - \mathbf{Q}\mathbf{m^*}_{(t\text{-}1)} = \mathbf{P}(\mathbf{r}_{(t\text{-}1)} - \mathbf{m^*}_{(t\text{-}1)}) - \mathbf{Q}\mathbf{m^*}_{(t\text{-}1)} = \\ &= (\mathbf{P} - \mathbf{Q})\mathbf{m^*}_{(t\text{-}1)} + \mathbf{P}\mathbf{r}_{(t\text{-}1)} \\ &= \mathbf{R}\mathbf{m^*}_{(t\text{-}1)} + \mathbf{P}\mathbf{r}_{(t\text{-}1)} \\ &= \mathbf{R}\mathbf{Q}^{t\text{-}1}\mathbf{m}_{(0)} + \mathbf{P}\mathbf{r}_{(t\text{-}1)} \end{aligned} \tag{8.17}$$

Here, the first term represents production of progeny from the initial group of individuals, while the second term represents ageing of and reproduction from their descendants.

8.2.3. Defining the Flow of Genetic Improvement from One Round of Selection

Vector $\mathbf{m}_{(t)}$ defines the proportions of genes coming from the original group of animals which, assuming additive genetic inheritance, is also the proportion of the breeding value of the original group of animals that is expected to be expressed in each sex-age class at time t. Thus, if the vector of genetic superiorities (i.e. mean breeding values relative to unselected animals in the same age-sex class) at time t is **s**, then

$$\mathbf{m}_{(0)} = \mathbf{s}$$

and, from equations 8.10 and 8.11, the vector of resulting responses at time t can be computed as: $$\mathbf{m}_{(t)} = \mathbf{P}\mathbf{m}_{(t\text{-}1)} = \mathbf{P}^{t}\mathbf{m}_{(0)}$$

Similar to before, vector $\mathbf{m}_{(t)}$ includes the occurrence of the original group of animals. Usually we are interested only in expression of genes in their progeny and other descendants. This can easily be accomplished by defining a new vector of responses, $\mathbf{m}^*_{(t)}$, which refers only to genetic expression of the original animals (parents) and ignores all descendants. Similar to before, $\mathbf{m}^*_{(t)}$ can be computed as $\mathbf{m}^*_{(t)} = \mathbf{Q}\mathbf{m}^*_{(t\text{-}1)} = \mathbf{Q}^{t}\mathbf{s}$, which is then subtracted from the original response vector, to obtain a vector $\mathbf{r}_{(t)}$ which defines the vector of responses at time t from selection decisions made at time 0 through reproduction alone. Similar to equation 8.17:

$$\mathbf{r}_{(t)} = \mathbf{R}\mathbf{m}^*_{(t\text{-}1)} + \mathbf{P}\mathbf{r}_{(t\text{-}1)} = \mathbf{R}\mathbf{Q}^{t\text{-}1}\mathbf{s} + \mathbf{P}\mathbf{r}_{(t\text{-}1)} \qquad (8.18)$$

Similar to equation 8.17, the first term represents transmission of genetic superiority from the originally selected individuals to their progeny, while the second term represents transmission of resulting response over age classes and generations through ageing of and reproduction from their descendants.

Again following the pig example given by Hill (1974), imagine that selection is for weight gain, which has $h^2 = 0.3$ and $\sigma_p = 70$ g/day. Assume that the best 1/40 males and the best 1/8 females are selected prior to 6 months of age. Then, genetic superiorities of selected 6-month-old males is 50 g/day ($ih\sigma_g$) and the genetic superiority of selected 6-month-old females is 35 g/day. Then, the vector of genetic superiorities at time 0 is:

$$\mathbf{s}' = [50 \;\; 0 \;\; 35 \;\; 0 \;\; 0]$$

If we are interested in following response from selection of males only, we can define $\mathbf{s}$ as:

$$\mathbf{s}' = [50 \;\; 0 \;\; 0 \;\; 0 \;\; 0]$$

Similarly, if we want to follow selection of females only, $\mathbf{s}' = [0 \;\; 0 \;\; 35 \;\; 0 \;\; 0]$

Table 8.3 shows resulting responses of live weight gain due to selection of this single group of males or females or with both sexes selected. Response is evaluated as the mean genetic merit of males and females entering the population in age class 1 in each time period (i.e. mean of elements 1 and 3 of $\mathbf{r}_{(t)}$). In the present example, genetic merit of males in age class 1 is the same as that of females in age class 1 at every time period, as can be seen from Tables 8.1 and 8.2. This is because parental origin of male genes is identical to that of female genes, as defined by rows 1 and $h + 1 = 3$ of $\mathbf{P}$. Although this need not be true in general, in many cases it will be.

Note that, similar to genetic contributions in Tables 8.1 and 8.2, responses to selection initially fluctuate but then stabilize. The eventual genetic contribution of 18.9 is equal to the asymptotic gain per time period for this breeding scheme, which can be computed as: $R = \frac{50 + 35}{2 + 2.5} = 18.9$ kg/six-month period. We will come back to this in the next section.

Table 8.3 Response of live weight gain when a single group of males and/or females are initially selected.*

Time	Response		
	Males Initially Selected	Females Initially Selected	Both Sexes Initially Selected
0	0	0	0
1	0	0	0
2	25.0	8.8	33.8
3	0	8.8	8.8
4	18.8	6.6	25.3
5	6.2	8.8	15.0
6	14.1	7.1	21.2
10	11.4	7.7	19.1
15	11.1	7.8	18.9
20	11.1	7.8	18.9

* Response is the mean genetic merit of males and females entering the population (age class 1) in each time period.

8.2.4 Defining the Flow of Improvement from Multiple Rounds of Selection

In most breeding programs, the intention would be to practice genetic selection uniformly for each successive group of animals recruited to the population in successive time periods. In an additive genetic system, the expected additive genetic merit of an animal from any one source can be added to the additive genetic merit obtained from all other sources to obtain the overall genetic merit of that animal, provided that all initial sources are expressed as deviations from the mean of that group prior to selection. Thus, the cumulate genetic merit of a given sex-age class at a given time due to selection in several time periods can be found by simply adding together the predicted merit due to each round of selection separately.

Thus if selection is practiced in each of n successive time periods starting at time period 0, the cumulate response vector, $\mathbf{R}_{(t)}$ is given by

$$\mathbf{R}_{(t)} = \mathbf{r}_{(t)} + \mathbf{r}_{(t-1)} \ldots + \mathbf{r}_{(t-n+1)} \qquad (8.19)$$

or, if selection takes place in every time period $\mathbf{R}_{(t)} = \mathbf{r}_{(t)} + \mathbf{r}_{(t-1)} + \ldots \mathbf{r}_{(1)}$ (8.20)

and selection taking place in any combination of time periods can be obtained by analogy.

Cumulate responses for the pig example assuming selection takes place in every time period are shown in Table 8.4. Extra response (genetic gain) from one time period to the next when both males and females are selected is also given in column 4 of Table 8.4. This genetic gain is equal to response in each time period due to a single round of selection at time $t = 0$ (column 3 of Table 8.3). Thus, the rate of approach to equilibrium response rate with continuous selection is

the same as the rate of approach to equal gene distribution (or equal genetic effect) in all groups of animals with a single round of selection.

Stabilized gain from one time period to the next is again equal to the asymptotic response to selection, i.e. 18.9. Here, response is the result of selection in all previous rounds of selection. In contrast, in Table 8.3, asymptotic response is the eventual effect of a single round of selection.

Table 8.4 Cumulate response[1] and Δg[2] with selection for live weight gain in every time period.

Time	Cumulative Response			Δg with Males and Females Selected
	Males Only Selected	Females Only Selected	Males and Females Selected	
1	0	0	0	0
2	25.0	8.8	33.8	33.8
3	25.0	17.5	42.5	8.8
4	43.8	24.1	67.8	25.3
5	50.0	32.8	82.8	15.0
6	64.1	39.9	104.0	21.2
10	107.5	71.3	178.8	19.1
15	163.0	110.2	273.1	18.9
20	218.5	149.1	367.6	18.9

[1] Response measured as the genetic merit of males and females recruited to the population (age-sex classes 1 and 3) in each time period.
[2] Δg is the genetic change from the previous time period to the current.

8.3 Discounted Gene Expression

Responses derived in previous sections can be converted to *discounted economic expressions* given the economic value of the trait, the discount rate, and the number of animals in each age-sex class that generate income. The discount rate in any given time period, t, is given by

$$d_{(t)} = \left(\frac{1}{1+\mathrm{r}}\right)^{t/m} \tag{8.21}$$

where r is the discount rate per annum and m is the number of time periods per year.

Let **n** be a vector with numbers of animals by sex-age group that express the trait in each given time. For the pig example, imagine that there are 100 sows breeding every time period (50 at age 12 months and 50 at 18 months) and each sow produces 4 male and 4 female progeny per litter. There are 400 males and 400 females entering the population each period. One in 40 males are selected for breeding, leaving $\frac{39}{40}$ x 400 = 390 males for slaughter at 6 months (age-sex class 1). The 10 males returned for breeding are slaughtered at 12 months (age-sex class 2). One in 8

females are returned for breeding, leaving $\frac{7}{8}$ x 400 = 350 females for slaughter. Thus, vector **n** is equal to: $\mathbf{n}' = [390 \;\; 10 \;\; 350 \;\; 0 \;\; 0]$

Similarly, let **v** be a vector with economic values per unit of genetic improvement for the trait in and per animal that expresses the trait in each sex-age group. These economic values can be derived as described in Chapter 7. For the sake of argument, the economic value of live weight gain is set at 0.01 $/g/day for slaughter males at both ages and for slaughter females at 6 months, but is zero for slaughter females at 18 months because of increased feed costs while maintaining the sow for breeding. Thus, vector v becomes:

$$\mathbf{v}' = [0.01 \;\; 0.01 \;\; 0.01 \;\; 0 \;\; 0]$$

Then, compute a vector $\mathbf{w} = \mathbf{n} \# \mathbf{v}$ (8.22)

where the operation # indicates element-wise multiplication. Vector **w** is a vector of economic benefits of a one unit genetic improvement of the trait for each sex-age group, across all animals that express the trait by sex-age group. In our example:

$$\mathbf{w}' = [390 \;\; 10 \;\; 350 \;\; 0 \;\; 0] \# [0.01 \;\; 0.01 \;\; 0.01 \;\; 0 \;\; 0] = [3.9 \;\; 0.1 \;\; 3.5 \;\; 0 \;\; 0]$$

The first element in this vector implies that a one unit genetic improvement in age class 1 of males results in $3.9 greater profit from this age group in a given time period.

The problem as defined equates to a commercial farmer practicing genetic selection within his own herd and not selling any breeding stock. (Note that this economic evaluation is not the same as that used by Hill (1974), in part to better illustrate the problem and in part to correct some minor inconsistencies in Hill's paper.)

With a single round of selection, returns at time *t* of genetic superiority created at time 0 is given by: $y^*_{(t)} = \mathbf{w}'\mathbf{r}_{(t)}$ (8.23)

With a discount rate for time *t* of $d_{(t)}$, the present value of these returns is equal to:

$$y_{(t)} = d_{(t)}\mathbf{w}'\mathbf{r}_{(t)} \quad (8.24)$$

The present value of cumulative returns at time t from one round of selection at time 0 is equal to: $Y_{(t)} = \sum_{i=1}^{t} y_{(i)}$ (8.25)

Discounted returns at time *t* if continuous selection is practiced are equal to:

$$y^c_{(t)} = d_{(t)}\mathbf{w}'\mathbf{R}_{(t)}. \quad (8.26)$$

and present value of cumulate returns from continuous selection at time *t* are given by:

$$Y^c_{(t)} = \sum_{i=1}^{t} y^c_{(t)} \quad (8.27)$$

For the present pig example, discounted returns at each time period and cumulative discounted returns over time are shown for a single round of selection and for continuous selection of both males and females in Table 8.5.

Table 8.5 Discounted returns and cumulative discounted returns at time *t* for a single round of selection and continuous selection of males and females for live weight gain in a pig nucleus of 100 sows, with a discount rate of 0.05.

		Discounted Returns ($)			
		Single Round of Selection		Continuous Selection	
Time	Discounting Factor (*d*)	At Time *t*	Cumulate to *t*	At Time *t*	Cumulate to *t*
1	.976	0	0	0	0
2	.952	238	238	238	238
3	.929	63	301	295	533
4	.907	171	472	459	992
5	.885	100	572	548	1540
6	.864	137	709	671	2213
10	.784	112	1165	1049	5858
15	.694	98	1680	1419	12258
20	.614	87	2138	1691	20207
25	.543	77	2542	1882	29265
30	.481	68	2900	2007	39073
40	.377	53	3497	2106	59828
50	.295	42	3965	2069	80770

Results in Table 8.5 illustrate the process whereby, as the discounting factor becomes smaller over time, returns per annum from a single round of selection continue to decrease even after genetic equilibrium has been achieved. Cumulate returns then increase at a diminishing rate as time progresses and would eventually plateau.

With continuous selection, discounted returns at time *t* increase over time at a diminishing rate and eventually reach a peak and then begin to decrease, though more slowly than returns from a single round of selection. In the present example, maximum discounted returns with continuous selection occurred at period 42 (= year 21). This is a reflection of the role of discounting in saying that future events (costs or returns) are inherently of less interest to us than current events. It should not be taken as meaning that animal breeding will be inherently less valuable 40 years from now than it is today. This being so, it raises the question of whether it is sensible to consider continuous selection schemes in terms of continuous investment appraisal or whether one should consider only returns from investment made over a finite time period.

8.4 Cost-Benefits and Cash Flow

The discounted returns illustrated in Table 8.5 can readily be used to estimate the cost-benefits of a breeding program over time. Often the costs (at today's prices) will be expected to be the same

from one period to the next but this need not be so. If we identify the cost of running the improvement program in period t at today's $(t = 0)$ prices as $c_{(t)}$ then the discounted cost in year t is:

$$C_{(t)} = \sum_{i=1}^{t} d_{(i)}c_{(i)} + c_{(0)} \qquad (8.28)$$

It is then straightforward to examine the cumulate discounted net returns, NR, over time t_1 from investment over time t_2 as:

$$NR_{t_1,t_2} = Y^{c}_{(t_1)} - C_{(t_2)} \qquad (8.29)$$

Also of interest is the change in NR over time and, in particular, determination of the point in time when the investment becomes profitable and estimation of the rate of increase in profit beyond that point. This is illustrated by considering the pig example. Imagine that the cost of selection in each time period was \$572, due to the costs of identifying individual pigs, weighing and recording, selection and so on, so that,

$$c_{(0)} = c_{(1)} = c_{(2)} \ldots = c_{(t)} = \$572$$

It is clear from column 2 of Table 8.5 that a single round of selection at time 0 would yield discounted cumulate returns of \$572 at period 5 (= 2½ years later). Thus the initial investment in selection is recouped after 2½ years, and yields a net discounted cumulate profit of $Y_{(t)}$ - 572 beyond that time. For example, after 10 time periods (= 5 years) the net cumulate discounted profit is 1165 - 572 = \$593.

Often, net profit is expressed in ratio form as the cost-benefit ratio, which is the ratio of cumulate discounted returns to cumulate discounted costs:

$$RR_{t_1,t_2} = Y^{c}_{(t_1)} / C_{(t_2)} \qquad (8.30)$$

In the present case, with a single round of selection, the cost-benefit ratio at time period 10 would be 1165/572 = 2.04.

It takes longer to obtain a net profit with continuous selection schemes than with a single round of selection. This is illustrated in Table 8.6 where cumulate returns from continuous selection, taken from Table 8.5, are combined with cumulate investment costs to obtain the cost-benefit ratio for this scheme and this is compared to the cost-benefit ratio for a single round of selection. In this case, with continuous selection, it takes 10 periods (5 yrs) before the cost-benefit ratio exceeds 1.0 and the program becomes profitable. The cost-benefit ratio is always lower than that for a single round of selection, though the two would have equal cost-benefit ratios at $t = \infty$. Again, these results raise a note of caution in using discounted cost-benefit measures for long periods of investment, since later investments are not given the opportunity to recoup returns when the returns' time horizon is truncated at the same point as the investment (cost) horizon.

Table 8.6 Discounted costs, returns and cost-benefit ratio for continuous selection and cost-benefit ratio for single round of selection of both sexes for live weight gain, when the cost of selection is $572 per time period.

Time	Single Round of Selection Cost-Benefit Ratio	Continuous Selection	
		Cumulate Discounted Costs	Cost-Benefit[1] Ratio
1	0	1130	0
2	.41	1674	.14
3	.53	2206	.24
4	.82	2725	.36
5	1.0	3231	.48
6	1.24	3726	.59
10	2.04	5586	1.05
15	2.94	7670	1.60
20	3.74	9515	2.12
25	4.44	11147	2.62
30	5.07	12593	3.10
40	6.11	15004	3.97
50	6.93	16894	4.78

[1] Using discounted returns from Table 8.5.

8.5 Expansion to More Complex Breeding Structures

8.5.1 Different Selection Intensities in Parents of Males and Females

In section 8.2.2 it was implicitly assumed that genetic superiority of parents (defined by **s**) was the same for both sexes of replacement breeders. In most situations this is true, but typically in dairy cattle it does not hold since male and female parents of breeding males are more intensely selected than parents of breeding females. This can be easily be accommodated by separately defining parental selection vectors, $\mathbf{s}_m$ and $\mathbf{s}_f$, and creating two reproduction matrixes, $\mathbf{R}_m$ and $\mathbf{R}_f$, which define the passage of genes to males and females by reproduction only. In our pig example, $\mathbf{R}_m = \begin{bmatrix} 0 & 1/2 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ and $\mathbf{R}_f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

so that $\mathbf{R}_m$ and $\mathbf{R}_f$ correspond to the rows 1 and h+1(=3) of **P**, with all other rows set to zero, or to rows 1 and h+1 of the full reproduction matrix **R**. Note that $\mathbf{R}_m + \mathbf{R}_f = \mathbf{R}$.

Then, using equation 8.18, response at time t from parents of males selected at time 0 can be computed as:

$$\mathbf{r}_{m(t)} = \mathbf{R}_m\mathbf{m}^*_{(t-1)} + \mathbf{P}\mathbf{r}_{m(t-1)}$$

$$= \mathbf{R}_m\mathbf{Q}^{t-1}\mathbf{s} + \mathbf{P}\mathbf{r}_{m(t-1)} \qquad (8.31)$$

The first term defines response obtained from transmission of genetic superiority created at time 0 through the parents of males selection path and the second term represents subsequent responses through ageing and reproduction of progeny.

Similarly, responses through selection of parents of females at time 0 can be computed as:

$$\mathbf{r}_{f(t)} = \mathbf{R}_f\mathbf{Q}^{t-1}\mathbf{s} + \mathbf{P}\mathbf{r}_{f(t-1)} \qquad (8.32)$$

Note that combined response from selection of parents of males and females is equal to:

$$\mathbf{r}_{(t)} = \mathbf{r}_{m(t)} + \mathbf{r}_{f(t)} = \mathbf{R}_m\mathbf{Q}^{t-1}\mathbf{s} + \mathbf{P}\mathbf{r}_{m(t-1)} + \mathbf{R}_f\mathbf{Q}^{t-1}\mathbf{s} + \mathbf{P}\mathbf{r}_{f(t-1)}$$

$$= (\mathbf{R}_m + \mathbf{R}_f)\mathbf{Q}^{t-1}\mathbf{s} + \mathbf{P}(\mathbf{r}_{m(t-1)} + \mathbf{r}_{f(t-1)})$$

$$= \mathbf{R}\mathbf{Q}^{t-1}\mathbf{s} + \mathbf{P}\mathbf{r}_{(t-1)}$$

which is identical to equation 8.18.

Responses can be further split into the four individual paths of selection, sires of males (*sm*), dams of males (*dm*), sires of females (*sf*), and dams of females (*df*), by defining separate reproduction matrices by path, $\mathbf{R}_{sm}$, $\mathbf{R}_{dm}$, $\mathbf{R}_{sf}$, $\mathbf{R}_{df}$. For example, for our pig breeding program,

$$\mathbf{R}_{df} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Response through this pathway can then be evaluated as: $\mathbf{r}_{df(t)} = \mathbf{R}_{df}\mathbf{Q}^{t-1}\mathbf{s} + \mathbf{P}\mathbf{r}_{df(t-1)}$ (8.33)

The above equations can also be used to evaluate the number of discounted expressions by pathway by replacing vector **s** with vector $\mathbf{m}_{(0)}$ that has element 1 at the appropriate place. For example, the number of discounted expressions from selection of dams of females in our pig example can be evaluated by setting

$$\mathbf{m}_{(0)} = [0\ 0\ 1\ 0\ 0]$$

and computing

$$\mathbf{r}_{df(t)} = \mathbf{R}_{df}\mathbf{Q}^{t-1}\mathbf{m}_{(0)} + \mathbf{P}\mathbf{r}_{df(t-1)} \qquad (8.34)$$

Resulting vectors $\mathbf{r}_{df(t)}$ can then be used to compute discounted expressions using the methods described in section 8.3.

Differential reproduction by pathway can result in different numbers of discounted expression by pathway and, therefore, in different discounted economic values. In most cases, these differences are, however, expected to be small.

8.5.2 Multiplier Tiers and Commercial Herds

The examples so far have been for selection in a single nucleus population or herd. In practice there may be several levels of multipliers before genetic improvement reaches commercial herds. This is easily accommodated since **P** can be defined to accommodate flow of genes between any number of groups of animals rather than just the two groups (males and females in the nucleus) so far considered.

The general structure of **P** is:

$$\mathbf{P} = \begin{matrix} \overbrace{\begin{bmatrix} group\,1 & group\,2 & . & \dots & group\,n \end{bmatrix}}^{\text{genes } FROM} \\ \left.\begin{matrix} group\,1 \\ group\,2 \\ \\ group\,n \end{matrix}\right\} \text{Genes } TO \end{matrix}$$

and each element of **P**, P_{ij}, describes the proportion of genes appearing in group-age class i at time t that originate from group-age class j at time t-1.

Vectors $\mathbf{m}_{(0)}$ (for discounted expression analysis) and **s** (for response analysis) are similarly augmented to match the dimensions of **P**, and matrices **Q** and **R** contains those elements of **P** describing the aging and reproduction of animals, respectively, with all other elements set to 0.

Consider a slightly modified version of Hill's (1974) extension to the pig example, whereby breeding males from the nucleus become parents of commercial animals at 18 months old and replacement females in the commercial herd come from the commercial herd, in which females have litters at 12, 18 and 24 months. Allowing for culling and other losses, females in commercial herds leave 1/2, 1/3 and 1/6 of their progeny at parities 1, 2 and 3 (ages 12, 18 and 24 months). All commercial males are slaughtered at 6 months. The **P** matrix then becomes:

$$\mathbf{P} = \begin{matrix} \\ nm \\ \\ \\ nf \\ \\ cm \\ \\ cf \\ \\ \\ \end{matrix} \overbrace{\left[\begin{array}{ccc|ccc|c|cccc} 0 & 1/2 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1/2 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/6 & 1/12 \\ \hline 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/6 & 1/12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]}^{\text{Genes } FROM} \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} \text{Genes } TO$$

$$\begin{matrix} & nm & & & nf & & cm & & cf & \end{matrix}$$

where *nm, nf, cm*, and *cf* indicate nucleus males, nucleus females, commercial males and commercial females.

There are now three age classes for nucleus males because they are kept till 18 months to sire commercial progeny. There is one age class of commercial males, the age at which they appear and are sent to market (6 months). Commercial males pass on no genes and appear here for the purposes of estimating economic returns and gene flow to commercial product. There are four age classes of commercial dams, representing their first appearance in the population and at three successive breedings.

Matrix **Q** is obtained by setting the elements of rows 1, 4, 7, 8, 9, 10, 11 of **P** to zero. And when both sexes are initially selected, **s** becomes

$$\mathbf{s}' = [50\ \ 0\ \ 0\ \ 35\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0] .$$

Table 8.7 gives the average genetic merit of animals entering the commercial herd with one round of selection and with continuous selection on both sexes in the nucleus.

Table 8.7 Genetic responses of commercial animals and cumulate discounted returns of commercial plus nucleus animals with selection on males and females for live weight gain.

	Responses (g/day)		Cumulate Discounted Returns ($)	
Time	1 Round of Selection	Continuous Selection	1 Round of Selection	Continuous Selection
1	0	0	0	0
2	0	0	238	238
3	25	25	647	879
4	0	25	816	1674
5	23.1	48.1	1222	2855
6	8.5	56.7	1468	4254
10	16.3	126.0	2764	12756
15	18.6	217.5	4273	28907
20	18.8	310.9	5628	49940
25	18.9	405.1	6833	74576

We will assume essentially the same economic conditions as before, but now with 200 breeding sows in the commercial herd, 1/8 sows in the commercial herd are kept for replacements. (There are 200 x 4 = 800 females born per time period, and ½ x 200 = 100 of breeding sows are in the youngest category at 18 months of age. Of the 800 females born, we therefore retain 100 for breeding; i.e. $\frac{100}{800} = \frac{1}{8}$ of females born.) Thus, there are 800 males and $\frac{7}{8}$ x 800 = 700 females in age class 1 in the commercial herd going to slaughter. Since breeding males in the nucleus are now kept until 18 months old, we will assume that the economic advantage in their live weight gain is zero. This gives

$$\mathbf{n}' = [390\ \ 0\ \ 0\ \ 350\ \ 0\ \ 0\ \ 800\ \ 700\ \ 0\ \ 0\ \ 0]$$

and

$$\mathbf{v}' = [0.01\ \ 0\ \ 0\ \ 0.01\ \ 0\ \ 0\ \ 0.01\ \ 0.01\ \ 0\ \ 0\ \ 0]$$

so that $\mathbf{w}' = \mathbf{n}' \# \mathbf{v}' = [3.9 \;\; 0 \;\; 0 \;\; 3.5 \;\; 0 \;\; 0 \;\; 8 \;\; 7 \;\; 0 \;\; 0 \;\; 0]$

The resulting discounted economic returns from both the nucleus and the commercial herd are also shown in Table 8.7.

8.5.3. Crossbreeding and Other Structures

It should be clear from the general structure of **P** given in section 8.5.2 that virtually any breeding structure can be accommodated by gene flow methods, provided the origin of genes can be defined for each class of animal in the population. Crossbreeding structures consist of two or more nucleus populations, with or without multipliers feeding into a commercial population. The purebred nucleus can also be open to genes coming back in from selected males or females from the purebred multipliers and commercial populations.

As an example, in the population examined in section 8.5.2, a proportion of the commercial sows could be selected for litter size in their first two parities and returned to the nucleus for one parity of breeding. If half the 18-month nucleus sows were replaced by 24-month commercial sows, the first and fourth rows of **P** would become

$$[0 \quad 1/2 \quad 0 \quad 0 \quad 1/4 \quad 1/8]$$

and all other rows would remain unchanged (in this case there are sufficient sows in the commercial population for age structure of breeding sows to remain unchanged at the commercial level). Assuming no correlation between growth rate and litter size, **s** is unaffected, since these commercial sows are not selected for growth. If you test this example, you should find that genetic progress in growth rate is lower than previously because a proportion of nucleus females selected for growth rate are replaced by commercial females not selected for growth rate. When examining the response in reproduction, elements of **s** are the breeding values for reproduction of the two selected groups in the initial example, or three groups when some sows in the nucleus come from the commercial population. Note that the elements of **s** describe breeding values of the selected group as a deviation from breeding values of that class of animal. Response in reproduction may increase or decrease depending on the breeding values of the three selected groups. If you try this as an example, you should find that the breeding value of the commercial sows has to exceed the genetic lag (see section 8.6.2) between the selected nucleus and unselected commercial sows before response is increased compared to when taking all sows from the nucleus.

8.6 Applications of Gene Flow Methods

8.6.1 Gene Flow Versus Equilibrium Response Rate

Gene flow methods, as outlined here, provide a method of estimating rates of response from two-path selection schemes and, using the elaboration developed in 8.5, four-path selection schemes. The equilibrium response rate (as $t \rightarrow \infty$) with continuous selection was shown by Hill (1974) to

be identical to that predicted by Rendel and Robertson's formula (see Chapter 3, equation (3.42)) for response per year,

$$R = \frac{S_{sm} + S_{dm} + S_{sf} + S_{df}}{L_{sm} + L_{dm} + L_{sf} + L_{df}} = \frac{\sum_i S_i}{\sum_i L_i}$$

With, for each path, genetic superiorities derived as : $S_i = i_i r_i \sigma_g$

The standardized rate of response from selection in any one path can be found by setting one term on the numerator to 1 and others to 0, to obtain $r = \frac{1}{\sum_i L_i}$. As illustrated earlier (section 8.2.2) this is also the equilibrium genetic contribution to the population from one round of selection.

Gene flow methods have the advantage over the asymptotic response derived in equation (3.42) in that:

1. the rate of response can be estimated in the early time periods, following initiation of a program but prior to attaining equilibrium response rates; and

2. genetic and economic responses in sub-populations that depend on genetic improvement in the nucleus (e.g. multipliers and commercial operations) is readily estimated along with genetic lags (see below).

The combination of 1) and 2) allows estimation of cost-benefits over time, the point at which the program becomes profitable in terms of discounted net returns - costs, and the point at which rate of returns exceed rate of costs (i.e. the point of positive cash-flow). Note that the geneflow equations for predicting cumulative response from time-period to time-period are equivalent to the recursive equations provided in Chapter 3 (e.g. equation 3.7).

8.6.2 Genetic Lag

Genetic lag is the difference in genetic merit between any two contemporary groups of animals; for example, between males entering the nucleus and commercial males, or between breeding females and females entering the commercial population. The genetic lag initially fluctuates over time but eventually reaches an equilibrium. In our pig breeding example, the genetic level for growth rate of pigs entering the commercial population at time period 20 (close to equilibrium) is 310.9 g/day (Table 8.7) compared to 367.6 g/day for pigs entering the nucleus, giving an equilibrium genetic lag of 367.6 - 310.9 = 56.7 g/day. Since the equilibrium response rate is 18.9 g/day (Table 8.4), this is equivalent to 56.7/18.9 = 3.0 time periods or 18 months of genetic improvement.

In general it is desirable to minimize genetic lag between nucleus and commercial populations, and gene flow methods can be used to explore the consequences of alternative breeding structures for genetic lag. Analytical procedures to study genetic lags in stabilized breeding programs were provided by Bichard (1971) and Guy and Smith (1981).

Genetic lag also affects economic weights, and hence selection indexes, since, as noted in Chapter 7, economic weights should be evaluated under the economic and management conditions when genes are expressed, not when selection takes place. Genetic lag can vary substantially between selection paths, and should be evaluated separately for each path. A striking example is the difference between sire to sire and sire to dam paths in a dairy cattle progeny testing scheme (see Table 3.1), where genetic lags between selection decisions and expression of genes in the commercial population are typically about 13 versus 5 years.

8.6.3. Relative Economic Weights

In many cases, economic weights for different traits initially will be estimated per expression, as in Chapter 7; for example, for growth traits the economic value is determined per slaughter pig, while for reproduction traits, value is estimated per farrowing per sow. Gene flow methods can then be used to determine the number of discounted gene expressions of each trait in a given population structure. The net economic value, to be used for deriving selection indexes, will be the economic value per expression multiplied by the discounted number of expressions.

To obtain the discounted number of expressions, the appropriate elements of the vector $\mathbf{m}_{(0)}$ are set equal to 1. The elements of **v** are set to the relative value per expression of the corresponding age-sex class, with a value of one given to the age-sex class for which the economic weight was initially estimated. For example, if the economic weight of milk yield is estimated per lactation for mature cows, elements of **v** corresponding to mature age classes would be one. The element for first lactation cows would be about 0.80, reflecting the lower milk yield and lower value of genetic expressions in first lactation. Cumulate discounted expressions are then estimated in the same way as cumulate discounted returns (see section 8.3) over the required time horizon, using an **n** vector appropriate for each trait in turn.

Considering the pig nucleus example, the vector **n** for growth rate was given in section 8.3 as

$$\mathbf{n}' = [390 \quad 10 \quad 350 \quad 0 \quad 0],$$

and $\mathbf{v}'$ would become, $\mathbf{v}' = [0.01 \quad 0.01 \quad 0.01 \quad 0 \quad 0]$

since expression in 6 and 12 month males and in 6 month females had full economic value, while expression in other age-sex classes had zero value. Appropriate vectors for litter size would be,

$$\mathbf{n}' = [0 \quad 0 \quad 0 \quad 50 \quad 50]$$

and $\mathbf{v}' = [0 \quad 0 \quad 0 \quad 1 \quad 1]$

which assumes that litter size is expressed with equal value in sows at 12 and 18 months.

Rerunning the example shown in Table 8.5, using $\mathbf{m}_{(0)} = [1 \ 0 \ 1 \ 0 \ 0]$, yields the discounted expressions for growth and reproduction following a single round of selection in males, females,

and both sexes together, shown in Table 8.8. It is immediately clear that there are many more expressions of growth rate than litter size. By time 20 (close to equilibrium), there have been about 8.17 discounted growth rate expressions for every litter size expression, whichever sex is initially selected.

Table 8.8 Cumulate discounted expressions for growth rate and litter size from one round of selection in one sex or both sexes.

	Growth rate			Litter size		
Time	Males Selected	Females Selected	Males and Females Selected	Males Selected	Females Selected	Males and Females Selected
1	0	0	0	0	0	0
2	352	176	529	0	0	0
3	352	348	701	23	12	35
4	604	473	1078	46	34	80
5	686	637	1324	62	54	116
6	866	767	1633	84	73	157
10	1388	1306	2695	155	155	301
15	1986	1906	3892	236	225	462
20	2516	2436	4953	308	298	606

In the shorter term, the relative number of expressions depends on the time horizon and the sex initially selected. Because of the different pattern of expression of genes for different traits over time, the relative number of expressions will also depend on the discount rate. As noted in 8.1.4, imposing high discount rates forces a short-term perspective which conflicts with the inherently long-term nature of genetic improvement. It is recommended, therefore, that low discount rates and long time horizons should generally be used when determining relative numbers of expressions for use in estimating economic weights. Similarly, short-term differences between selection paths should generally be ignored. In closed single nucleus systems, each selection path makes the same genetic contribution to commercial expression at equilibrium. In open systems and systems with crossbreeding, different selection paths can make substantially different genetic contributions to commercial expression of different traits and these differences should be incorporated into estimates of economic weights.

8.6.4 Limitations of Gene Flow Methods

The principal limitations to gene flow methods are that the proportion of genes arising from different age classes of parents in the previous time period, i.e. elements of matrix **P**, must be known in advance, along with the genetic merit of selected parents in each pathway. With continuing genetic progress, within a sex, the mean genetic merit of age class *i*-1 will be higher than that of age class *i*, and so on. Also, if information on animals accumulates with age, accuracy of selection, r_{HI}, will increase with age. There may also be changes in genetic variance, σ_g^2, over time. Different selection responses at different ages can be dealt with by adding

elements to vector **s**. The elements at older ages should represent the difference in response between the older age and the previous age at selection. However, one critical question is what should be the contribution of each age class to the following generation for optimum genetic response; a question that gene flow methods cannot tackle directly since they require constant proportions over time. Similarly, it is difficult to optimize complex breeding structures where there may be several methods and sources of genetic improvement with flow of genes between all sources and to commercial animals. Here, one design question would be what are the optimum proportions of genes originating from different sources and how does this affect genetic merit of selected parents from different sources? Again, these optima would change over time in the early years of a program. Optimal proportions selected could be derived using the bisection methods of Chapter 3 for selection across age classes. Some of these problems can be tackled by coupling gene flow to a nonlinear optimization routine, allowing parameters of interest to vary. But such routines often become complex, do not always converge, and require much computing.

Caution should also be used carrying out economic evaluations over prolonged periods. The methods outlined assume linear relationships between genetic change and economic value, an assumption that may be reasonable for small genetic changes (see Chapter 7) but is less reasonable for the genetic changes that accumulate over long periods of time. And, as noted earlier, there is uncertainty over appropriate discount rates, particularly for predicting returns in the distant future. These difficulties argue that, in every case, it would be wise to examine the sensitivity of results obtained to the values assumed for the various input parameters.

8.7 Investment appraisal in competitive markets

The methods discussed previously, evaluate returns from a breeding program in terms of the effect of genetic change on profit of commercial production. For example, the effect of an increase in genetic value of milk yield on profit of cows on a dairy farm. Commercial breeding programs that operate in a competitive market, however, do not derive their income from increased profit of commercial production but from increased market share for their germplasm. Although in a perfect market, the market share that a company's germplasm is able to attain should be directly related to the profit which that germplasm is able to generate in commercial production, such conditions often do not exist.

8.7.1 Economic perspectives in competitive markets

Dairy cattle breeding is a clear example where there is intense global competition for germplasm from progeny-tested bulls and individual bulls are sold on the basis of their estimated breeding values in competition with other companies or countries (see Figure 8.4). In addition, there is competition for contracting bull dams and all competitors have access to semen from all progeny-tested sires for use as bull sires. Thus, in this situation, an AI firm's breeding program is not closed but is part of a single global breeding program, in which, at equilibrium, all AI firms improve at the same rate but with genetic lags, depending on the effectiveness of each firm's improvement program (see Figure 8.5). This also implies that program improvements will have less of an impact on returns than they would have in a closed system.

Based on these considerations, commercial breeding firms must look at breeding programs from a different perspective, as illustrated in Figure 8.6. Important components then are:

1) procurement of superior germplasm
2) product development
3) product marketing

For example, for a conventional dairy cattle progeny testing program for dairy cattle, procurement of superior germ plasm includes sourcing of bull dams and bull sires from the available global cow and bull populations for production of young bulls or, if bull calves have already been produced by individual producers, sourcing of bull calves. The product development phase involves the progeny-testing of these young bulls.

Figure 8.4. A Competitive Global AI Industry

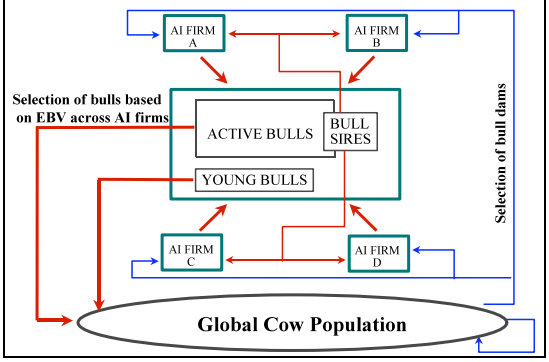


Figure 8.5. Impact of Improving Effectiveness of Selection by Firm A in a Competitive versus a Closed Market

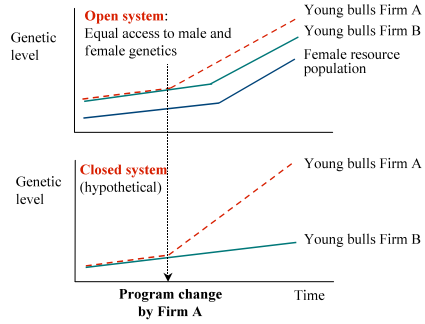


Figure 8.6. Breeding Company View of AI Progeny Testing Programs

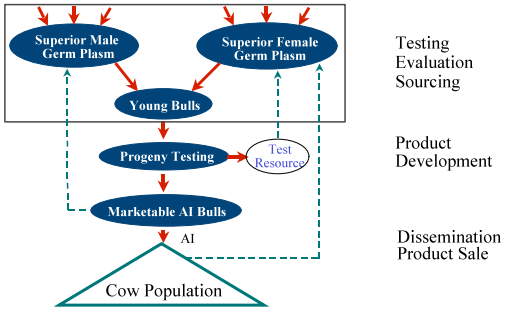
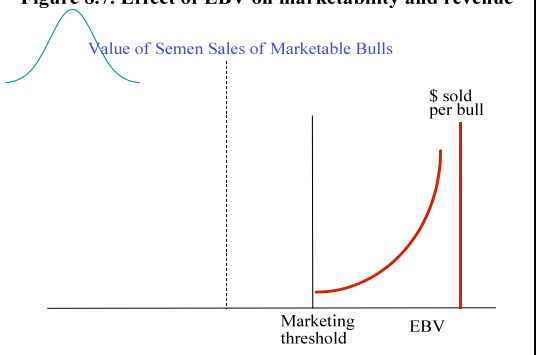


Figure 8.7. Effect of EBV on marketability and revenue

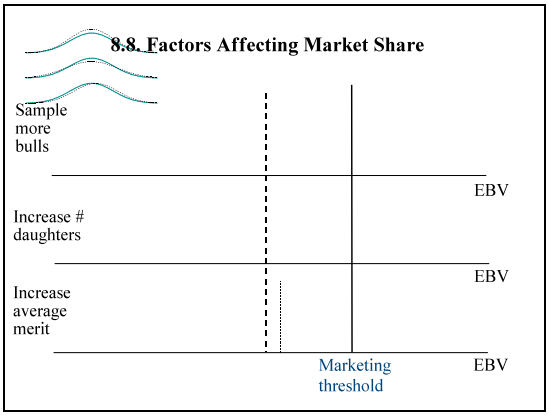


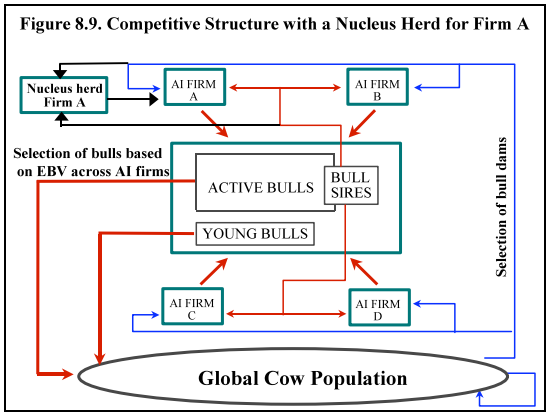
In such programs, returns are generated from the sale of germ plasm from marketable bulls. Only a limited number of bulls are required to breed the population, so that only those bulls above a certain threshold are likely to be saleable. A similar situation may also arise in some beef cattle and sheep breeding markets where animals (or their semen) are sold for breeding on the basis of their EBV. The situation is illustrated graphically in Figure 8.6. Only bulls above the threshold are saleable. For marketable bulls, there will also be non-linear relationship between the value of product sold and EBV of the bulls: more semen is sold from the top marketable bulls, plus it is sold at a higher price per unit.

In such a competitive market for breeding stock, there are three ways to increase market share in terms of share of value of germ plasm sold in the global market:

1) increase the size of the program
2) increase the mean of the germ plasm that is entered in the product development phase
3) increase the differentiation of germ plasm during the product development phase.

The impact of these three strategies on market share are illustrated in Figure 8.8 for a dairy cattle progeny testing program. For such a program, increasing the size of the program (1) amounts to increasing the number of bulls tested, increasing the mean of germ plasm entered the product development phase (2) amounts to increasing the mean genetic value of young bulls entered, and increasing differentiation of the germ plasm during product development (3) amounts to increasing progeny group size. The latter will increase the accuracy of EBV following progeny test, which increases the variance of EBV ($\sigma^2_{\hat{g}} = r^2\sigma^2_g$).





In this situation increasing the mean of the genetic value of procured germ plasm can be achieved in two ways:

a) increasing effectiveness of selection of superior germ plasm from the resource population
b) increasing genetic progress in the resource population

Objective a) can be achieved by applying the principles outlined in Chapter 3 by selecting the best sires and dams based on EBV from all available candidates, regardless of age or accuracy. With regard to objective b), in a conventional progeny testing program, an individual AI firm has limited impact on genetic gain in that population. In addition, all AI firms are in direct competition with each other for procurement of superior bull dams and thus source from the same population of selection candidates. Addition of a nucleus herd changes this situation. As illustrated in Figure 8.9 access to an AI firm's nucleus herd is restricted to that AI firm. This allows that firm to have some degree of protection of its female genetic resources. If selection procedures in the nucleus are more effective than in the general population, this will increase genetic gain and genetic lags, as illustrated in Figure 8.5.

The impact of alternative breeding programs and of nucleus herds on market share were studied by Dekkers and Shook (1990a,b) using a semi-stochastic simulation program (Dekkers and Shook 1990c). In this model, cows in the population were modeled deterministically as age

groups with defined means and variances of breeding values. Sires appeared in the model as individuals (i.e. stochastically) with an EBV based on pedigree information and a daughter average performance. This process provided the actual number of sires achieving saleable status for each organization in each time period for a given replicate. Running the program many times allowed estimation of mean and variance of performance of a given selection strategy.

The impact of selection strategies on market share can, however, also be evaluated using a complete deterministic model. The deterministic model would model the mean and variance of true and estimated breeding values of each sex-age group. For bulls, means and variances would be modeled by AI firm. Then, assuming multivariate normality, determination of the number of marketable bulls provided by an AI firm would be obtained by determining the unique truncation point across distributions of EBV for all available age groups and AI firms such that the correct number of bulls is selected. Multiple truncation procedures described in Chapter 3 can be used for this purpose.

Let n_{ijt} be the number of marketable bulls from age group i of AI firm j at time t, which are selected from a Normal distribution with mean $\bar{\hat{g}}_{ijt}$ and variance $r_{ijt}^2\sigma_{g_{ijt}}^2$. Also, let $\phi(\hat{g},t)$ represent the functional relationship between EBV of a marketable bull and value of semen sold from that bull in a particular time period. Then, returns from semen sales from age group ij at time t, R_{ijt}, can be determined by integrating the relationship between EBV and value of sales over the truncated distribution:

$$R_{ijt} = n_{ijt} \int_{\hat{g}=\hat{g}_{M_t}}^{\infty} f(\hat{g} \mid \bar{\hat{g}}_{ijt}, r_{ijt}^2\sigma_{g_{ijt}}^2)\phi(\hat{g},t)\ \partial\hat{g}$$

where $\hat{g}_{M_t}$ is the marketing threshold for time t.

Returns per age group can be summed over age groups within AI firm to determine total returns at a given time t, discounted to determine the present value of those returns, and summed over time periods.

8.7.2 Example of economic optimization of progeny group size

Dekkers et al. (1996) used the semi-stochastic model of Dekkers and Shook (1990) to optimize progeny group size for a fixed testing capacity for young bulls in a competitive market (Figure 8.9). The principal question asked was, what is the combination of number of bulls sampled and number of daughters tested per bull that would maximize the net profits of an AI organization that is in competition with three other companies for sale of bulls into a market requiring 36 bulls, each selling 25,000 doses of semen per 6 mo. period? The base situation was each AI firm testing 60 bulls per annum, with 60 daughter records per bull. The performance of one organization, which varied its sampling program, was evaluated, while all other organizations maintained the original sampling policy. Selection for net economic merit with h^2 = 0.25 was assumed.

Semen prices were assigned to an individual bull, k, based on the following linear or quadratic function of EBV: $\phi(\hat{g}_k, t) = p_{min} + b_t(\hat{g}_k - \hat{g}_{M_t})^q$

where p_{min} is the semen price assigned to the lowest ranking marketable bull across A.I. firms in a given time period ($4), and $(\hat{g}_i - \hat{g}_{M_t})$ is the difference in EBV between the i^{th} marketable bull and the lowest ranking marketable bull in time t. Exponent q is equal to 1 and 2 for the linear and quadratic price functions. Coefficient b_t was determined for each time period such that the average semen price remained constant ($15).

For each A.I. firm, discounted gross returns from semen sales were computed per semi-annual cohort of young bulls by discounting and summing over time the semi-annual returns from semen sales for each marketable bull in the cohort: $R_{ij} = 25000 \sum_t \sum_{k \varepsilon B_{t,ij}} \phi(\hat{g}_{ijk}, t)(1+r)^{-(t-1)/2}$

where R_{ij} is the total discounted gross return from the cohort of bulls sampled by firm j and born in semi-annual period i, r is the annual discount rate (5%), t is a semi-annual period in which bulls from cohort ij are marketable, and $B_{t,ij}$ is the set of bulls from cohort ij that are marketable at time t, and. Factor 25000 represents the number of doses sold per marketable bull per half year.

Total discounted sampling costs per cohort (C) were computed as: $C = N(F + VD)$, where N is the number of bulls sampled, F is the fixed cost per bull sampled, which includes all costs for a young bull associated with purchase, housing, feeding, etc., D is progeny group size, and V is the variable cost per daughter record, which mainly consists of incentives to producers for use of young bull semen. Similar to returns, costs were discounted to the time of birth of the cohort of young bulls at 5% per year.

The model was run for twenty combinations of numbers of bulls sampled and progeny group size for AI firm A, keeping the program of the other three AI firms at the base level of 60 bulls sampled and 60 daughters per bull tested.

Summary data for number of bulls marketed and gross returns per cohort were then analysed using response surface methodology by fitting the following quadratic response surface to the twenty data points: $Y(N_k, D_l) = = b_1 N_k + b_2 N_k^2 + b_3 D_1 + b_4 D_1^2 + b_5 N_k D_1 + e_{k1}$

where $Y(N_k, D_l)$ is the mean response for firm A when it samples N_k bulls per year with D_l progeny per bull.

Response surfaces for discounted net returns per cohort were obtained by subtracting $C = N(F+VD)$ from the response surface for gross returns or, equivalently, by subtracting F and V from parameter estimates for b_1 and b_5.

In many cases, the number of cows that is available to an A.I. firm for insemination with young bull semen is limited. Test capacity was defined in terms of the number of young bull daughters per annual cohort of bulls sampled by an A.I. firm: $T = ND$. Optimum utilization of a fixed test capacity in terms of number of bulls to sample versus progeny group size was investigated by reformulating the estimated response surface equations by substituting $N = T/D$:

$$Y(D,T) = \mu + b_1 T/D + b_2 T^2/D^2 + b_3 D + b_4 D^2 + b_5 T$$

Optimum progeny group size for a given test capacity T was derived by maximizing $Y(D,T)$ with regard to D, which resulted in the following quartic polynomial for optimum progeny group size (D^*):

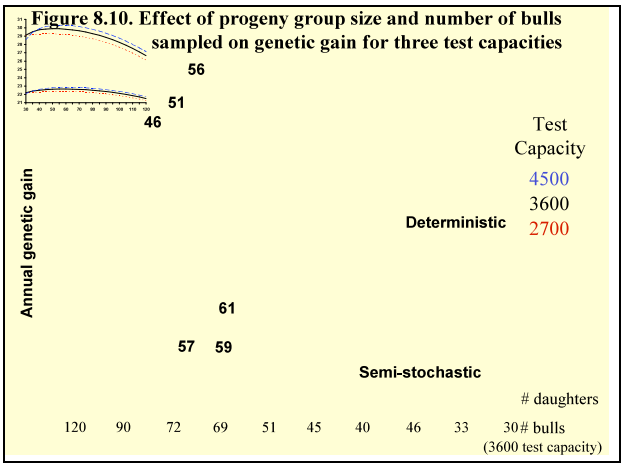
$$2b_4 D^{*4} + b_3 D^{*3} - b_1 T D^* - 2b_2 T^2 = 0$$

with a shadow value for T of:

$$\lambda_T = b_1 / D^* + 2b_2 T / D^{*2} + b_5$$

Solutions for D^* were obtained by using Maple V (Maple V, 1994, Waterloo Maple Software, Waterloo, ON, Canada). Optima and shadow values for net returns per cohort were obtained by replacing estimated parameters b_1 and b_5 for gross returns by (b_1 - F) and (b_5 - V).

Figure 8.10 shows the effect of progeny group size and numbers of bulls sampled on genetic gain for fixed test capacities. To determine the effect on genetic gain, breeding programs were changed for all four AI firms. Genetic gain was also estimated using a deterministic model based on the asymptotic equations developed in Chapter 3 but ignoring the Bulmer effect.



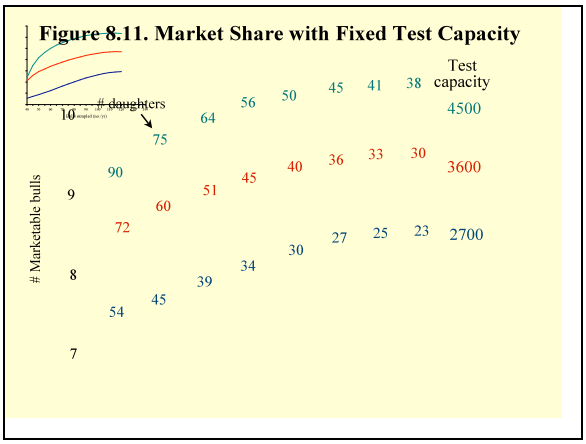
Increasing progeny group size (and decreasing number of bulls sampled) had a greater effect on deterministic predictions of genetic gain than on genetic gain predicted based on the semi-stochastic model; when progeny group size was greater than 50, the effect of reducing the number of bulls sampled on selection intensity outweighed increases in accuracy for deterministic predictions. This was less the case for semi-stochastic predictions. The reason is that relative increases in accuracy with progeny group size are larger when the effects of selection are accounted for (see Chapter 4).

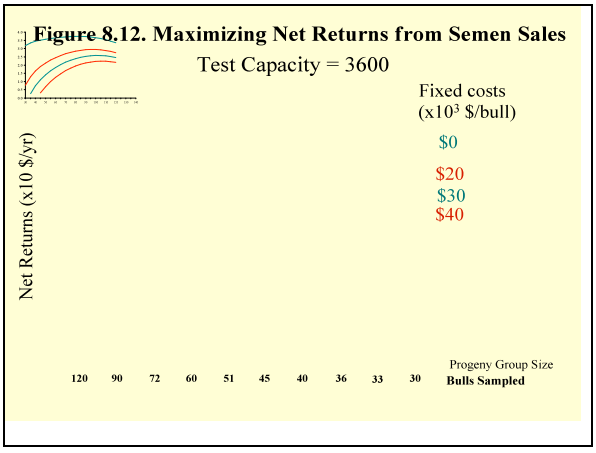
For the deterministic method, optimum progeny group size was 46, 51, and 56 for test capacities of 2700, 3600, and 4500 young bull daughters (Figure 8.10). Optimum progeny group size was larger under the stochastic method (57, 59, and 61 daughters). For both methods, small to moderate deviations from optimum progeny group size had a small effect on genetic gain.

Figure 8.11 shows how market share, in terms of number of marketable bulls, depends on progeny group size and number of bulls sampled for a fixed test capacity. Progeny group size that resulted in the maximum number of marketable bulls increased with test capacity: optimum progeny group size was 22, 31, and 40 for test capacities of 2700, 3600, and 4500. For a given number of bulls sampled, differences between lines in figure 8.11 reflect the effect of progeny group size on

market share through its effect on accuracy. As expected, the effect of progeny group size was smaller for larger progeny group sizes (left side of Figure 8.11).

Figure 8.12 shows the effect of progeny group size and number of bulls sampled on discounted net returns for a fixed test capacity of 3600 and a quadratic price function, for varying fixed costs per bull. For a fixed test capacity, total cost for young bull daughter incentives are equal to *TV* and unaffected by the number of bulls sampled and progeny group size. Incentive costs per daughter do, therefore, not affect the shape of the contour lines in Figure 8.12, nor the optimum progeny group size. Changing the number of bulls sampled does, however, affect total fixed costs. A fixed cost of $0 gives discounted gross returns. Increasing fixed cost per bull increased the optimum progeny group size (Figure 8.12). Curves were, however, relatively flat around the optimum. For typical fixed costs per bull in Canada of $30,000, optimum progeny group size was 102.





Global optima for other cost scenarios and for the linear price function are in Table 8.9. The optimum number of bulls to sample was highly sensitive to the cost of sampling, but sampling cost had a limited effect on optimum progeny group size. Comparing optima for the linear and quadratic

price function (Table 8.9) shows that price function had almost no effect on the optimum number of bulls to sample but optimum progeny group size was slightly lower for the linear price function.

Table 8.9. Optimal progeny group size for a fixed test capacity (from Dekkers et al. 1996)

	Test Capacity					
	2700		3600		4500	
	Fixed costs per bull ($x10^3$)					
Deviation from base[1]	$20	$30	$20	$30	$20	$30
	Optimum progeny group size					
None	98	102	97	102	97	103
Linear price function	92	97	91	98	91	98
Population size +20%	96	100	95	100	95	100
Population size -20%	100	104	100	105	100	107
Semen price +20%	97	100	95	100	95	100
Semen price -20%	100	104	100	105	100	107
Interest 8%	100	104	99	104	100	106
One competitor at 100	99	102	99	103	99	105
	--Extra profit ($x10^4$ $/yr) at optimum versus at 60 daughters/bull--					
None	49	66	49	73	56	86
Linear semen price	28	44	28	50	34	61
One competitor at	54	72	56	80	61	92
	----------Shadow value of test capacity ($/daughter)---------					
None	376	274	338	238	289	195
Linear semen price	397	287	352	246	305	207
Population size +20%	454	348	416	313	377	278
Population size -20%	259	161	229	134	200	109
Semen price +20%	495	389	448	344	398	300
Semen price -20%	259	161	229	134	200	109
Interest 8%	282	183	251	155	219	128
One competitor at 100	261	163	242	145	222	129

[1] In the base situation population size is 950,000 cows, semen price is based on a quadratic function of estimated breeding value, average semen price is $15, interest rate is 5% per year, and the three competing AI firms sample 60 bulls with 60 daughters each.

Table 8.9 also shows the additional profit per annual cohort of bulls of sampling at the optimum instead of at 60 daughters per bull. Sampling at 102 daughters per bull instead of at 60 increased discounted net return by $730,000 per annual cohort or by 38%. Reducing fixed costs to $20,000

per bull reduced the optimum progeny group size by five daughters and also reduced the economic benefit of moving to the optimum scheme.

Optimum progeny group size was little affected by test capacity (Table 8.9). However, the economic benefit of optimising progeny group size was greater for larger test capacity, at least in absolute terms. Table 8.9 also shows the shadow value of test capacity, which is equal to the extra profit that can be expected when increasing test capacity by one daughter under an optimum design. Shadow values do not incorporate incentive costs (V) but represent the maximum incentives that could be paid to increase test capacity by one daughter without reducing profit. Shadow values were lower for larger test capacities and for larger fixed costs per bull. Shadow values were, however, greater than the $180 incentive that was on average provided to producers by Canadian A.I. organizations. Table 8.9 also shows that the optimum was little affected by population size, semen price, or interest rate.

Market share of an A.I. firm depends not only on the firm's own breeding program but also on the breeding program of its competitors. Previous results were for situations in which competitors conducted the base breeding program of sampling 60 bulls with a progeny group size of 60. Optimum progeny group size was, however, little affected when one of the competitors sampled bulls with a progeny group size of 100 daughters instead of 60 (Table 8.9). Although profit was significantly lower when one competitor sampled at 100 daughters per bull instead of at 60, the economic benefit to firm A of sampling at the optimum progeny group size instead of at 60 daughters per bull was little affected by the breeding program of its competitor. This represents the opportunity cost of not changing to the optimum of 100 progeny per bull.

Inbreeding and its Impact on Design of Breeding Programs

Jack Dekkers

Inbreeding = mating of individuals that are related by ancestry
- **→ may carry alleles that are identical-by-descent (IBD)** (vs. by state or IBS)
- **→ increases probability that progeny will by homozygous**

Inbreeding coefficient = probability individual's pair of alleles at a locus are IBD
= coefficient of coancestry of parents

Coefficient of coancestry individuals *x* and *y*
= prob(a random allele from *x* (at a given locus) is IBD to a random allele from *y*)

Additive genetic relationship *x*, *y* = 2 x coefficient of coancestry between *x* and *y*

Effects of inbreeding → increased homozygosity
- Increased incidence of recessive disorders
- Inbreeding depression → reduced phenotypic performance
- Loss of genetic variance → reduction in rates of genetic improvement

Genotypic frequencies and mean performance in a population with inbreeding coefficient F for a single gene with 2 alleles with inbreeding coefficient F

P = freq(A_1) q = freq(A_2)

Genotype	Frequency	Value	Frequency x value
A_1A_1	p^2 +pqF	+a	p^2a +pqaF
A_1A_2	2pq -2pqF	d	2pqd -2pqdF
A_2A_2	q^2 +pqF	- a	$-q^2$a -pqaF
		Sum = M_F	= a(p-q)+2dpq-2dpqF = a(p-q)+2dpq(1-F)

Without inbreeding: mean = M_0 = a(p-q)+2dpq

Inbreeding depression = $M_0 - M_F$ = -2dpqF

Summed over loci (no epistasis): $M_0 - M_F = -2F\Sigma dpq$

Impact of inbreeding on genetic variance:

Infinitesimal genetic model

No inbreeding: $\sigma^2_{g_{(t+1)}} = \frac{1}{4}(1-k_s r^2_{s_{(t)}})\sigma^2_{g_{(t)}} + \frac{1}{4}(1-k_d r^2_{d_{(t)}})\sigma^2_{g_{(t)}} + \frac{1}{2}\sigma^2_{g_{(o)}}$ $\sigma^2_{g_{(o)}}$ = base pop. var.

With inbreeding: Mendelian sampling variance = $(1-\bar{F}_{s(t)})\frac{1}{4}\sigma^2_{g_{(o)}} + (1-\bar{F}_{d(t)})\frac{1}{4}\sigma^2_{g_{(o)}}$

$= (1-\frac{1}{2}(\bar{F}_{s(t)} + \bar{F}_{d(t)}))\frac{1}{2}\sigma^2_{g_{(o)}}$

$$\sigma^2_{g_{(t+1)}} = \frac{1}{4}(1-k_s r^2_{s_{(t)}})\sigma^2_{g_{(t)}} + \frac{1}{4}(1-k_d r^2_{d_{(t)}})\sigma^2_{g_{(t)}} + (1-\frac{1}{2}(\bar{F}_{s(t)} + \bar{F}_{d(t)}))\frac{1}{2}\sigma^2_{g_{(o)}}$$

Only Mendelian sampling variance is affected by inbreeding, depending on inbreeding coefficient of parents, rather than inbreeding of the progeny

PREDICTION OF RATES OF INBREEDING

$$\Delta F = \frac{F_{t+1} - F_t}{1 - F_t} = \frac{1}{2N_e} \qquad N_e = \frac{1}{2\Delta F}$$

N_e = Effective population size = number of individuals that would give rise to a rate of inbreeding ΔF if bred as an idealized population

Idealized population

- **Random mating** (incl. selfing), **No selection**
- **Discrete (non-overlapping) generations**
- **Random distribution of family size** – each individual has equal probability to contribute a progeny

Factors affecting rate of inbreeding in a closed non-idealized population

In a population that is not under selection:

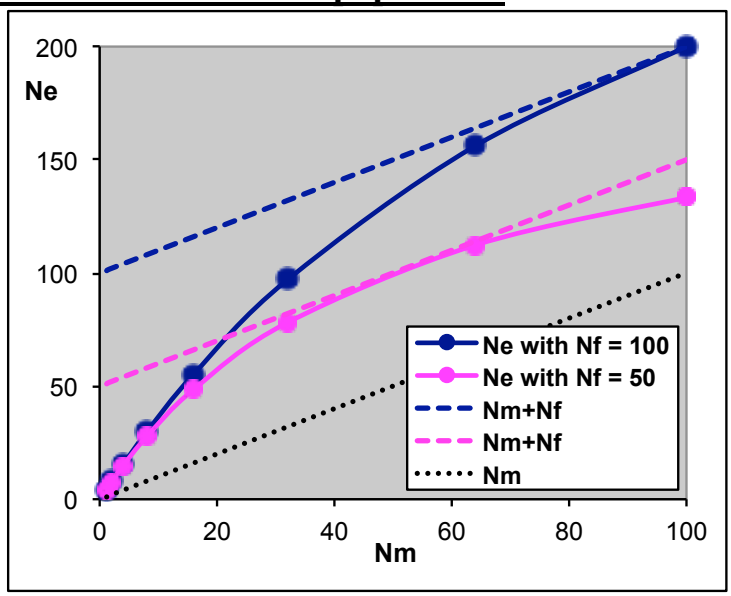
- **# males (N_m) and females (N_f) used for breeding**
  - ←→ population size
  - ←→ selection intensity

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f} \quad \text{(Wright, 1922)}$$

$$\Delta F = \frac{1}{8N_m} + \frac{1}{8N_f}$$

→ N_e is less than the # parents; $\leq N_m + N_f$

→ N_e is driven primarily



by the smaller of N_m and N_f

- **Variance of family size** ←→ unequal use of parents (and their progeny)
  - family size = number of progeny that become breeding parents
  (Hill, 1979 Genetics 92:317)

$$N_e \approx \frac{8N}{V_{km} + V_{kf} + 4} \quad \rightarrow \quad \Delta F = \frac{1}{2N_e} = \frac{V_{km} + V_{kf} + 4}{16N}$$

N = Total population size (½N males, ½N females)

V_{km} = Var(# progeny per male)

V_{kf} = Var(# progeny per female)

V_{km} and V_{kf} affected by unequal use of individuals for breeding
- **selection**
- **differential use of selected individuals**

Mean family size = 2 (each parent → 2 progeny to maintain population size)

Idealized population: distribution of family size = Binomial ≈ Poisson

→ $V_{km} = V_{kf}$ = mean family size = 2 → $N_e = \frac{8N}{2+2+4} = N$

Variance of family size can be reduced (by the breeder) by ensuring that all selected parents equally contribute breeders for the next generation

- within family selection – select best male and best female from each fullsib family → $V_k = 0$ → $N_e \approx 2N$

- **Generation Interval** ←→ shorter → greater rate of inbreeding per year

$$N_e \approx \frac{8N_cL}{V_{km} + V_{kf} + 4} \quad \rightarrow \quad \Delta F/yr = \frac{1}{2N_e}/L = \frac{V_{km} + V_{kf} + 4}{16N_cL^2}$$

*Most of the above equations calculate inbreeding per generation

N_c = total # progeny per year

L = average generation interval (across males and females)

Selection increases inbreeding through: (Verrier et al. 1990)

- **Probability of co-selection of relatives** ←→ correlation of the selection criterion between relatives
- **Inheritance of selective advantage** – progeny of good parents are more likely to be selected themselves, as are their descendants
  → increased variance of family size

*Half sib and full sib records will increase the correlation of EBV's; progeny records will decrease the correlation. All of these equations can be found in 6.1-6.2 and can be calculated in stepEBV.

More accurate methods to predict rates of inbreeding in populations under selection

In part based on notes from Bijma and van Arendonk

See Wray and Thompson (1990 Genet. Res. 55:41), Verrier et al. (1990)

Previous methods are 'single generation' methods

- account for differential contributions of ancestors to future generations through differential numbers of progeny that become breeding parents
- do not account for additional differences in an ancestral contributions through differential numbers of grand progeny that become breeding parents

ΔF ~ variance of long-term genetic contributions among ancestors (Wray & Thompson '90)

Theory of long-term genetic contributions

Wray and Thompson 1990. Genet. Res. 55:41
Woolliams and Bijma 2000 Genetics 154:1851
Bijma et al. 2000 Genetics 156:361
Woolliams et al. 1999 Genetics 153:1009
Bijma and Woolliams 2000 Genetics 156:361
Bijma et al. 2001 J.Anim.Sci. 79:840

$r_{i,t_1}(j,t_2)$ = **Genetic contribution** of ancestor *i* born at generation t_1 to an individual *j* born at generation t_2 ($t_2 > t_1$)l

= proportion of genes of *j* expected to derive by descent from ancestor *i*.

Note: Full-sibs share ½ of their genes but make no genetic contribution to each other.

$r_{i,t_1}(t_2)$ = **Mean genetic contribution** of ancestor *i* born at generation t_1 to generation t_2

= the average proportion of genes among individuals in generation t_2 contributed by ancestor *i*

- $E(r_{i,t_1}(t_2)) = \frac{1}{2N_m}$ for male ancestors (N_m = # male ancestors)

  $= \frac{1}{2N_f}$ for female ancestors

- $r_{i,t_1}(t_2)$ differ between ancestors due to differences in # **progeny** and differences in the **selective advantage** of descendents

$$\sum_i r_{i,t_1}(t_2) = 1$$

- as t_2- t_1 increases, contributions from a given ancestor stabilize and become similar across individuals in generation t_2 → $\underset{j}{\mathrm{Var}}(r_{i,t_1}(j,t_2)) \to 0$
- t_2- t_1→ infinity, genetic contributions from a given ancestor are the same for all individuals in time t_2

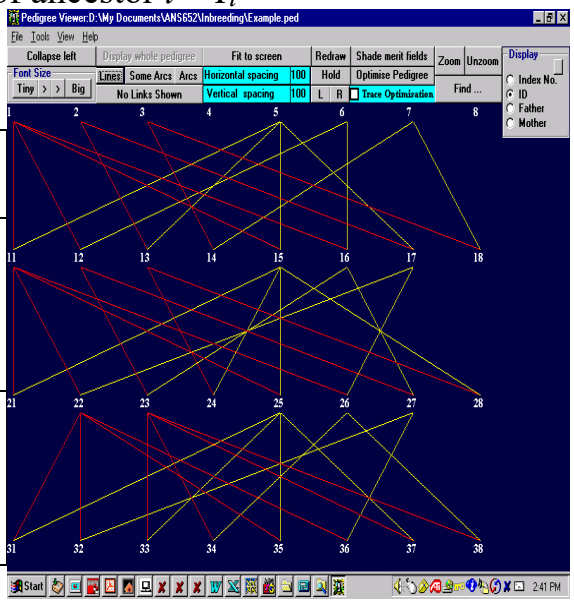
= **long-term genetic contribution** of ancestor i = r_i

Pedigree to illustrate concept of genetic contributions

Each generation contains 4 males and 4 females.

Base population:1-4=males ; 5-8=females

	Generation 1			Generation 2			Generation 3		
sex	ind	sire	dam	ind	sire	dam	ind	sire	dam
Male	11	1	5	21	11	15	31	22	25
	12	1	6	22	11	17	32	22	27
	13	2	5	23	12	16	33	23	25
	14	3	7	24	13	15	34	23	26
Female	15	1	5	25	11	15	35	22	25
	16	1	6	26	11	17	36	22	27
	17	2	5	27	12	16	37	23	25
	18	3	7	28	13	15	38	23	26



Contribution of each ancestors to each offspring: $r_{i,t_1}(j,t_2)$

and **mean contribution** of each ancestor: $r_{i,t_1}(t_2)$

	Ancestors							
***Offspring*[1]**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
				Generation 1				
11/15	0.5	0	0	0	0.5	0	0	0
12/16	0.5	0	0	0	0	0.5	0	0
13/17	0	0.5	0	0	0.5	0	0	0
14/18	0	0	0.5	0	0	0	0.5	0
Mean contribution	***0.25***	***0.125***	***0.125***	***0***	***0.25***	***0.125***	***0.125***	***0***
				Generation 2				
21/25	0.5	0	0	0	0.5	0	0	0
22/26	0.25	0.25	0	0	0.5	0	0	0
23/27	0.5	0	0	0	0	0.5	0	0
24/28	0.25	0.25	0	0	0.5	0	0	0
Mean contribution	***0.375***	***0.125***	***0***	***0***	***0.375***	***0.125***	***0***	***0***
				Generation 3				
31/35	0.375	0.125	0	0	0.5	0	0	0
32/36	0.375	0.125	0	0	0.25	0.25	0	0

33/37	0.5	0	0	0	0.25	0.25	0	0
34/38	0.375	0.125	0	0	0.25	0.25	0	0
Mean contribution	***0.406***	***0.104***	***0***	***0***	***0.313***	***0.187***	***0***	***0***

[1] Each generation consisted of a full sib male and female, which have equal contributions. Contributions can be derived from pedigree. Contributions across ancestors sum to 1. Contributions of a given ancestor to descendants become less variable over time.

The variability of genetic contributions will tend to be variable until they stabilize which is the "long term contribution

Use of long-term genetic contribution theory to predict ΔF

Rate of inbreeding is related to the variance of long-term contributions among ancestors

- Asymptotic ΔF = ¼ * sum of squares of long-term contributions

$$\Delta F = \tfrac{1}{4}\Sigma r_i^2$$

Example: 20 selected parents per generation (ignoring that there are two sexes). Pedigree analysis quantifies the contribution of each parent to a particular generation. Their contribution will sum to 1; genetic contributions always sum to 1 per generation.

Consider two extreme cases:

1) the contribution of each individual is the same, $r = 0.05$ for all individuals

   → $\Delta F = \tfrac{1}{4}(0.05^2 + 0.05^2 + .. 0.05^2) = 0.0125$ = 1.25% per generation

2) contributions differ between individuals: $r = 0.25$ for 4 best parents $r = 0$ for rest

   → $\Delta F = \tfrac{1}{4}(0.25^2 + 0.25^2 + .. 0^2) = 0.0625$ = 6.25% per generation.

   ➔ variance in the contributions of ancestors → higher inbreeding

Example: If there were 2 male and 2 female ancestors in a generation with contributions 3/8, 1/8, 5/16 and 3/16 (note the 2 males sum to ½ and the two females sum to ½) then the estimate of ΔF attributed to that generation is

$$\Delta F = \tfrac{1}{4}\left\{\left(\frac{3}{8}\right)^2 + \left(\frac{1}{8}\right)^2 + \left(\frac{3}{16}\right)^2 + \left(\frac{6}{16}\right)^2\right\} = 0.072$$

If the population were in a steady state we would expect approximately the same answer every generation and the average over generations would be the expected ΔF.

The importance of relationship $\Delta F = \tfrac{1}{4}\Sigma r_i^2$ is:

- It is general and applies to both selected and unselected populations.
- It relates ΔF to terms that can be found in the relationship (A) matrix.
- Predictive forms can be developed from the relationship.
- Strictly it is an approximation, but the proportional error (an underestimate) is of the same order as those previously developed for unselected populations.
- Its form will lead to insights into how optimal selection schemes work.

A problem for prediction of ΔF is that it requires estimation of the variance of long term contributions

Use of long-term contributions to predict rates of inbreeding (Woolliams and Bijma, 2000):

$E(\Delta F) = E\{\frac{1}{4}\Sigma r_i^2\} = \frac{1}{4}T^*E[r_i^2]$ T = # selection candidates

$E[r_i^2]$ = expectation of square of contributions across all candidates

- Requires mean and variance of long-term contributions

BUT: Woolliams et al. ('99) showed: ΔF ~ square of *expected* long-term genetic contributions of selected parents

$\Delta F = \frac{1}{4}\Sigma r_i^2 \rightarrow \Delta F = \frac{1}{2}NE(r)^2$

➔ prediction of the *variance* of long term genetic contributions not needed

Following Woolliams et al. (1999), the (long-term) genetic contribution of an ancestor can be predicted by regression on its breeding value, using the model:

$$E(r_i \mid g_i) = r_i = \alpha + \beta g_i$$

r_i = expected genetic contribution of ancestor *i*

α = expected genetic contribution for an average ancestor

ß = regression coefficient of the genetic contribution on the BV (g_i) of the ancestor

For discrete generations: α is determined by the number of parents:

For male ancestors, $\alpha = \frac{1}{2}N_s$

For female ancestors $\alpha = \frac{1}{2}N_d$

ß describes that:

- selective advantage influences the selection decisions in offspring generation
- selective advantage is inherited →has an influence beyond offspring generation

➔ Two mechanisms must be described to enable the prediction:

1) better parents have on average more offspring that are selected as parents.
2) the selected offspring of better parents are on average better, which also affects the genetic contributions.

In short, the procedure (implemented in SelAction) is as follows:

1) A regression model is used to predict the long-term contribution of selected parents,

$$E(r) = \alpha + \beta(g - \bar{g})$$

E(r) = expected contribution given the true BV of an individual
α = the contribution of an individual with an average BV
β = increase of the contribution of parents with a higher BV.

**Genup calculates the long-term contributions, shows pedigree trees

The second term accounts for parents with high BV having more selected offspring.

α and β can be derived mathematically (see Woolliams et al. 1999).

2) Calculate the square of the expected contributions of selected parents:

$$E(r)^2 = \alpha^2 + \beta^2\sigma_A^2(1 - k\rho^2)$$ (Woolliams et al. 1999)

$\sigma_A^2(1\text{-}k\rho^2)$ = genetic variance of <u>selected</u> parents; ρ = accuracy

The above gives E(r)^2, but in fact we need to calculate E(r^2):

E(r)^2 = square of the expected contributions of selected parents.
E(r^2) = expectation of the squared (actual) contribution of selected parents.

Under certain conditions E(r^2) = 2[E(r)]^2 , leading to the following result to predict ΔF:

$$\Delta F = \tfrac{1}{2} N\mathrm{E(r)}^2$$

N = # parents E(r)^2 = square of the expected contributions of selected parents, predicted as given above

Note: ¼ in ΔF= ¼Σr$_i^2$ is replaced by ½ because we have replaced the square of the actual contributions Σr^2 by the square of the expected contributions, NE(r)^2.

Woolliams et al. (1998) also extended the method to overlapping generations.

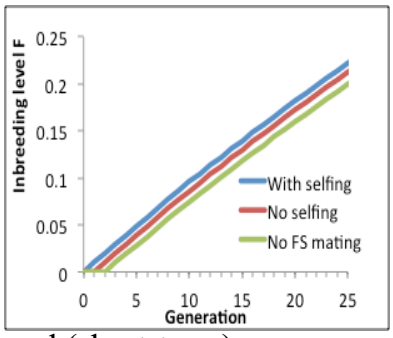
Design of breeding programs with controls on inbreeding

<u>Short-term</u> response is maximized by (based on ΔG = i r σ_g):

- selection on BLUP EBV – i.e. maximize accuracy r
- select only the best individuals – i.e. maximize intensity i - given repro rates

But this may not maximize long(er)-term response because it leads to higher ΔF.

Strategies to control inbreeding



- (Mate selected parents such that inbreeding of progeny is minimized)
  - limited effect on long-term rates of inbreeding
- Select more animals - increase population size → increased costs
  - reduce selection intensity →reduced (short-term) response
- Reduce probability of co-selection of relatives
  - impose restrictions on selection of relatives (e.g. 1/full-sib family)
  - increase h^2 in genetic evaluation (affects both pedigree and progeny info)
    - decrease weight on pedigree information
    - control the average relationship among selected parents
      - cost factor on average relationship (Brisbane and Gibson 1994)
      - constraint on average relationship (Meuwissen 1997, JAS 75:934)
- Introduce outside genetics

Toro & Perez-Enciso 1990, GSE:

Optimal weight given to family information

Table I. Expected and observed cumulative selection response, R_E and R_o, and inbreeding coefficient (%), F_E and F_o, after 5 generations of selection, as a function of the weight given to family information, λ. The initial additive variance was 50.

Standard errors ranged from 0.24 to 0.28 (R_o), and from 0.16 to 0.43 (F_o).

	λ	R_E	R_o	F_E	F_o
$h^2 = 0.10$	0	6.64	6.26	12.86	10.65
	1	13.80	11.74	13.63	14.65
	2	17.22	13.79	19.32	21.45
	3	18.45	15.27	23.81	26.28
	4	18.80	14.97	26.87	30.03
	5	18.88	14.89	28.98	32.39
	6	18.85	14.88	30.48	34.37
	7	18.77	14.46	31.55	35.06
	6.33	18.83	15.00	30.85	34.52
$h^2 = 0.30$	0.0	12.16	11.81	12.86	10.06
	1.0	23.72	16.87	15.38	12.97
	1.5	26.14	19.62	19.08	17.22
	2.0	27.27	21.62	22.24	25.06
	2.5	27.65	21.50	24.71	26.90
	3.0	27.73	21.73	26.64	29.06
	3.5	27.68	21.70	28.14	30.08
	4.0	27.57	21.36	29.36	31.94
	3.73	27.64	21.09	28.71	31.92

Restriction on the distribution of family size

Table II. Expected and observed cumulative selection response, R_E and R_o, and inbreeding coefficient (%), F_E and F_o, after 5 generations of selection, as a function of family size. The initial additive variance was 50, $h^2 = 0.10$.

Standard errors ranged from 0.21 to 0.32 (R_o), and from 0.16 to 0.43 (F_o).

Case	*Distribution of family size*	R_E	R_o	F_E	F_o
1	4 4 0 0 0 0 0 0	17.42	12.77	42.76	41.40
2	4 3 1 0 0 0 0 0	18.17	13.94	35.81	35.88
3	4 2 2 0 0 0 0 0	17.87	13.85	33.26	33.88
4	4 2 1 1 0 0 0 0	17.78	14.85	30.59	31.94
5	3 3 2 0 0 0 0 0	17.30	13.72	30.59	31.17
6	3 3 1 1 0 0 0 0	17.21	14.34	27.80	28.60
7	4 1 1 1 1 0 0 0	16.38	13.48	27.80	28.62
8	3 2 2 1 0 0 0 0	16.91	14.99	24.87	26.56
9	3 2 1 1 1 0 0 0	16.24	14.32	21.79	24.06
10	2 2 2 2 0 0 0 0	14.91	12.66	21.79	22.89
11	2 2 2 1 1 0 0 0	14.85	13.18	18.57	20.24
12	3 1 1 1 1 1 0 0	14.23	12.78	18.57	20.17
13	2 2 1 1 1 1 0 0	13.56	12.22	15.20	16.81
14	2 1 1 1 1 1 1 0	10.83	9.65	11.66	13.27
15	1 1 1 1 1 1 1 1	5.90	5.54	7.96	9.18
Opt.		18.83	15.00	30.85	34.52

Table III. Expected and observed cumulative selection response, R_E and R_o, and inbreeding coefficient (%), F_E and F_o, after 5 generations of selection, as a function of family size. The initial additive variance was 50, $h^2 = 0.30$.

Standard errors ranged from 0.17 to 0.27 (R_o), and from 0.16 to 0.43 (F_o).

Case	R_E	R_o	F_E	F_o
1	24.48	18.85	42.76	40.61
2	26.17	20.17	35.81	35.06
3	25.93	20.77	33.26	32.75
4	26.03	20.38	30.59	31.77
5	25.35	20.53	30.59	30.79
6	25.45	20.84	27.80	28.48
7	24.29	19.95	27.80	29.31
8	25.21	20.42	24.87	26.26
9	24.49	20.66	21.79	24.19
10	22.27	18.86	21.79	21.74
11	22.71	19.21	18.57	20.15
12	21.89	18.86	18.57	20.47
13	21.13	19.08	15.20	17.89
14	17.51	16.28	11.66	13.44
15	10.81	10.56	7.96	9.23
Opt.	27.64	21.09	28.71	31.92

Minimum coancestry matings

The observed genetic progress attained during the first 5 generations of selection, both with random, R_R, and minimum coancestry matings, R_{MC}, together with the corresponding inbreeding coefficients, F_R and F_{MC}, are shown in Table IV (λ_{op} was used). The selection response obtained was similar in both cases, as expected in a strictly additive model. However, minimum coancestry matings dramatically reduced inbreeding, compared with random mating. Nevertheless, it should be noted that this reduction was mainly due the one generation delay in the initial appearance of consanguinity.

Table IV. Observed cumulative selection response after 5 generations of selection with random mating, R_R, minimum coancestry mating, R_{MC}, and mate selection, R_{MS}, together with their respective inbreeding coefficient (%) F_R, F_{MC} and F_{MS}. The initial additive variance was 50.

Generation	R_R	R_{MC}	R_{MS}	F_R	F_{MC}	F_{MS}
$h^2 = 0.10$						
1	3.71	3.86	2.76	8.49	0.00	3.01
2	6.58	6.53	5.74	15.89	8.54	7.87
3	9.48	9.08	8.53	22.54	14.10	13.56
4	12.27	11.98	11.64	28.88	19.08	18.73
5	15.00	14.10	14.45	34.52	24.08	23.77
$h^2 = 0.30$						
1	5.26	5.65	4.96	7.87	0.00	3.03
2	9.54	9.64	9.90	14.44	7.21	8.00
3	14.51	16.68	14.38	21.01	12.48	13.19
4	17.43	17.79	18.49	26.43	17.03	17.81
5	21.09	21.34	22.31	31.92	21.78	22.92

Standard errors in the fifth generation ranged from 0.23 to 0.28 (R_R and R_{MC}), 0.80 (R_{MS}) and 0.40 (F_R, F_{MC} and F_{MS}).

Mates selection

Table IV shows the observed response, R_{MS}, and the inbreeding coefficient, F_{MS}. It can be seen that, while conforming with inbreeding restrictions, response was not smaller than that attained under the optimum unrestricted scheme, R_R.

Quinton, Smith, Goddard. 1992. Comparison of selection methods at the same level of inbreeding. J. Anim. Sci. 70: 1060.

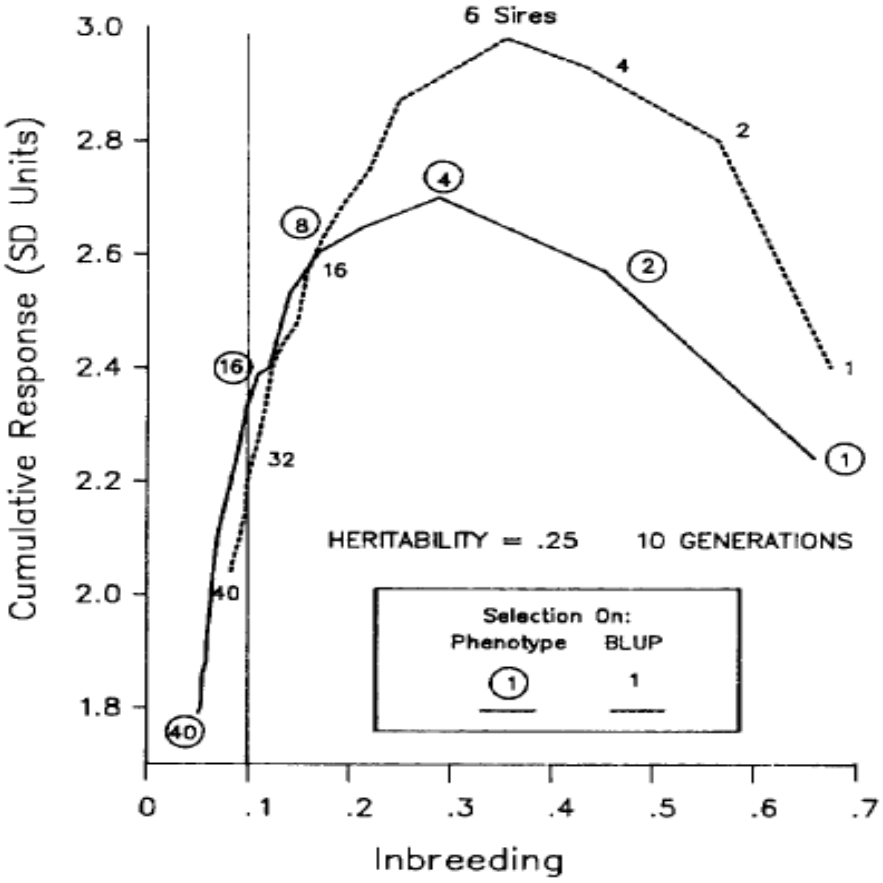


Figure 1. The average (100 replicates) simulated cumulative genetic response and inbreeding after 10 generations of phenotypic and best linear unbiased prediction (BLUP) selection for a range (1 to 40) in the number (S) of sires selected (heritability .25, S/100 males selected, 50/100 females selected). Numerical results are given in Table 2.

ABSTRACT: Animal geneticists predict higher genetic responses to selection by increasing the accuracy of selection using BLUP with information on relatives. Comparison of different selection methods is usually made with the same total number tested and with the same number of parents and mating structure so as to give some acceptable (low) level of inbreeding. Use of family information by BLUP results in the individuals selected being more closely related, and the levels of inbreeding are increased, thereby breaking the original restriction on inbreeding. An alternative is to compare methods at the same level of inbreeding. This would allow more intense selection (fewer males selected) with the less accurate methods. Stochastic simulation shows that, at the same level of inbreeding, differences between the methods are much smaller than if inbreeding is unrestricted. If low to moderate inbreeding levels are targeted, as in a closed line of limited size, then selection on phenotype can yield higher genetic responses than selection on BLUP. Extra responses by BLUP are at the expense of extra inbreeding. The results derived here show that selection on BLUP of breeding values may not be optimal in all cases. Thus, current theory and teaching on selection methods are queried. Revision of the methodology and a reappraisal of the optimization results of selection theory are required.

Villaneuva and Woolliams (1997). Optimization of breeding programs under index selection and constrained inbreeding. Genet. Res. Camb. 69:145

Objective = maximize response (over planning horizon) with constraint on ΔF

Parameters to optimize:

- Population size
- # sires and dams to select
- Selection criterion to use (emphasis on family info)
- Mating strategy

Population size = 200

Maximize average response from 5 - 20 generations by optimizing

- # sires selected
- # dams/sire
- weight on family vs. own performance

Table 1. *An example of the maximization procedure for $N = 200$, $h^2_{(0)} = 0{\cdot}3$ and $\Phi_{(5,20)} = \Delta\bar{G}_{(5,20)} - \lambda\Delta F$. Hence, for a restriction of $\Delta F \leqslant 1\%$, the scheme for $\lambda = 7{\cdot}4$* would be expected to give the greatest value of $\Delta\bar{G}_{(5,20)}$ *by using 30 sires (N_s) with a mating ratio (d) of 1 and a relative weight ($b_2 = b_3$) of 1·04 for the family means*

λ	$\Phi_{(5,20)}$	$\Delta\bar{G}_{(5,20)}$	ΔF	N_s	d	$b_2 = b_3$
0·0	0·322	0·322	0·03179	16	1	1·63
1·0	0·295	0·318	0·02336	19	1	1·47
2·0	0·274	0·312	0·01910	21	1	1·33
3·0	0·256	0·304	0·01612	23	1	1·25
7·3	0·201	0·276	0·01020	29	1	1·01
7·4	0·200	0·273	0·00986	30	1	1·04
55·6	−0·009	0·132	0·00253	67	1	0·74
55·7	−0·009	0·130	0·00249	68	1	0·76

	h^2=0.1			**h^2=0.3**		
	Constraint on Inbreeding			**Constraint on Inbreeding**		
	None	**ΔF≤1%**	**ΔF≤0.25%**	**None**	**ΔF≤1%**	**ΔF≤0.25%**
ΔF/generation	**2.09**	**1.00**	**0.25**	**2.00**	**1.00**	**0.25**
ΔG in generation 20	**0.109**	**0.100**	**0.047**	**0.278**	**0.258**	**0.128**
# sires	**22**	**32**	**69**	**21**	**30**	**68**
# dams/sire	**1**	**1**	**1**	**1**	**1**	**1**
Relative weight on family info	**2.12**	**1.60**	**1.07**	**1.43**	**1.06**	**0.76**
Optimal weight based on sel. index	**9.6**					

Optimal Contribution Selection

Selection while Controlling Inbreeding in Operational Programs

Based on Meuwissen (1997). J. Anim. Sci. 75: 934.

Meuwissen (1997) developed a method to directly control long-term ΔF while maximizing ΔG by formulating selection as a constrained maximization program:

$$\text{Max } \bar{g}_{t+1} = \mathbf{c}_t' \hat{\mathbf{g}}_t \qquad \text{Subject to } \mathbf{Q}'\mathbf{c}_t = \frac{\mathbf{1}}{\mathbf{2}}$$

$$\tfrac{1}{2}\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t = \bar{C}_{t+1}$$

$\bar{g}_{t+1}$ = mean BV in the next generation

$\hat{\mathbf{g}}_t$ = vector of BLUP EBV of selection candidates in generation t

$\mathbf{c}_t$ = vector of contributions of selection candidates to the next generation

$\mathbf{Q}$ = known incidence matrix for sex (the first column contains 1 for male candidates and the second column a 1 for female candidates)

$\frac{\mathbf{1}}{\mathbf{2}} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ which ensures that contributions of males and of all females sum to ½

$\mathbf{A}_t$ = additive genetic relationship among selection candidates in generation t.

$\bar{C}_{t+1}$ = average coancestry among all progeny in generation t+1

= ½ weighted average genetic relationships among selected parents = $\tfrac{1}{2}\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t$.

= set equal to ΔF(t+1) when objective is to restrict rate of inbreeding per generation to ΔF and generation 0 is non-inbred

⇨ Maximize $H_t = \mathbf{c}_t'\hat{g}_t - \lambda_0\left(\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t - 2\bar{\mathbf{C}}_{t+1}\right) - \left(\mathbf{c}_t'\mathbf{Q} - \tfrac{1}{2}'\right)\lambda_v$ for $\mathbf{c}_t$, λ_0, and λ

λ_0, and λ are LaGrangian multipliers, and $\boldsymbol{\lambda}_v' = [\lambda_0, \lambda]$.

Solving this system for $\mathbf{c}_t$ results in:

$$\mathbf{c}_t = \frac{\mathbf{A}_t^{-1}\left(\hat{g}_t - \mathbf{Q}\lambda_v\right)}{2\lambda_0}$$

In order to obtain $\mathbf{c}_t$, values for λ_0 and λ are needed.

Constraint $\mathbf{Q}'\mathbf{c}_t = ½$ → $\mathbf{Q}'\mathbf{A}_t^{-1}\mathbf{Q}\lambda_v = \mathbf{Q}'\mathbf{A}_t^{-1}\hat{g}_t - \mathbf{1}\lambda_0$

Constraint $\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t/2 = \bar{C}_{t+1}$ → $8\bar{C}_{t+1}\lambda_0^2 = \mathbf{Q}'\mathbf{A}_t^{-1}\hat{g}_t - \mathbf{1}\lambda_0$

Solving for λ_0 → $\lambda_0^2 = \dfrac{\hat{g}_t'\left(\mathbf{A}_t^{-1} - \mathbf{A}_t^{-1}\mathbf{Q}\left(\mathbf{Q}'\mathbf{A}_t^{-1}\mathbf{Q}\right)^{-1}\mathbf{Q}'\mathbf{A}_t^{-1}\right)\hat{g}_t}{8\bar{C}_{t+1} - \mathbf{1}'\left(\mathbf{Q}'\mathbf{A}_t^{-1}\mathbf{Q}\right)^{-1}\mathbf{1}}$

The value for λ_0 is used in the previous equation to obtain the value for λ. These two values can now be used to obtain $\mathbf{c}_t$. This $\mathbf{c}_t$ may contain negative values for some animals with a poor EBV. Negative values of $\mathbf{c}_t$ can be constrained to 0 by eliminating those animals.

A negative right hand sight for the last equation implies that $\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t/2 = \bar{C}_{t+1}$ cannot be met. So it is impossible to find a solution for $\mathbf{c}_t$ for which the average coancestry between parents is less or equal to the desired level. The minimum average relationship that can be obtained by minimizing $\mathbf{c}_t'\mathbf{A}_t\mathbf{c}_t$ under the constraint $\mathbf{Q}'\mathbf{c}_t = ½$. This leads to the following minimum: $.25\mathbf{1}'(\mathbf{Q}'\mathbf{A}_t^{-1}\mathbf{Q})^{-1}\mathbf{1}$.

Table 1. The average coancestry of the parents of generation t, the inbreeding in generation t, and the genetic gain from generation t – 1 to t, when the average coancestry was limited to .025 (t – 1) and genetic contributions were optimized within each generation for both sexes[a]

Generation (t)	Coancestry parents	Inbreeding	Genetic gain
2	.025	0	.380
3	.050	.029	.322
4	.075	.052	.293
5	.100	.076	.318
6	.125	.100	.287
7	.150	.127	.303
8	.175	.150	.301
9	.200	.175	.311
10	.225	.202	.315

[a]Average of 100 replicated simulations of the breeding scheme. The standard errors of the inbreeding and genetic gain were approximately .0011 and .011, respectively. The coancestry did not vary, because it was constrained.

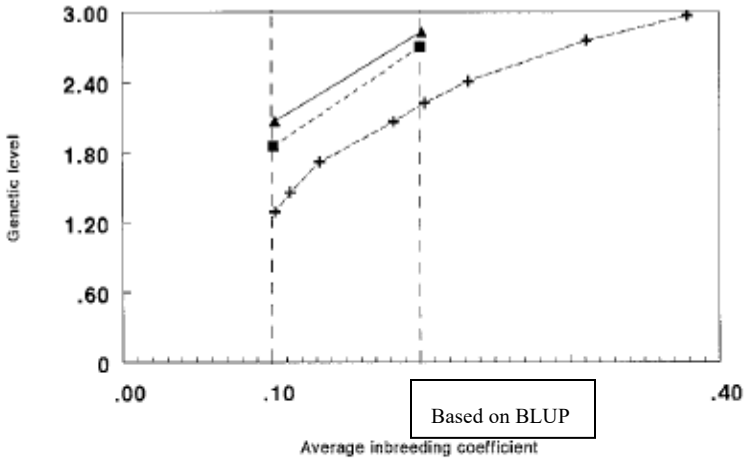


Figure 1. Genetic and inbreeding levels in generation 10 with inbreeding levels constrained to .1 and .2 and with optimal genetic contributions (▲), optimal selection of sires and dams but with equal contributions of selected sires and selected dams (■), and BLUP selection with selection of (from left to right) 32, 30, 26, 20, 18, 16, 12, and 10 sires and equal numbers of dams (+).

Table 2. Optimal numbers of parents selected when the number of offspring per sire and per dam are equal. For comparison, the numbers according to Wright's (1931) random mating formula are given

	ΔF_{init} = .025[a]		ΔF_{init} = .0125[a]	
Generation	Sires	Dams	Sires	Dams
2	10	10	21	20
3	10	10	19	19
4	10	10	18	19
5	9	9	18	18
6	9	9	18	17
7	9	9	17	17
8	8	8	17	17
9	8	8	16	17
10	8	8	17	16
Wright's formula	10	10	20	20[b]

[a]ΔF_{init} = the initial rates of inbreeding; in later generations ΔF increased to .0303 and .0137, respectively.

[b]Wright's formula is $\Delta F = 1/8n_s + 1/8n_d$, where the numbers of males (n_s) and females (n_d) are assumed equal.

Optimal contribution selection was extended to overlapping generations
by Meuwissen and Sonesson (1998). J. Anim. Sci. 76: 2575

Table 1. Parameters of the closed nucleus breeding schemes

Constraint on inbreeding	.5 or .25%/yr
No. of new progeny per yr (males and females)	256 or 512
Size of unrelated base population	5 × (No. of new progeny)
No. of years evaluated	20
Involuntary culling rate of males and females	.3
Voluntary culling rate	negligible
Age at which females completed lactation records	2, 3, and 4 yr[a]
No. of test daughters of bulls with unrelated cows outside the nucleus	0 or 100
Age at which progeny test became available	5 yr[a]
Reproductive rate of males and females within nucleus	unlimited
No. of sires and dams selected in BLUP selection schemes	equal and such that inbreeding constraint holds
Genetic, and permanent and temporary environmental variances	.3, .2, .5

[a]When the animals are selected for this information, the offspring are born 1 yr later (i.e., the generation interval is 1 yr longer than the age at which the information becomes available).

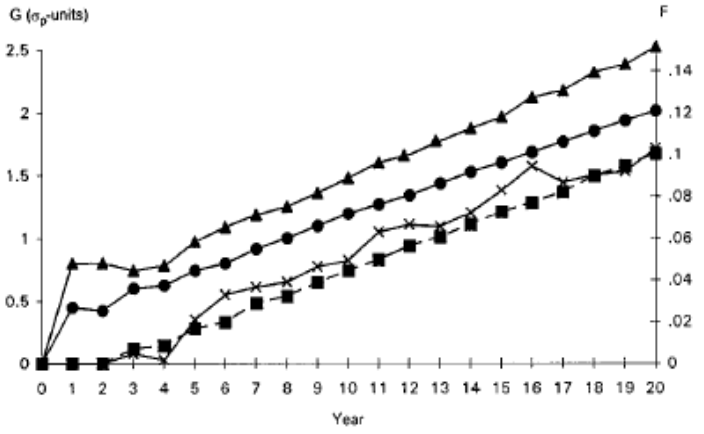


Figure 1. Genetic level (G) and inbreeding coefficient (F) for optimal contribution selection (OC) and BLUP-EBV selection of 64 sires and 64 dams. Averages of 50 simulations with 256 new progeny per year without progeny test of young bulls (G-OC = ▲; G-BLUP = ●; F-OC = ×; F-BLUP = ■).

Table 2. Genetic level (G) and inbreeding coefficients (F) at yr 20 when nucleus herds were selected with the optimal contribution method and with selection for BLUP-EBV[a]

New progeny per yr, no. of animals	Size of progeny test, no. of records	Optimal Contribution G, σ_p-units	Optimal Contribution F	BLUP G, σ_p-units	BLUP F
	Constraint on ΔF per year = .005				
256	0	2.52	.10	2.01	.10
	100	3.12	.08	2.30	.09
512	0	2.83	.11	2.43	.11
	100	3.46	.09	2.73	.09
	Constraint on ΔF per year = .0025				
256	0	2.24	.05	1.63	.05
	100	2.83	.04	1.97	.05
512	0	2.65	.05	2.05	.05
	100	3.27	.04	2.35	.05

Table 3. Number of animals selected and generation intervals at yr 20 with the optimal contribution and BLUP selection, where numbers selected were chosen to achieve the inbreeding constraint[a]

New progeny per yr, no. of animals	Size of progeny test, no. of records	Optimal Contribution Selected sires/dams, no. of animals	Optimal Contribution Gen. Interval sires/dams, yr	BLUP Selected sires/dams, no. of animals	BLUP Gen. Interval sires/dams, yr
	Constraint on ΔF per year = .005				
256	0	19.2/5.3	2.7/4.7	64/64	2.7/3.3
	100	2.8/4.5	6.2/4.5	64/64	3.2/3.1
512	0	24.6/5.3	2.9/4.5	80/80	2.5/3.2
	100	2.7/5.8	6.0/4.4	80/80	3.1/3.0
	Constraint on ΔF per year = .0025				
256	0	34.9/11.4	2.9/4.5	105/105	2.8/3.3
	100	7.6/10.1	6.0/4.4	95/95	3.1/3.1
512	0	45.2/13.2	2.6/4.5	130/130	2.6/3.2
	100	6.4/9.7	6.0/4.5	125/125	3.1/3.1