ACCOUNTING FOR SCALE EFFECTS IN GENETIC EVALUATION OF NEW ZEALAND DAIRY CATTLE

A thesis presented in partial fulfilment
of the requirements for the degree
of Master of Agricultural Science
in Animal Science
at
Massey University
Palmerston North
New Zealand

NICOLAS LOPEZ-VILLALOBOS
1994

Massey University Library
Thesis Copyright Form

Title of thesis: ACCOUNTING FOR SCALE EFFECTS IN GENETIC EVALUATION OF NEW ZEALAND DAIRY CATTLE.

(1) ✓(a) I give permission for my thesis to be made available to readers in Massey University Library under conditions determined by the Librarian.

(b) I do not wish my thesis to be made available to readers without my written consent for ... months.

(2) ✓(a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.

(b) I do not wish my thesis, or a copy, to be sent to another institution without my written consent for ... months.

(3) ✓(a) I agree that my thesis may be copied for Library use.

(b) I do not wish my thesis to be copied for Library use for ... months.

Signed [signature] Lopez-Villalobos

Date 21 Feb 1994

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME AND ADDRESS DATE

ACKNOWLEDGEMENTS

The author gratefully acknowledges the guidance and assistance, above and beyond the call of duty, provided by his supervisors, Dr. Dorian J. Garrick and Professor Hugh T. Blair throughout his two years study at Massey University. Dr. Garrick has contributed greatly to the author's understanding of linear models and theoretical aspects of genetic evaluation of dairy cattle. Professor Blair has increased the author's interest in genetic improvement of farm animals. They have given to the author the sense of working to achieve goals even though those seemed to be unattainable.

The author acknowledges the friendly environment provided by Professor Robert D. Anderson, Dean of the Faculty of Agricultural and Horticultural Science and Professor Stuart N. McCutcheon, Head of the Department of Animal Science.

The author acknowledges the assistance of Livestock Improvement Corporation of the New Zealand Dairy Board in providing the data for this research. The author would like to acknowledge the analyses provided by Dr. Bevin Harris and the useful discussions concerning theoretical and practical aspects of this research.

Acknowledgement is given to Professors C. W. Holmes and D. D. S. Mackenzie. They have contributed to the author's knowledge concerning milk production from pastures and the physiological basis of milk production, respectively.

Thanks are due to Mr. R. G. Sherlock, and Dr. S. Peterson from whom the author received technical help in computer problems.

Thanks are given to those people that helped the author in learning English. Special thanks are given to Professor D. Crabbe of the English Language Institute, Victoria University.

The author appreciates the friendly environment provided by his fellow graduate students in the Department of Animal Science. These students made the two years an enjoyable time.

The author would like to acknowledge the financial support provided by the Ministry of Foreign Affairs and Trade of New Zealand, El Consejo Nacional de Ciencia y Tecnología de México and La Universidad Autónoma Chapingo. The funding provided by these institutions were invaluable for the author's studies.

The author is grateful to his parents; Sr. Manuel López-Avila and Sra. Hermelinda Villalobos de López, his brothers; Manuel, J. Martín, Arturo, Ignacio and Sergio and his sisters; Hortencia, Maria del Rosario and Estela for their love and encouragement over the years.

Finally, the author would like to express special thanks to his wife, Silvia, for her understanding, support and encouragement throughout the study.

ABSTRACT

It is well known that variation in lactation yields tends to increase with average production. Failure to account for this scale effect may cause overestimation of genetic merit for sires with a majority of daughters in high-variation herds and vice-versa. The current system of sire evaluation in New Zealand overcomes this problem by expressing daughters performance as a proportion of contemporary average performance. The objectives of this study were to quantify the magnitude of scaling (heterogeneous variance), and to identify methods to stabilise the variance of milkfat yields for use in the genetic evaluation system of dairy cattle through best linear unbiased prediction (BLUP) using an animal model across breeds.

Lactation records of dairy cows calving between 1986 and 1989 were obtained from the Livestock Improvement Corporation of the New Zealand Dairy Board. There were milkfat yields from 2,004,854 lactations in 83,805 contemporary groups (herd-year-age; HYA). The data were divided into three equal-sized subsets based on HYA mean; these being (kg milkfat ± sd) High (H), 172 ± 28; Medium (M), 152 ± 26; and Low (L), 139 ± 25.

The methods investigated for the accounting of scaling were: adjustment by the HYA sd (SD-adjustment); scaling by the HYA mean (MEAN-correction); and natural logarithmic transformation (LOG-transformation) of milkfat yield. The overall correlation between HYA means and HYA sd's was 0.44. This value was reduced to 0.31 in SD-adjusted, -0.27 in the MEAN-corrected and -0.24 in the LOG-transformed data. Ideally, the transformed data should exhibit independence between the mean and standard deviation.

Breeding values of sires were separately estimated from each data subset using a mixed model. Product-moment and rank correlations between breeding values for sires estimated from the independent subsets and with variable minimum number of daughters were in the overall comparisons (L-M, L-H and M-H) lower than expected correlations, reflecting inaccuracies in sire evaluation when scaling is ignored. Product-moment and rank correlations were similar for SD-adjustment and MEAN-correction, but LOG-transformation reduced the calculated correlations in the L-M, L-H and M-H comparisons.

Estimates of the genetic correlations between production in pairs of environments were obtained from the ratio of observed to expected correlations. These estimates ranged from 0.82 to 1.01 for the linear yields. Estimates of genetic correlations were similar for SD-adjusted and MEAN-corrected data, but for LOG-transformed data these were reduced, especially in the L-H comparison which ranged from 0.77 to 0.87.

Results confirm the problem of scaling on genetic evaluation of New Zealand dairy cattle. MEAN-correction and LOG-transformation methods are not appropriate because they tend to overcorrect the scaling problem. SD-adjustment is not satisfactory but seems to be more appropriate than no adjustment. An alternative method is proposed based on a Bayesian approach, which takes into account any relationship between variance and mean.

Keywords: dairy cattle, BLUP, scaling effect.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1
INTRODUCTION

In a broad sense, factors determining milkfat yield of a dairy cow are classified into two groups. The first group is the inherited ability of the cow to produce milkfat, of which one half comes from her sire and the other half comes from her dam. The second group is the set of environmental factors which will affect the cow's level of production, but which are quite independent of her inherited ability. Generally, this last group of factors is comprised in a herd-year-age group called a contemporary group (Van Vleck et al. 1987). Thus, cows of the same age daughters of a dairy bull located in different environments (herd-years) will have different yields due to the effect of specific factors in each of the environments. It is assumed that genetic and residual variances across contemporary groups are equal. Several studies, however, have indicated a positive relationship between production level and estimates of genetic and residual variances and heritability. Failure to account for this scale effect may cause misranking in the genetic evaluation of animals.

The possibility of this scaling problem in dairy production has been studied in New Zealand since the progeny testing of bulls was established (see for example, Stichbury, 1957). A matter of discussion has been if breeding values of dairy sires will rank in the same order based on daughters milked over a wide range of environmental conditions. In that time (1957), 214 dairy bulls were surveyed in low and high producing herds. In both cases it was showed that a bull's daughter average varied according to the level of the herd in which it was used, but the difference from expectancy figure, in general, remained constant.

Similarly, Wickham (unpubl.) cited by Holmes and Wilson (1984) presented data of the productive performance of the progeny of one dairy sire used in the artificial insemination for commercial herds. The differences in milkfat yield between the daughters of the bull and their contemporaries at three levels of production increased as the level of production increased. However, the rank of the sire in the different producing levels was not studied.

One of the features of the present New Zealand system for sire evaluation is the use of ratios of the daughter average to contemporaries average rather than deviations of daughter average from contemporaries average. The ratio breaks a possible relationship between the mean and the variance and adjusts for unequal error variances (Everett and

Jones, 1985). This was the case of the dairy bull above mentioned. The proportion of the average milkfat production of the sire's daughters and the average of contemporaries remained constant as the level of production increased.

In the genetic evaluation of dairy cattle, therefore, there are two practical situations that merit attention. First, if for a given trait the variance increases as the mean increases, animals to be parents of the next generation will be selected in a greater proportion from those environments with high levels of production (higher variability) than from environments with low levels of production (lower variability). This is more important when selecting cows than when selecting bulls because sires are generally evaluated in several herds with variable levels of production. Second, if heterogeneity of variance across environments is not taken into account for the genetic evaluation of the animals, the ranking of animals on the basis of estimated breeding values would change across environments.

In 1991 a technical review of the current sire evaluation procedures was initiated and a prototype animal model based on best linear unbiased prediction (BLUP) procedures had been developed, which will be adopted in 1995 (Harris, et al. 1993). The prototype animal model can readily account for heterogenous residual and genetic variance if they were found.

In regard to the above questions, this thesis had the following objectives:

1. To provide evidence of heterogeneity of variance for milkfat production measured through the relationship between the mean and the variance.

2. To determine the effect of heterogeneity of variance on the ranking of sires on the basis of breeding values estimated at different production levels through a single repeatability animal model across breeds using BLUP procedures.

3. To identify the best of three methods to stabilise the variance of milkfat yields.

CHAPTER 2
DAIRY FARMING IN NEW ZEALAND

New Zealand claims to be the most efficient country in the world in growing and converting pasture, through the grazing animal, into meat, milk, fibre and by products for human use. The basic elements of this efficiency are temperate climate favouring all-year grass growth, a high level of education among farmers, applied research and the rapid adoption of advanced technology.

2.1 Climatic Conditions

The country is mostly hilly and mountainous with relatively little easily-cultivable land. It stretches through more than 12° of latitude from the subtropical north to the colder southern regions; within this altitude and latitude range there is very wide variation in climate (Coop, 1987).

Rainfall distribution is controlled mostly by the mountains. The west coast of the South Island and the mountain regions of the North Island receives over 2,400 mm rainfall/annum. The northern and western regions of the North Island receives 1,000-2,000 mm, which together with mild winters is very favourable for pasture growth, and it is here that dairy farming predominates. The east cost of both islands have a rainfall of 500-1,000 mm, which is also good for pasture production, but the rainfall is less reliable and the winters are colder. These regions are more suitable for sheep farming than dairy farming. Mean temperature during January arises to 19°C and the mean temperature in July, in some regions, lowers to 2°C (Coop, 1987).

2.2 Dairy Industry Structure

The New Zealand Dairy industry is export orientated. Only 10% of the milk produced is consumed within the country, the remaining 90% being processed for export as butter, cheese, dried milk powders, and other products. Dairy production is a highly organised industry owned and controlled by the producers in a largely cooperative movement. The factory-supply farmers supply milk to 33 manufacturing cooperative dairy companies, the directors of which are elected by the local farmers. The directors of the companies elect the directors of the New Zealand Dairy Board, which is the main

organisation responsible of marketing and dictating new policies in all aspects of the dairy industry (Coop, 1987).

2.2.1 Milk Production Systems

Two systems of dairy farming are practiced in New Zealand:

2.2.1.1 *Town Supply Dairy Farms*

Cows calve in spring and in autumn, or throughout the year. For the season 1990/91 were reported 177,100 cows in 1265 herds (Livestock Improvement, 1991a). The farmer is paid per litre of milk supplied (provided that the milk satisfies certain minimum composition standards, namely 3.25 percent milkfat, 8.50 percent solids non fat). Milk is used for consumption without significant processing. Even in this situation, pasture provides most of the lactating cow's feed requirements (Holmes, 1986).

2.2.1.2 *Seasonal or Factory Supply Dairy Farms*

Cows calve in spring time (July to September), and they are dried off during the winter (May to June) (Holmes and Wilson, 1984; Wickham, 1993). For the season 1990/91 were reported 2,225,045 cows distributed in 13,420 herds. The farmer receives payment based on quantity of milkfat and protein with a penalty for milk volume (Livestock Improvement, 1991a).

2.2.2 Operating Structures

The main operating structures found on New Zealand dairy farms are owner-operators, sharemilkers and contract milkers. Owner-operators are considered to be farmers who either own and operate their own farms or else employ a manager to operate the farm, for a fixed wage. They receive all the farm receipts, although they may then have to pay wages. Owner-operators comprise the largest operating group, accounting for 68% of the farms in 1990/91 (Livestock Improvement, 1991a).

Sharemilking has traditionally been the first step by which a young person can eventually accumulate enough capital to buy a farm of his own. Sharemilking involves operating a farm on behalf of the farm owner, for an agreed share of the farm receipts (as opposed to a set wage). Three types of sharemilking are commonly used: 29%, 39% and 50% agreements (Livestock Improvement, 1991a)

2.2.3 Herd Production Statistics

In the period of early settlement over 120-150 years ago, dairy cattle formed the most important livestock industry, providing subsistence for the local market. The cattle were mostly Shorthorns (Coop, 1987). In 1932 were recorded 1,292,873 cows, of which two thirds were Jersey or Jersey grades (New Zealand Jersey Cattle Breed Association, 1932). This breed was dominant until the 1960's when the Holstein-Friesian began to offer serious competition because of its better beef characteristics and higher volume of milk (Coop, 1987).

Nowadays, Holstein-Friesian, Jersey, Ayrshire and Holstein-Friesian/Jersey crossbred are the dominant genetic groups. Other breeds of dairy cattle present in smaller numbers in New Zealand include Milking Shorthorn, Guernsey and Brown Swiss (Livestock Improvement, 1991a).

An analysis of the trends in the New Zealand dairy industry indicates improvement due mainly to genetic gain and improvements in farm management. For the season 1980/81 the herd size has 132.7 cows, each producing 144 kg milkfat at a stocking rate of 2.1 cows per hectare, so that milkfat per hectare was 310 kg milkfat (Livestock Improvement, 1991a). For the season 1990/91 herd size increased to 165.8 cows and stocking rate increased to 2.4 cows per hectare. However milkfat per cow increased to 148 kg and milkfat per hectare also increased to 351 kg (Livestock Improvement, 1991a). For the New Zealand dairy farm, an indicator of economic efficiency is milkfat produced per hectare instead of milkfat produced per cow.

2.3 Pasture Production

The climatic conditions and freedraining soils allow a milk production system based almost entirely on pastures of perennial ryegrass (*Lolium perenne*) and white

clover (*Trifolium repens*). Other grass species such as yorkshire fog (*Holcus lanatus*), timothy (*Phleum platense*), paspalum (*Paspalum dilatatum*), and kikuyo (*Penisetum clandestinum*) may also be present (Coop, 1987).

For New Zealand conditions, Cooper (1970) based on the conversion rate of light to pasture growth, estimated a potential herbage production of 37 t DM/ha for rye grass-white clover pastures. Under cutting conditions, pastures provided with nutrients and water, herbage yields have been estimated in 24.7 t DM/ha (Brougham, 1959). Under grazing conditions, however, pastures top dressed with superphosphate, in about 300-600 kg/ha and receiving no nitrogenous fertiliser, produce only about 12 to 16 t DM/ha in New Zealand (Holmes, 1982; Bryant et al. 1982; Radcliffe and Baars, 1987). Radcliffe and Baars (1987) reported the seasonal pattern of perennial ryegrass/clover pastures. Of the total pasture growth, 39% occurs in spring (September to November), 26% in summer (December to February), 21% in autumn (March to May), and 14% in winter (June to July).

Although the climate permits some winter pasture growth, winter feed deficits occur and farmers supplement the cows' diet with silage and hay, which are harvested during periods of surplus, or deferred pasture.

2.4 Calving Date

A seasonally concentrated calving pattern is recognised as a feature of dairy farming in New Zealand. According to MacMillan et al. (1984) this pattern is part of a system which involves the maximum utilisation of pasture dry matter *in situ*; with limited conservation of pasture as hay or silage; very little cropping, except as part of a developed program; and almost no use of high energy or protein supplements. Ideally, calving is planned to commence in late winter, with a large proportion of the herd calving during the first four weeks, and the remainder over the next 6 to 12 weeks. This means that calving should be completed by the time the herd will have reached its peak in demand for pasture dry matter, and the seasonal flush in pasture growth will have commenced to meet this demand (MacMillan et al. 1984).

Herds which calve earlier than the "optimum" period are likely to be underfed in early lactation, but to have longer lactations. Herds which calve later than the "optimum" are likely to be well fed in early lactation, but to have shorter lactations

(Holmes, 1986). The terms "early" and "late", although described in relation to calendar dates, are related to herd requirements and pasture growth rates in spring (Holmes and MacMillan, 1982).

Management practices are used to avoid underfeeding in early lactation when calving date is too early. One alternative is increasing pasture growth through the strategic application of nitrogen fertiliser. Dairy farms with high stocking rate apply between 10 and 50 kg nitrogen/ha during September and October (Bryant, 1983; Roberts and Thompson, 1989; Thompson et al. 1991). Another practices such as grazing management (Bryant, 1990), supplementation (Holmes and MacMillan, 1982) and irrigation (Hutton, 1978) would make less severe the period of underfeeding and in this way achieving higher milk production.

2.5 Stocking Rate

Stocking rate is defined as the number of animals per unit of area of land. It has been recognised as one powerful management tool in the grazing system (McMeekan, 1956) determining milkfat yield per hectare (Holmes and Parker, 1992). Very low and very high stocking rates can depress herbage production because of the effects of under grazing and overgrazing on rates of photosynthesis and rates of senescence (Hodgson, 1990). As stocking rate increases, degree of defoliation increases, and a consequence, there is low photosynthetic activity caused by a reduction in leaf area. As a result of this, there is a reduction in pasture growth (Stockdale and King, 1980). On the contrary, the rate of herbage growth increases as stocking rate is reduced, but this effect is eventually offset by increasing losses to senescence so that net herbage production reaches a plateau and eventually starts to decline again at low stocking rate.

High stocking rate can also cause changes in pasture composition, such as a decreased proportion of dead material (Hodgson, 1990), an increased concentration of crude protein (Holmes, 1987), an increased digestibility (Stockdale and King, 1980; Hodgson, 1990), and an increased proportion of clover and decreased proportion of erect grasses (such as cocksfoot, *Dactylis glomerata*); these changes generally increase the feeding value of the herbage (Holmes and MacMillan, 1982; White, 1987).

The direct effect of stocking rate on animal production is affecting herbage utilisation per hectare. Even though higher stocking rate decreases pasture intake per

animal, pasture intake per hectare is increased as well as pasture utilisation, and therefore more animal product per hectare is achieved (King and Stockdale, 1980). Hence, increased animal production per hectare is achieved through increasing pasture utilisation (Bryant, 1980).

Mathematical models for the relationship between animal output and stocking rate have been suggested. Among the more cited in the literature are those proposed by Mott (1960); Peterson, Lucas and Mott (1965); Conniffe, Browne and Walshe (1970); and Jones and Sandland (1974). A debate arises about the quantitative relationship between performance per animal with stocking rate. As a result, it is difficult to define biological and economic optima for stocking rate. Nevertheless, in New Zealand, linearity has been assumed to study the effect of stocking rate on milk production (Holmes, 1980; Holmes and MacMillan, 1982; Ahlborn and Bryant, 1992).

Data obtained in New Zealand indicate that an increase of one cow per hectare (over the range 2 to 5 cows/ha) caused a decrease of 18 kg milkfat produced per cow, but an increase of 70 kg milkfat produced per hectare (Holmes and MacMillan, 1982). The biological and economic optima stocking rate were estimated at 5.4 cows/ha and 3.75 cows/ha, respectively (Wright and Pringle, 1983), and these values were higher than the actual average of 2.1 cows/ha (Holmes and Parker, 1992).

2.6 Grazing Management

Different systems of grazing management can influence both the amount of pasture eaten per cow daily, and the way in which the pasture is grazed. These can, in turn, influence the subsequent production of the animals and of the pastures (Holmes, 1980). Grazing management is, therefore, an important factor affecting the productivity of a dairy farm.

The New Zealand dairy farm is managed under a rotational grazing systems. The classical experiments with dairy cattle carried out by McMeekan (1956) and McMeekan and Walshe (1963) showed that at high stocking rate, milk production per hectare was strongly in favour of rotational grazing rather than set stocking grazing.

One of the reason why rotational grazing system has been used in the New Zealand dairy farm is that, through it, the matching of the pasture growth and feed

requirement curves has been encouraged. Further, rotational grazing allows that available herbage can be utilised more effectively in those periods of short herbage growth rate (Campbell, 1966).

The whole milking herd is shifted to one fresh paddock each day and amount of dry matter assigned to each animal depends, among other factors, on the requirement of the animal. The herd is rotated round the paddocks available on a 15-20 day rotation in spring and early summer, but this lengthens later to 30 days in summer, to 60 days in late autumn and 100 days in winter (Coop, 1987; Holmes, 1987; Holmes and Wilson, 1984).

2.7 Genetic Improvement

General objective of the New Zealand breeding program is to breed cattle which provide the highest economic returns (Wickham, 1993). Therefore, breeding objectives need consider both income (like milk, meat, or surplus stock), and expenditure (like feed, fertiliser, labour, repairs and interest on capital). The challenge for breeders is to select and breed cattle which incur greater profits. Traits of importance to farm profit are: milk production and its components, survival and related traits, labour costs, and beef characteristics of calves and slaughter cows (Holmes and Wilson, 1984).

Optimal genetic improvement demands an industry structure and operational environment in which breeding decisions are made on the basis of appropriate information and are consistent with objectives derived from long-term breeding programs.

2.7.1 Herd testing

Herd testing in New Zealand began as early as 1909 and its use was fostered by the then Department of Agriculture. The practice grew and group herd testing was established through the country by the mid 1920s under the name of Dominion Group Herd testing Federation Inc. In principle this system of testing was operated by and for commercial herd owners. Additional systems of testing, certificate of record, and official herd test were administered for pedigree herds by the Department of Agriculture. In 1936 some 28 cooperative herd testing associations were licensed and in 1939, the then 27 licensed herd testing associations were amalgamated to form 6 Herd Improvement

Associations under the name of New Zealand Dairy Board. In 1988 until nowadays, the official organisation responsible for herd recording is the Livestock Improvement Corporation, a wholly owned subsidiary of the New Zealand Dairy Board (Macdonald Committee Report, 1992).

The objective of herd testing is to obtain data on milk, milkfat and protein which, following suitable analyses, enables objective comparison to be made between individual cows, both on a within- and between-herd basis. For the season 1980/81 42.2% of the herds and 44.8% of the cows were under the herd testing system. These numbers increased to 62.4% of the herds and 68.2% of the cows for the season 1990/91 (Livestock Improvement, 1991a).

2.7.2 National Database

Samples of milk from individual cows are collected regularly by sampling officers or by the farmer (self-sampling) and processed through central laboratories providing information on milk yield, protein, milkfat, and somatic cell counts. Sampling is based on measured yields obtained over a 24 hour period, and samples are collected from consecutive evening and morning milking. The frequency of testing may be, either four, six or eight weekly (providing up to 12, 8 or 6 test during the season respectively), or else either two or three test total for the season (called double and triple testing).

Additional to the productive performance of the animals, there is recorded on the individual file for cows, animal events such as calving dates, mating dates and drying off dates. This information has allowed to the New Zealand Dairy Board the formation of a National Database, which is used for a various purposes. Some examples include:

cow: ancestry and performance data,
herd: management reports,
sire: progeny test information,
breed: Industry statistics.

(Macdonald Committee Report, 1992).

2.7.3 Artificial Insemination

Early investigations and experiments concerning artificial insemination in New Zealand started about 1939. In the season 1943/44 experimental work of artificially inseminating 1000 cows was carried out. The results showed that 80% of the cows were pregnant after a mating season of 3 months (Dalton and Rumble, 1985).

As a consequence of the seasonal spring calving, mating period is also seasonal. Therefore, demand of semen is concentrated for a period of 3 months and fresh semen is used rather than frozen semen. Using fresh semen has allowed reducing the semen concentration in each straw and so the top bulls can be used more intensively. The average number of inseminations per bull per year has steadily increased from 34,200 in 1966 to 65,770 in 1983 (Ahlborn-Breier et al. 1987) and for the season 1990/91, using a sperm concentration of 1 million per straw, over 150,000 inseminations per bull were achieved (Vishwanath, 1992).

The Livestock Improvement Corporation offers two main artificial breeding services. "Premier Sires" in which most semen used is in liquid form (i.e., fresh) and allows greater utilisation of bulls. Farmers have the option of inseminating cows themselves or having an artificial breeding technician to do the inseminations. "Nominated Service" gives the farmers the opportunity of choosing individual bulls to be used. This service uses frozen semen. Many of the bulls in the Premier Sire Service are also available through the Nominated Service (Livestock Improvement, 1991b).

Deep frozen semen is also supplied by other companies such as Ambreed NZ Ltd, Wrightson Breeding Services Ltd, which are also involved in the extensive progeny testing. Another other private companies supply small amount of deep frozen semen imported from the USA, Canada and Australia under the regulatory conditions determined by the Ministry of Agriculture and Fisheries. Nowadays, artificial insemination is widely used and most of the dairy farmers breed their cattle through artificial insemination. For the season 1990/91, 1,858,966 cows were artificially inseminated (Livestock Improvement, 1991a).

2.7.4 Breeding Program

Rendel and Robertson (1950) indicated that a dairy cattle population can be genetically improved along any one or more of four separate pathways, namely:

(i) Bulls to breed bulls.
(ii) Bulls to breed cows.
(iii) Cows to breed bulls.
(iv) Cows to breed cows.

In New Zealand, the idea of the breeding plan is to use all four pathways in the best possible combination. A Sire Proving Scheme has been developed with two main aims, to accurately identify bulls of outstanding genetic merit, and to achieve widespread use of chosen sires (Macdonald Committee Report, 1992).

Approximately 155 bulls are progeny tested annually. Ayrshire, Holstein-Friesian and Jersey bulls calves from contract mating, along with others selected from outside the contract mating scheme but which meet the stringent selection criteria, are purchased to become part of each year's intake for the Livestock Improvement Sire Proving Scheme (Livestock Improvement, 1991b). Semen from the yearling Sire Proving Scheme bulls is used in about 70,000 cows in approximately 450 Sire Proving Scheme herds throughout New Zealand. The inseminations generate about 85 daughters for each of the Holstein-Friesian, 60 for each of the Jersey and 50 for each of the Ayrshire bull. Each bull's daughters are spread through 30 to 40 Sire Proving Scheme herds (Macdonald Committee Report, 1992).

Once the required inseminations have taken place, bulls are not used again until progeny test results for production and traits other than production are available. This information is used for selecting the best bulls for extensive use in the Premier Sires team and Nominated Semen Services of the Livestock Improvement Corporation. The best 5-10% of the bulls which entered the Sire Proving Scheme are finally selected for extensive use as bulls to breed cows. The best 50% of these young sires are used in contract mating as bulls to breed bulls.

From the cows under the herd testing system, only a small proportion satisfy the requirement to become dams to breed sires (active cows). The main requirement is that

those cows must have three generations of identified artificial breeding proven sires to one breed (at least seven-eighths purebred). The active cow population size has been small but is now increasing rapidly. For the season 1990/91 only 89,000 active cows were recorded and they are expected to increase to 217,000 for the season 1992/93 (Macdonald Committee Report, 1992).

2.8 Sire evaluation

The exceptionally high dilution rate of fresh semen as practised in New Zealand has allowed a very high selection differential on the bulls to breed bulls selection pathway (Cunningham, 1983). Due to this, major efforts have been put in developing mathematical procedures to estimate the genetic merit of dairy sires.

The main method for sire evaluation in dairy cattle is progeny testing. The value of a sire in a breeding program, is judged by the mean value of its progeny. If a sire is mated to a number of cows taken at random from the population then its breeding value is twice the mean deviation of the progeny from the population mean (Falconer, 1960). Therefore, the daughter average should usually enable the estimation of the breeding value with a high reliability. Unfortunately, in reality there are many fixed and random effects which influence the performance of the daughters and thus mask the breeding value of a sire (Dempfle, 1984). Some of the factors which influence the daughter performance are:

- breeding value of the daughter, which is made up of half of the breeding value of the sire and half of the breeding value of the dam,
- herd management,
- nutrition level during lactation,
- age at calving,
- days open,
- number of times milked per day,
- length of lactation.

In the derivation of a computing formulae for sire evaluation two steps are involved (Garrick, 1991). First, definition of the model to describe the processes that influence the phenotype. The model includes a model equation, relevant means and variance parameters and perhaps distributional assumptions. Second, definition of the

criterion for identifying a "good" predictor. For example, defining the prediction error as the difference between true genetic merit and the predict genetic merit, then the best predictor (BP) is that which minimises the variance of these errors.

2.8.1 Sire Evaluation up to 1969

A system of sire evaluation had been developed in New Zealand in the 1960s (Wickham, 1984), in which the result of a bull proof was defined as the genetic superiority expressed in pounds milkfat of the bull daughters over those of an "average" bull, if both bulls were mated to cows of a similar genetic level. Later refinements of this system were correction for number of daughters and lactations, refinements in age corrections, and number of herds in which the bull had been surveyed (Shannon, 1974). This system was based on the following statistical model:

$$y_{ijk} = a_i + s_{ij} + e_{ijk}$$

where

y_{ijk} = the age corrected record of the k^{th} daughter of the j^{th} sire in the i^{th} year,

a_i = genetic value of the stud in the i^{th} year,

s_{ij} = the effect of the j^{th} sire,

e_{ijk} = the random error unique to y_{ijk}.

Thus, the breeding value of the bull by definition was $a_i + s_{ij}$ which is a similar approach to the concept of genetic groups considered in sire evaluation through best linear unbiased prediction methods (Thompson, 1979; Quaas and Pollak, 1981; Pollak and Quaas, 1983).

A detailed description of the method was given by Searle (1964) when comparing the New Zealand method of sire evaluation with those of Great Britain and New York State. The procedure to estimate the breeding value of an artificial insemination sire was on the basis of the difference from expectancy. The genetic merit of an artificial

insemination sire was estimated using the following procedure:

Estimated sire merit = 2 rating + breed average

where

$$\text{Rating} = F[\overline{X} - (B + 0.9(\overline{S} - B)]$$
$$\text{Rating} = F[\overline{X} - 0.9\overline{S} - 0.1B]$$

and where

B = the breed average of all cows in the particular region,

$\overline{X}$ = the average production of the daughters of the sire being evaluated,

$\overline{S}$ = the overall mean of the herdmate averages to which the daughters of the sire were compared,

0.9 = the regression of daughter average on true herd average (effectively intra-sire regression of daughter average production on herdmate average),

F = the regression of the sire's future daughter production on his estimated true daughter average.

Anderson (1974) discussed this method of genetic evaluation in regard to the biases in the estimation. Age corrections offered an opportunity for increasing the accuracy of sire evaluation but also biases were found. The intra-sire regression of daughter production on herdmate production was an adjustment for the non-random usage of the sires being evaluated amongst herds of differing production levels. For New Zealand conditions, the value was estimated in 0.9. It implied that, on average, 20% of the between-herd difference in production was genetic in origin (Rae, 1971 as cited by Anderson, 1974). Thus, a possible source of bias in sire evaluation resulting from genetic differences between herds was overcame by the use of an intra-sire regression of daughter average on herdmate average. Subsequent studies (Brumby, 1961) suggested that the regression should be unity since in the case of milkfat yield, genetic differences between herds were not-existent.

The formula for obtaining the regression of a sire's future daughter production on his estimated true daughter average was

$$F = \frac{\frac{1}{4}nh^2}{[1 + (n-1)\frac{1}{4}h^2]}$$

where

h^2 = heritability for milk fat
n = number of daughters of the sire

It was assumed a value of heritability for milkfat of 0.25, so that a simplified formula was

$$F = \frac{n}{n + 15}.$$

Because the above formula is valid only if first lactations of the sire's daughters are used, possible biases existed.

2.8.2 Current Sire Evaluation

The New Zealand Dairy Board continued to use the method of sire evaluation as above outlined for a number of years before introducing major changes in 1970 (N. Z. Dairy Board, 1970) and some modifications have been occurring during the intervening years (Garrick et al. 1993). Currently, the measure of the genetic quality of an animal is the breeding index. The final estimate of the breeding index of a bull is calculated from information about the breeding indexes of its parent and the results of its progeny test. Breeding indices are calculated for milk, milkfat, milk protein and traits other than production. A total index has been developed to display, in one figure, the economic merit of a sire for farm income. This total index takes into account all traits; production, efficiency, management and conformation, weighted by economic values (Livestock Improvement, 1991b).

Breeding indices are expressed on a percentage scale with a base of 100 which was the average breeding index of the cows in 1960 (Holmes and Wilson, 1984). The average breeding index of the Holstein Premier Sires team used in the season 1991/92 averaged 151, 140 and 144 units for milkfat, protein and total, respectively (Livestock

Improvement, 1991b). These data reflect a significant increase in the genetic improvement obtained in New Zealand dairy cattle population. In Holmes and Wilson (1984) an outline of the method is given. In general terms, the estimation of a breeding index for a bull (BIB) is

$$BIB = BIA + R[(2 \times CC) - 100 - BIA]$$

where BIA is the breeding index from ancestry data which is the initial estimate of a young bull's breeding index. BIA is calculated from the breeding indexes of his parents as

$$BIA = \tfrac{1}{2}BI \text{ of the sire} + \tfrac{1}{2}BI \text{ of the dam}$$

with reliability (p) equal $\frac{1}{4}$ reliability of sire's BI + $\frac{1}{4}$ reliability of dam's BI.

R is the regression coefficient for the regression of the breeding value of the average future offspring of a bull on the average production of his daughters. It is derived from theory of selection index and assumes a heritability value of 0.25 for milkfat yield and a environmental correlation of zero. Thus, the resulting regression coefficient is

$$R = \frac{(1-p)\sum w}{(1-p)\sum w + 15}$$

where the weights (w) are calculated as

$$w = \frac{n_1 n_2}{n_1 + n_2}$$

and where, for each herd-year, n_1 is the number of daughters and n_2 is the number of contemporaries.

CC is the contemporary comparison, which is not calculated as difference but a proportion of averages. A computational expression is

$$CC = \Sigma \left[w \frac{\overline{D}}{\overline{C}} (50 + \tfrac{1}{2}\overline{BIC}) \right]$$

where $\overline{D}$ is the average yield of daughters, $\overline{C}$ is the average yield of contemporaries and $\overline{BIC}$ is the average breeding index of sires of contemporaries.

The above mathematical procedure is based on the theory of selection index using the following statistical model

$$y_{ijkl} = H_i + G_j + s_{jk} + e_{ijkl}$$

where

y_{ijkl} is the production record of a cow;

H_i is the effect of a particular herd and year in which the record was made;

G_j is the effect of the group of sires in the artificial breeding scheme of which the cow's was member;

s_{jk} is the effect of the sire, taken as deviation from the average of the group he was in; and

e_{ijkl} is the deviation of the particular production record that expected on the basis of the other listed effects.

The method is, therefore, a contemporary comparison modified to combine prior knowledge of the breeding indexes of individuals and contemporaries with new phenotypic information (Shannon, 1974; Wickham and Stichbury, 1980). The ratio between daughters average and contemporaries average breaks a possible relationship between the mean and the variance and adjusts for unequal error variances and should yield similar results to the log transformation used in some other evaluations systems (Everett et al. 1982).

The advantages of the New Zealand contemporary comparison system are based on the well-designed progeny testing program, which removes biases for preferential mating (Garrick et al. 1993), and it appears to be cost-effective (Everett and Jones, 1985). Weakness in the system include selection biases from the use of all records and the possible improper handling of new bulls with little pedigree information (Everett and Jones, 1985). The system, also, ignores son's evaluations (Garrick et al. 1993).

2.8.3 A Prototype Sire Evaluation

Methodology of mixed liner models, with the properties of best linear unbiased prediction (BLUP) developed by Dr. Henderson (1950, 1963, 1973), was proposed as applicable to dairy sire evaluation in New Zealand by Anderson (1974). In 1991 a technical review of the current sire evaluation procedures was initiated and a prototype animal model based on BLUP procedures had been developed, which will be adopted in 1995 (Garrick et al. 1993; Harris et al. 1993). The proposed system is based on a repeatability model across breeds and it has the following features:

(i) The practical problems of biases for genetic trend are solved.
(ii) The breeding value of all animals are concurrently estimated.
(iii) The use of information of relatives is maximised.
(iv) Genetic merit of the herdmates of the daughters of a sires are taken into account in the estimation of the breeding value.
(v) Information on crossbred progeny is utilised.

CHAPTER 3
BEST LINEAR UNBIASED PREDICTION

Genetic evaluation through BLUP procedures as developed by Henderson since 1949 (Van Vleck, 1992), was the consequence of an attempt to combine the power of generalised least-squares with the appealing features of selection index, applied by Hazel (1943) in animal breeding. Fixed effects are estimated and random effects are predicted simultaneously by solving a set of equations. Genetic evaluations obtained using BLUP have many desirable features. These include, the method can account for factors such as, sires coming from different populations using genetic groups (Quaas and Pollak, 1981; Pollak and Quaas, 1983; Quaas 1988; Westell et al. 1988), cow culling, association between sire and herd values, differential treatment, and assortative mating with and without selection (Henderson, 1984; 1990b). BLUP by mixed model methods is now the standard method for evaluation in dairy cattle, beef cattle, and swine in most nations (Henderson, 1990a).

3.1 Linear Model

Different statistical models can be used in the genetic evaluation of dairy cattle. A model equation which may be used for the computation of breeding values of dairy animals for a single trait may be:

$$y_{ijk} = h_i + u_j + e_{ijk} \qquad (3.1)$$

where

y_{ijk} is the k^{th} observation from the j^{th} animal in the i^{th} subclass,

h_i is the mean level of performance for the i^{th} subclass,

u_j is the additive genetic component relating to the performance record,

e_{ijk} is the residual effect, including random environmental and non-additive genetic effects, corresponding to the y_{ijk} record.

Effects of the subclasses are commonly considered as fixed effects while additive genetic effects are considered as random effects. Models containing fixed and random effects are called mixed models (Searle, 1971) and were formally described by Eisenhart (1947). The above model equation (3.1) can be rewritten in matrix notation (denoted by bold letters) as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{3.2}$$

where

- $\mathbf{y}$ is a vector of all observations,
- β is a vector that contains subclass effects,
- $\mathbf{X}$ is the incidence matrix associating effects in β to $\mathbf{y}$,
- $\mathbf{u}$ is a vector containing genetic effects,
- $\mathbf{Z}$ is the known matrix associating genetic effects in $\mathbf{g}$ to $\mathbf{y}$, and
- $\mathbf{e}$ is a vector of residual effects, one effect corresponding to each observation in $\mathbf{y}$.

The model is not completely specified until the distributional properties of the effects in the model equation are defined. The usual definitions for the model equation (3.2) involve the following expectations and variances.

The expectation of $\mathbf{y}$, $\mathbf{u}$ and $\mathbf{e}$ are assumed to be

$$E\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{X}\beta\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

with variance-covariances matrices

$$\text{var}\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

$$\text{and } \text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \tag{3.3}$$

where covariances among the y's, **ZGZ'**, are introduced by having random effects in common. In a common animal breeding application for a single trait analysis $\mathbf{G} = \sigma_a^2\mathbf{A}$, whereas $\mathbf{R} = \sigma_e^2\mathbf{I}$, therefore

$$\text{var}\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\alpha^{-1}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\sigma_e^2$$

$$\text{with } \alpha^{-1} = \frac{\sigma_a^2}{\sigma_e^2}$$

where

A is the numerator relationship matrix between individuals in **g**,

σ_a^2 is the additive genetic variance for the trait,

I is an identity matrix,

σ_e^2 is the residual variance.

The above assumptions have the following implications:

(i) Genetic values are all from the same distribution and have common genetic variance, in the absence of inbreeding.
(ii) Each residual effect has the same variance, and residual effects are mutually uncorrelated.
(iii) random effects **u** and **e** are assumed to have zero covariance, equivalent to assuming no genotype-environment interaction.

A particular case of the model equation (3.1) is that model including the fixed effect of herd-year group (h) and the random effect of sire (s):

$$y_{ijk} = h_i + s_j + e_{ijk} \qquad (3.4)$$

or in matrix notation

$$\mathbf{y} = \mathbf{Xh} + \mathbf{Zs} + \mathbf{e}$$

where

y is the vector of the records,
h is the vector of fixed herd-year effects,
X is the incidence matrix relating herd-year to records,
s is the vector of random sires effects,
Z is the incidence matrix relating sires to daughters' records, and
e is a vector of random residual effects.

This model has been referred to as a **sire model** and was the common model used in the genetic evaluation in several countries. An assumption of the model is that all dams are unrelated and that sires are mated at random cows. It may be clear that those assumptions are often not true. Dams may have more than one progeny, and animals may have more genetic covariances than only through their sires. Furthermore, it is clear that not all dams are average dams. Expectations of different dams are often not equal, particularly when we have practised selection, or when we have dams of different genetic merit.

To avoid those assumptions, it is necessary to include equations for dams as well. A more general analysis is obtained when each observation is written as a function of the breeding value of the animal that made that record (rather than a function of the breeding value of the sire) as considered in Model Equation (3.1). A model with an equation for each animal that made an observation is called an **animal model**. In this model it is essential to include the genetic relationships between the various animals, so that in this way, genetic evaluation accounts for covariances between related animals.

An animal model, hence, requires usually many equations (one per animal, at least), and solving such a model requires a lot of computing effort (time and costs). Since 1980 (Quaas and Pollak, 1980), it has become feasible to construct and solve an animal model with the computer. Currently, animal models are being considered and implemented for dairy populations with up to several millions animals on minicomputers (Wiggans et al. 1988a, 1988b). Many earlier animal models used supercomputers.

3.2 Mixed Model Equations

Consider the model equation (3.2). Henderson (1963; 1973; 1974; 1975b) described various criteria that are desirable in predicting breeding values of animals. The most desirable of these, is minimisation of squared errors of prediction. It is possible, provided **G** and **R**, of equation (3.3), are known, to derive a method to predict **u**, that has the following desirable properties:

(i) Is unbiased in the sense that the predictor $\hat{\mathbf{u}}$ has the same expectation as the expectation of the unknown variable **u**, i.e.,

$$E(\hat{\mathbf{u}}) = E(\mathbf{u}).$$

(ii) Minimises the variance of the error of prediction in the class of linear unbiased predictors, i.e.,

$$\text{Predictor error variance} = \text{var}(\hat{\mathbf{u}} - \mathbf{u}) = \min.$$

(iii) Maximises the correlation between the predictor and the predictand in the class of linear unbiased predictors, i.e.,

$$r_{\hat{\mathbf{u}}\mathbf{u}} = \max$$

(iv) When the distribution is multivariate normal,

a. Yields the maximum likelihood and the best linear unbiased estimator of the conditional mean of the predictand.

b. In the class of linear, unbiased predictor, maximises the probability of a correct pairwise ranking.

Henderson (1963) applied selection index theory combined with least squares method to find the best linear unbiased estimators (BLUE) of β and to use these estimators, β^{o} in predicting **u** satisfying the above desirable criteria. Hence, the method seeks a predictor **a'y** of **m'g**, such that

$$E(\mathbf{m}'\mathbf{g} - \mathbf{a}'\mathbf{y})^2 = \min,$$

and

$$E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\beta\mathbf{X}$$

Applying these restrictions to the model (3.2), the BLUE of β can be obtained from the generalised least squares equations

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta^{o} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

where $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ generally is not of full rank. A solution, denoted by β^{o} instead of $\hat{\beta}$ to indicate that β has many solutions, can be obtained from

$$\beta^{o} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{3.5}$$

$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ being a generalised inverse of $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})$.

The best linear unbiased predictors of $\mathbf{u}$ can be obtained from $(\mathbf{y} - \mathbf{X}\beta^{o})$ as:

$$\hat{\mathbf{u}} = \mathbf{G}\,\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta^{o}) \tag{3.6}$$

this expression is essentially the regression of $\mathbf{u}$ on $\mathbf{y}$ after adjustment of $\mathbf{y}$ for fixed effects $(\mathbf{y} - \mathbf{X}\beta^{o})$. The difficulty in applying this method is that $\mathbf{V}$ is, in practice, often large and non-diagonal, therefore difficult to invert. An alternative method was suggested by Henderson (1973) to simultaneously solve for β^{o} and $\hat{\mathbf{u}}$ without the need of computing $\mathbf{V}^{-1}$. The same BLUE of β in (3.5) and the same BLUP in (3.6) can be obtained by maximising for variation in β and $\mathbf{u}$ the joint density function of $\mathbf{y}$ and $\mathbf{u}$. Differentiating with respect to β and $\mathbf{g}$ and equating to zero gives the following equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \beta^{o} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{3.7}$$

which are called Henderson's Mixed Model Equations (MME). Henderson et al. (1959) proved that β^{o} of (3.7) are BLUE as from generalised least squares and Henderson (1963) proved the $\hat{\mathbf{u}}$ are BLUP. The advantage to these equations lies in the fact that for many applications $\mathbf{R}^{-1}$ and $\mathbf{G}^{-1}$ are feasible to compute because of simple structure, as above pointed out, frequently $\mathbf{R}$ is $\sigma_e^2\mathbf{I}$, and $\mathbf{G}$ is $\sigma_a^2\mathbf{A}$.

Equations (3.7) corresponds to a very general model and **u** can comprise several random factors. Given the assumptions described in the definition of the model (3.2), the Mixed Model Equations reduce to

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \beta^{o} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \tag{3.8}$$

with

$$\alpha = \frac{\sigma_e^2}{\sigma_a^2}$$

3.3 Variance of BLUE and BLUP

Variance of prediction errors is the variance of differences between estimated breeding value ($\hat{\mathbf{u}}$) and the true breeding value (**u**). They can be computed directly from the elements of the inverse matrix of the mixed model equations (Henderson, 1974; Henderson, 1975a; Henderson, 1984). Using the simple equations (3.8), the variance of a linear function of the BLUEs and BLUPs are

$$\text{var}\{\mathbf{k}'\beta^{o} + \mathbf{m}'(\mathbf{u} - \hat{\mathbf{u}})\} = \begin{bmatrix} \mathbf{k}' & \mathbf{m}' \end{bmatrix} \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha \mathbf{A}^{-1} \end{bmatrix}^{-} \begin{bmatrix} \mathbf{k} \\ \mathbf{m} \end{bmatrix} \sigma_e^2$$

In denoting the above matrix of coefficients as

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha \mathbf{A}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\beta\beta} & \mathbf{C}_{\beta u} \\ \mathbf{C}_{u\beta} & \mathbf{C}_{uu} \end{bmatrix}$$

a generalised inverse of the above matrix may, therefore, be represented as

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha \mathbf{A}^{-1} \end{bmatrix}^{-} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta u} \\ \mathbf{C}^{u\beta} & \mathbf{C}^{uu} \end{bmatrix}.$$

Henderson (1975a) showed the following useful results

$$\begin{aligned}
&\text{var}(\mathbf{k}'\beta^{o}) = \mathbf{k}'\mathbf{C}^{\beta\beta}\mathbf{k}\sigma_e^2, \\
&\text{var}(\hat{\mathbf{u}}) = (\mathbf{G} - \mathbf{C}^{\mathbf{uu}})\sigma_e^2, \\
&\text{var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C}^{\mathbf{uu}}\sigma_e^2, \\
&\text{cov}(\hat{\mathbf{u}}, \mathbf{u}') = \text{var}(\hat{\mathbf{u}}) = (\mathbf{G} - \mathbf{C}^{\mathbf{uu}})\sigma_e^2
\end{aligned} \tag{3.9}$$

where

$$\begin{aligned}
\mathbf{C}^{\beta\beta} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-} \\
\mathbf{C}^{\mathbf{uu}} &= [\mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^{-1} - \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Z}]^{-1}.
\end{aligned}$$

Thus, prediction error variances (PEV) of genetic evaluations can be obtained from the diagonal elements of a generalised inverse of the coefficient matrix as

$$\text{PEV} = \text{var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C}^{\mathbf{uu}}\sigma_e^2$$

The reliability (REL) of predicted breeding values is often expressed as the squared correlation between estimated and true breeding values $(r_{\hat{\mathbf{u}}\mathbf{u}})^2$. As in the case of the prediction errors variances, reliability of the estimated breeding values can also be derived from the generalised inverse of the coefficient matrix of the Mixed Model Equations. Let c^{ii} represent the i^{th} diagonal element of $\mathbf{C}^{\mathbf{uu}}$, and using result of (3.9) the reliability of predicted breeding values can be expressed as

$$\text{REL} = (r_{\hat{\mathbf{u}}\mathbf{u}})^2 = \frac{[\text{cov}(\hat{\mathbf{u}}, \mathbf{u}')]^2}{\text{var}(\hat{\mathbf{u}})\ \text{var}(\mathbf{u})} = 1 - c^{ii}\alpha$$

In most applications, diagonal elements of $\mathbf{C}^{\mathbf{uu}}$ are usually not estimated. To date, no completely satisfactory procedures to obtain diagonal elements of $\mathbf{C}^{\mathbf{uu}}$ exist. Alternative methods, based on iterative procedures, have been presented by VanRaden and Freeman (1985), Greenhalgh et al. (1986), Robinson and Jones (1987), Misztal and Wiggans (1988), Meyer (1989) and VanRaden and Wiggans (1991).

3.4 The Numerator Relationship Matrix

Wright (1922) defined the coefficient of relationship as the correlation between additive genetic values, which was derived by path analysis. For use in a mixed model analysis it is the covariance structure that is needed, not the correlation structure. Henderson (1976) presented a recursive method to create the numerator relationship matrix, **A**, required in the prediction of breeding values through the Mixed Model Equations in (3.7).

Let u_i and u_j represent the additive genetic values of the i^{th} and j^{th} animals. The covariance between these animals is

$$\text{cov}(u_i, u_j) = a_{ij}\sigma_a^2.$$

where a_{ij} is the coefficient of genetic additive relationship between the i^{th} and j^{th} animals. Then, representing **A** as

$$\mathbf{A} = \{a_{ij}\}$$

the a_{ij} elements can be recursively computed from

$$a_{ij} = \frac{1}{2}(a_{i,j'} + a_{i,j''}) \qquad \text{for } j \neq i$$

where j' is the sire and j" is the dam of j respectively. The diagonal elements of **A** can be computed from

$$a_{ii} = 1 + \frac{1}{2}a_{i'i''} = 1 + F_i$$

where i' is the sire and i" is the dam of i, respectively and F_i is the inbreeding coefficient of i. A feature of this method is that the animals need be sorted in chronological order (i.e., parents precede progeny).

What is needed, however, is not **A** but its inverse. The amazing fact is that $\mathbf{A}^{-1}$ can be structured from a list of sires and dams without ever setting up **A**. Henderson

(1975c, 1976) discovered how to do this from a list of animals and their parents, and described the relationships between the inbreeding coefficients of the parents and the contributions made by each animal to $\mathbf{A}^{-1}$.

Quaas (1976) arrived at the same results through another recursive method for the derivation of $\mathbf{A}^{-1}$ and pointed out that this method may have some computational advantages. To construct $\mathbf{A}^{-1} = \{a^{ij}\}$, the contributions from the i^{th} animal with parents j and k are

d_{ii}^{-1} to a^{ii}

$-\frac{1}{2}\, d_{ii}^{-1}$ to a^{ij}, a^{ik}, a^{ji}, a^{ki}

$\frac{1}{4}\, d_{ii}^{-1}$ to a^{jj}, a^{jk}, a^{kj}, a^{kk}

where the d_{ii} are the diagonal elements of a matrix **D**, which is recursively computed according to the following rules:

Number of parents identified	Value of d_{ii} in the matrix **D**
Both parents	$1 - \frac{1}{4}\, a_{jj} - \frac{1}{4}\, a_{kk} = \frac{1}{2} - \frac{1}{4}\, F_j - \frac{1}{4}\, F_k$
One parent	$1 - \frac{1}{4}\, a_{jj} = \frac{3}{4} - \frac{1}{4}\, F_j$
Neither parent (base animal)	1

Therefore, diagonals elements in $\mathbf{A}^{-1}$ are the sum of own contribution of the animal and those of its progeny. Offdiagonal elements in $\mathbf{A}^{-1}$ arise between mates and between parents and offspring.

The use of $\mathbf{A}^{-1}$ increases the accuracy of breeding value estimation, because its inclusion in the Mixed Model Equations reduces the variance of prediction error, (Henderson, 1975c) particularly of young animals with few records. For example in sire evaluation, when two sires are related, progeny of one sire provide additional clues to the estimate of breeding value of the other sire and vice versa.

3.5 Genetic Groups

In the obtention of BLUPs for breeding values, a basic assumption is that E[**u**]=**0**, i.e., all animals are distributed about a mean of zero. Quaas et al. (1984) pointed out two situations in which this assumption is not valid. First, animals might came from two or more genetically distinct populations, say, sires from different countries. And second, when analysing records accumulated for several years, the breeding values of animals might have changed. To account for these problems in the genetic evaluation, different grouping strategies have been proposed. There are, however, no generally accepted criteria for defining groups in a particular application. Arbitrary definitions have been used, for example, bulls from a particular stud entering service in the same year, birth date, pedigree information, or geographic region.

Thompson (1979) defined genetic group effect for an animal as a combination of group effects for its ancestors. The Model Equation (3.1) means that breeding value of an individual (u_j) is the average of its parents' breeding values ($0.5u_s + 0.5u_d$) plus a random deviation (ϕ_j) caused by Mendelian sampling of the parental genotypes during gametogenesis (Quaas, 1988). This can be formulated as:

$$u_j = 0.5u_s + 0.5u_d + \phi_j$$

Thus, the vector of breeding values, **u**, in Model Equation (3.2) can be represented as:

$$\mathbf{u} = \begin{bmatrix} \mathbf{P}_b & \mathbf{P} \end{bmatrix} \begin{bmatrix} \mathbf{u}_b \\ \mathbf{u} \end{bmatrix} + \phi$$

$$\mathbf{u} = \mathbf{P}_b\mathbf{u}_b + \mathbf{P}\mathbf{u} + \phi \qquad (3.10)$$

where the matrix $\begin{bmatrix} \mathbf{P}_b & \mathbf{P} \end{bmatrix}$ relates progeny to parents; each row contains two nonzero elements (0.5) in the columns pertaining to sire and dam. If, for example, the sire is known there is a corresponding 0.5 in **P**, otherwise it is unknown and, thus, it is in $\mathbf{P}_b$. Sires and dams located in the $\mathbf{P}_b$ matrix are assigned a "phantom" identification and are called base animals.

Manipulation of (3.10) gives:

$$(\mathbf{I}-\mathbf{P})\mathbf{u} = \mathbf{P}_b\mathbf{u}_b + \phi$$

$$\mathbf{u} = (\mathbf{I}-\mathbf{P})^{-1}[\mathbf{P}_b\mathbf{u}_b + \phi]$$

this means that the vector of breeding values, u, is can be written as a function of the relationship between the animals to evaluate and their known and unknown parents. Quaas (1988) derived the mean and variance of $\mathbf{u}$ when it is a function of $\mathbf{u}_b$ and ϕ, these being:

$$E(\mathbf{u}) = (\mathbf{I}-\mathbf{P})^{-1}\mathbf{P}_b\mathbf{Q}_b\mathbf{g}$$

$$= \mathbf{Q}\mathbf{g}, \text{ and}$$

$$\text{var}(\mathbf{u}) = (\mathbf{I}-\mathbf{P})^{-1}\mathbf{D}\,(\mathbf{I}-\mathbf{P})^{-1}$$

where

$\mathbf{D} = \text{diag}\{0.25m_i + 0.5\}$ for $m_i = 1, 2 =$ the number of base parents of the i^{th} individual.

$\mathbf{Q} = (\mathbf{I}-\mathbf{P})^{-1}\mathbf{P}_b\mathbf{Q}_b$, $\mathbf{Q}_b$ being an incidence matrix relating base animals to their respective base populations means, and

$\mathbf{g}$ = a vector of genetic groups.

Consequently, $E(u_i) = \sum_j q_{ij}g_j$, i.e., the breeding value of an animal is a function of its ancestors.

Once having derived the mean and variance of $\mathbf{u}$, an equivalent model to (3.2) would be employed (Quaas and Pollak, 1981; Pollak and Quaas, 1983; Quaas, 1988) such as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{Q}\mathbf{g} + \mathbf{Z}\mathbf{u}^* + \mathbf{e}$$

with

$$E\begin{bmatrix}\mathbf{u}^*\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix} \text{ and } \text{var}\begin{bmatrix}\mathbf{u}^*\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

As pointed out by Quaas (1988), in this model, groups are just another set of fixed effects; **ZQ** and **g** could be included in **X** and β, respectively. The mixed model equations can be formed as follows:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}ZQ} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z+G^{-1}} & \mathbf{Z'R^{-1}ZQ} \\ \mathbf{Q'Z'R^{-1}X} & \mathbf{Q'Z'R^{-1}Z} & \mathbf{Q'Z'R^{-1}ZQ} \end{bmatrix}\begin{bmatrix}\beta^o\\ \hat{\mathbf{u}}^*\\ \hat{\mathbf{g}}\end{bmatrix}=\begin{bmatrix}\mathbf{X'R^{-1}y}\\ \mathbf{Z'R^{-1}y}\\ \mathbf{Q'Z'R^{-1}y}\end{bmatrix}$$

and solved for β^o, $\hat{\mathbf{u}}^*$, and $\hat{\mathbf{g}}$ and then $\hat{\mathbf{u}} = \text{BLUP}(\mathbf{u}) = \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{u}}^*$. However, this can be costly in forming the equations and in solving them iteratively. Quaas and Pollak (1981) proposed a transformation to solve directly for additive genetic merit. Such a modification is called the QP transformation and lead to the following transformed mixed model equations (Quaas, 1988):

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} & \mathbf{0} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z+G^{-1}} & \mathbf{-G^{-1}Q} \\ \mathbf{0} & \mathbf{-Q'G^{-1}} & \mathbf{Q'R^{-1}Q} \end{bmatrix}\begin{bmatrix}\beta^o\\ \hat{\mathbf{u}}\\ \hat{\mathbf{g}}\end{bmatrix}=\begin{bmatrix}\mathbf{X'R^{-1}y}\\ \mathbf{Z'R^{-1}y}\\ \mathbf{0}\end{bmatrix}$$

This approach to grouping was made computationally feasible by Quaas (1988) who discovered simple rules for constructing group equations and group contributions to animal equations. Westell and Van Vleck (1987) and Westell et al. (1988), also provided an alternative derivation of these rules and applied this to the genetic evaluation of large dairy cattle populations. Arnold et al. (1992) applied this grouping strategy to genetic evaluation of beef cattle across breeds.

Theoretically the use of the relationship matrix can be thought of as alternative to grouping sires according to time of entry into the stud (Thompson, 1979). The use of the relationship matrix should reduce the need for grouping animals.

3.6 Computing Strategies to Solve the Mixed Model Equations

Different mathematical procedures may be used in the solution to the mixed model equations in (3.8). Attempts to compare and improve upon various computational strategies are found readily in the literature (Ufford et al. 1979; Quaas and Pollak, 1981; Blair and Pollak, 1984; Hudson, 1984; Van Vleck and Dwyer, 1985; Schaeffer and Kennedy, 1986a, 1986b; Misztal and Gianola, 1987; Garrick, 1988). In this section a brief description of the mathematical bases are given. In practice, a simple compute implementation may involve more than one of the following techniques.

3.6.1 Absorption

A numerical method to reduce the number of equations may be to absorb one or more equations into the remaining equations. The mixed model equations of (3.8) can be rewritten in a full equation system as follows

$$\begin{array}{lcll} \mathbf{X'X}\beta^{o} + \mathbf{X'Z}\,\hat{\mathbf{u}} = \mathbf{X'y} & & (a) \\ \mathbf{Z'X}\beta^{o} + (\mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{u}} = \mathbf{Z'y} & & (b) \end{array} \qquad (3.11)$$

equations (a) of (3.11) are called fixed effects equations, and similarly, equations (b) of (3.11) are called random effects equations. With the aim of reducing the number of equations, the vector of fixed effects can be computed as

$$\beta^{o} = (\mathbf{X'X})^{-}(\mathbf{X'y} - \mathbf{X'Z}\,\hat{\mathbf{u}})$$

which once substituted in (b), the vector of random effects, $\hat{\mathbf{u}}$, can be computed as

$$\hat{\mathbf{u}} = (\mathbf{Z'PZ} + \alpha\mathbf{A}^{-1})^{-1}\mathbf{Z'Py} \qquad (3.12)$$

where $\mathbf{P} = (\mathbf{I} - \mathbf{X}(\mathbf{X'X})^{-}\mathbf{X'})$ is known as the absorption matrix. Hence, the fixed effects equations have been absorbed into those of random effects equations. This procedure has been well illustrated by Searle (1971) in the solution to general linear models. In the case of genetic evaluation, however, the size of the matrix to be inverted in (3.12) is still too large, its order is equal to the number of animals, so that, it is not always possible to directly invert such matrix because direct inversion of large matrices requires large

computing time and is sensitive to the accumulation of rounding errors that the computer makes (Searle, 1982). To avoid this problem, other alternative computational strategies exist.

3.6.2. Gauss Elimination

Mixed model equations of (3.8) can be represented as $\mathbf{Cb} = \mathbf{r}$, where $\mathbf{b} = \{b_i\}$ is the vector of unknowns breeding values, $\mathbf{r} = \{r_i\}$ is a vector of known values after absorption of the right hand side, and $\mathbf{C} = \{c_{ij}\}$ is the matrix of coefficients. Providing that the system is consistent and C being of full rank, Gauss elimination method allows computing a solution for **b** without directly inverting the matrix C. This method, with suitable modifications, forms the basis for reliable and efficient computer programs for solving system of equations in practice (Ortega, 1987).

The matrix C, through the LU decomposition, can be represented as $\mathbf{C} = \mathbf{LU}$, where **L** is unit lower triangular matrix (all diagonal elements are unity) and **U** is a upper triangular matrix. Derivation of **L** and **U** is illustrated in several matrix algebra books, see for example, Ortega (1987) and Schawarz et al. (1973). Therefore, the system $\mathbf{Cb} = \mathbf{r}$ can now be represented as

$$\mathbf{LUb} = \mathbf{r}$$

setting $\mathbf{Ub} = \mathbf{y}$,

then the original equation system become into two equations as follows

$\mathbf{Ly} = \mathbf{r}$ for **y** with **r** given, and

$\mathbf{Ub} = \mathbf{y}$ for **b** with the auxiliary vector **y** given.

Thus, taking advantage of the triangular structures of **L** and **U**, these two system of equations are equivalent to the problem of solving two simple equations in succession, namely to find the auxiliary vector **y** and from it the solution vector **b**. Because of the sequence in which the equations are used and the respective unknowns solved for, the method is termed **forward and backward substitution**. Another characteristic of the

matrix **C**, further of full rank, implicated in the feasibility of the Gaussian elimination method is that **C** is not necessarily a symmetric matrix but positive definitive (see Searle (1982) for the concept of positive definitive).

3.6.3 Cholesky Decomposition

In a similar way to the Gaussian elimination, the mixed model equations (3.8) can be represented as $\mathbf{Cb} = \mathbf{r}$. Because of in genetic evaluation, the matrix **C** is generally a symmetric positive definitive matrix, the Gauss elimination, doing use of the LU decomposition, may be simplified through the Cholesky decomposition, in which the matrix **C** is decomposed into two symmetric triangular matrices, namely,

$$\mathbf{C} = \mathbf{TT}'$$

where $\mathbf{T} = \{t_{ij}\}$ is a lower triangular matrix obtained by using the following rules given by Quaas et al. (1984):

$$t_{ii} = (c_{ii} - \sum_{k=1}^{i-1} t_{ik}^2)^{1/2}$$

$$t_{ij} = (c_{ij} - \sum_{k=1}^{j-1} t_{ik} t_{jk}) / t_{jj}$$

for i=1, 2, ..., n and j=1, 2, ..., n.

Proceeding in the same way as the Gaussian elimination, the equation system $\mathbf{Cb} = \mathbf{r}$, can be now represented as $\mathbf{TT'b} = \mathbf{r}$. Setting $\mathbf{T'b} = \mathbf{y}$ then the original equation system become into two set of equations,

$$\mathbf{Ty} = \mathbf{r}$$

and

$$\mathbf{T'b} = \mathbf{y}$$

The solution for **y** in the first set of equations is obtained in a forward substitution and with **y** computed, the solution for **b**, in the second set of equations, is obtained in a backward substitution. As above showed, hence, Cholesky's method for solution of symmetric definitive system of mixed model equations is just a symmetric

modification of the Gauss algorithm, in which the elimination was carried out without maintaining symmetry.

3.6.4 Iterative Methods

Solution to the mixed model equations in (3.7) through the LU and Cholesky decomposition, computationally require large memory in the computer, so that, the solution for a large equation system may not always be feasible. However, iterative methods have been alternatively used in the solution of large equation systems involved in genetic evaluation through BLUP procedures. One of the most known are the Gauss-Seidel iteration (Van Vleck and Dwyer, 1985; Schaeffer and Kennedy, 1986a, 1986b) and Jacobi iteration (Misztal and Gianola, 1987; Garrick, 1988).

Taking the representation of the mixed model equations of (3.7) denoted as $\mathbf{Cb} = \mathbf{r}$, as already defined in the Gauss elimination, $\mathbf{C}$ can be decomposed as the sum of three matrices, namely $\mathbf{C} = \mathbf{L} + \mathbf{D} + \mathbf{L}'$, where $\mathbf{L}$ is the strictly lower triangular matrix of the subdiagonal elements, $\mathbf{D}$ the diagonal elements and $\mathbf{L}'$ the matrix of the elements to the right of the diagonals. Thus, another way of describing the mixed model equation system is

$$(\mathbf{L} + \mathbf{D} + \mathbf{L}')\mathbf{b} = \mathbf{r}$$

equivalently,

$$(\mathbf{L} + \mathbf{D})\mathbf{b} = \mathbf{r} - \mathbf{L}'\mathbf{b}. \qquad (3.13)$$

Equation (3.13) gives place to represent the general **Gauss-Seidel iteration** scheme to find a solution for $\mathbf{b}$ through iterative steps, in which, the previous step is used to find a subsequent step. This can be written as

$$b_i^{(k+1)} = b_i^k + \{r_i - \sum_{j=1}^{i-1} c_{ij} b_j^{(k+1)} - \sum_{j=1}^{n} c_{ij} b_j^k\} / c_{ii} \qquad (3.14)$$

where b_i^k represents the value of the i^{th} solution after the k^{th} round of iteration, c_{ii} is the i^{th} diagonal of $\mathbf{C}$, r_i is the i^{th} right-side in $\mathbf{r}$, and $b_i^{(k+1)}$ is the i^{th} solution of $\mathbf{b}$ in the

$(k+1)^{th}$ round of iteration (Quaas et al. 1984; Van der Werf et al. 1991). Gauss-Seidel iteration is guaranteed to converge if **C** is real, symmetric and positive definitive. If instead of converging, the solutions 'blow up" the cause is most likely that **C** was set up incorrectly and is not non-negative definitive (Quaas et al. 1984).

Equation (3.13) can also be rewritten as

$$\mathbf{Db} = \mathbf{r} - \mathbf{L'b} - \mathbf{Lb}$$

so that, a solution for the vector **b** can be obtained by solving

$$\mathbf{b} = \mathbf{D}^{-1} (\mathbf{r} - \mathbf{L'b} - \mathbf{Lb})$$

equivalently

$$\mathbf{b} = \mathbf{D}^{-1}\mathbf{r} - \mathbf{D}^{-1}(\mathbf{L'} + \mathbf{L})\mathbf{b}. \qquad (3.15)$$

Equation (3.15) gives place to represent **Jacobi iteration** (Berger et al. 1989) scheme to find a solution for **b** through iterative steps. This can be written as

$$\mathbf{b}^{k+1} = \mathbf{D}^{-1}\mathbf{r} - \mathbf{D}^{-1}(\mathbf{L'} + \mathbf{L})\mathbf{b}^{k}$$

or, in scalar form

$$b_i^{(k+1)} = b_i^k + (r_i - \sum_{j=1}^{i-1} c_{ij} b_j^k - \sum_{j=1}^{n} c_{ij} b_j^k) / c_{ii}.$$

Jacobi iteration, at difference to Gauss-Seidel iteration, does not use the most recently information to obtain $b_i^{(k+1)}$, that is, the entire vector b is updated simultaneously at the end of one round of iteration. This method is not guaranteed to converge (Garrick, 1988; Misztal and Gianola, 1987).

There exist several approaches of iteration, which differ in efficiency (number of rounds before solutions are reached) and numerical stability (chance of never reaching solutions). Modifications of the Gauss-Seidel method of iteration are known as

successive under relaxation (**SUR**) and successive over relaxation (**SOR**) (Van Vleck and Dwyer, 1985). These modifications are used to accelerate the convergence process. The iterative method of SUR or SOR can be described (Van Vleck and Dwyer, 1985) as:

$$b_i^{(k+1)} = b_i^k + \theta\{ r_i - \sum_{j=1}^{i-1} c_{ij} b_j^{(k+1)} - \sum_{j=1}^{n} c_{ij} b_j^k \} / c_{ii}$$

where θ is the over-relaxation parameter. It can be shown the method is convergent only for $0 < \theta < 2$. If $\theta < 1$, the method is known as under-relaxation (Garrick, 1988).

A difficulty with these methods is determining when to stop and accept solutions. Some criteria are (i) maximum absolute change in any solution, (ii) sum of squares of changes in solutions and (iii) standardised stopping point. The first two are not scale free so that a value to compare them to can only be arrived at by experience. The third criterion was defined by Van Vleck an Dwyer (1985) and it is computed as

$$p = \frac{(\mathbf{e}'\mathbf{e})^{\frac{1}{2}}}{(\mathbf{r}'\mathbf{r})^{\frac{1}{2}}}$$

where $\mathbf{e} = \mathbf{r} - \mathbf{C}\mathbf{b}^n$, n being the number of iterations.

The stopping point, p, may be some constant. A constant value accepted by some experienced research workers has been $p = 1 \times 10^{-4}$. For ranking animals a less precise stopping point may be required than when genetic evaluations are used to predict genetic trend (Blair and Pollak, 1984; Van Vleck and Dwyer, 1985).

3.6.5 Block Iteration

Block iteration is another approach to the Gauss-Seidel iteration obtaining more rapid convergence (Quaas et al. 1984) and can be quite useful if there is a "natural" way of blocking which tends to put the larger elements into the diagonal blocks. An example of this natural blocking is the multiple traits analysis. Consider the genetic evaluation of dairy sires for two traits, say milkfat and milk protein. Ordering the data by traits within sires, the mixed model equations can still be written as in (3.6) with the following

modifications:

$$\mathbf{R} = \mathbf{I}_n * \mathbf{R}_S$$
$$\mathbf{G} = \mathbf{A} * \mathbf{G}_S$$

and where

$\mathbf{R}_S$ is the variance-covariance matrix of residual effects for multiple traits for animals having the same sire. It is obtained as $\mathbf{R}_S = \mathbf{P} - \mathbf{G}_S$ ($\mathbf{P}$ being the phenotypic variance-covariance matrix for the two traits);

$\mathbf{A}$ is the numerator relationship matrix between the sires;

$\mathbf{G}_S$ is the variance-covariance matrix of sire effects for multiple traits. It is obtained as $\mathbf{G}_S = \frac{1}{4}\mathbf{G}$ ($\mathbf{G}$ being the additive genetic variance-covariance matrix for the two traits);

$\mathbf{I}_S$ is an identity matrix of order equal to the number of animals.

Now the mixed model equations have a symmetric form and the submatrix $\mathbf{Z'R^{-1}Z + G^{-1}}$ has $\underline{s}$ blocks of matrices of order 2x2. Before solving for $\hat{\mathbf{g}}$, fixed effects equation can be absorbed into sire equations leaving symmetric equations of the form as illustrated

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1n} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2n} \\ \vdots & \vdots & & \vdots \\ \mathbf{C}_{n1} & \mathbf{C}_{n2} & \cdots & \mathbf{C}_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{bmatrix}$$

Analogous to the formulation of the general scheme of the Gauss-Seidel iteration, the solution for the b_i block in the $(k+1)^{th}$ iteration can be represented as

$$\mathbf{b}_i^{(k+1)} = \mathbf{b}_i^k + \mathbf{C}_{ii}^{-1}\{ \mathbf{r}_i - \sum_{j=1}^{i-1} \mathbf{C}_{ij}\mathbf{b}_j^{(k+1)} - \sum_{j=1}^{n} \mathbf{C}_{ij}\mathbf{b}_j^k \}.$$

The above expression is similar to (3.11) but instead of estimating individual elements of vectors, blocks of the vector solution are estimated. The inverse of each submatrix $\mathbf{C}_{ii}$ need only be computed once. When the order of each block is small, an inverse procedure may be as efficient as an indirect procedure such as a forward and backward procedure based on a Cholesky decomposition of $\mathbf{C}_{ii}$ and the inverse is computed only once and not for each round.

3.7 Example of the use of the BLUP procedure

The best way to illustrate the flexibility of the BLUP procedure may be to give a numerical example. Consider the case where milkfat yield records of cows with variable number of lactations in different year-herd groups (Table 3.1) are used to obtain the estimated breeding value (EBV) for all animals and the estimated producing ability (EPA) for the cows with records.

Table 3.1. Milkfat yield of cows in different lactations for the example of genetic evaluation through BLUP procedure.

Cow no.	Sire no.	Dam no.	Lactation no.	Herd-year group	Milkfat yield (kg)
4	1	2	1	1	98
			2	2	132
5	1	?	1	1	100
6	?	2	1	1	124
7	1	4	1	1	69
			2	2	75
8	?	?	1	1	88
			2	1	138
9	3	7	1	2	116

Parameters values for heritability and repeatability for milkfat yield are assumed to be 0.25 and 0.50, respectively. As can be seen, nine animals are involved in the analysis,

two sires and seven cows. A repeatability model (Henderson, 1975d) appears appropriate to be used for these data. This might be

$$y_{ijk} = h_i + u_j + p_j + e_{ijk}$$

where y_{ijk} is the k^{th} record of the j^{th} animal in the i^{th} herd-year group; h_i is the effect of the i^{th} herd-year group; u_j is the additive genetic value of the j^{th} animal; p_j is the permanent environmental effect associated with all records of the j^{th} animal; and e_{ijk} is the temporary environmental effect associated with the y_{ijk} record. Herd-year effects are considered as fixed effects and the others effect are considered as random effects.

In matrix notation the above model can be rewritten as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{p} + \mathbf{e} \qquad (3.16)$$

where now, **y** is the vector of observations; **h** is the unknown vector of herd-year effects, **u** is the vector of unobservable genetic effects; **p** is the vector of unobservable permanent environmental effects; **e** is the vector of residual errors corresponding to the vector of observations; and **X**, **W** and **Z** are matrices describing the structure of the data. The distributional properties of this model are

$$E\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{Xh}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

and

$$\text{var}\begin{bmatrix}\mathbf{u}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{I}\sigma_a^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where σ_a^2 is the additive variance for milkfat yield; σ_p^2 is the variance of permanent

effects; and σ_e^2 is the variance of environmental effects.

The BLUPs of **u** and **p**, as well as the BLUE of estimable functions of **h**, can simultaneously be obtained by solving the mixed model equations derived by Henderson (1975d). These are

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \alpha\mathbf{A}^{-1} & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \kappa\mathbf{I}^{-1} \end{bmatrix} \begin{bmatrix} \beta^{o} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix} \tag{3.17}$$

where

$$\alpha = \frac{\sigma_e^2}{\sigma_a^2},$$

and

$$\kappa = \frac{\sigma_e^2}{\sigma_p^2}.$$

The sum of the solutions $\hat{u}_j + \hat{p}_j$ predicts the producing ability for the j^{th} animal, $\hat{q}_j$, (Van Vleck, 1979). Lush (1945) defined heritability and repeatability as

$$h^2 = \frac{\sigma_a^2}{\sigma_y^2}, \text{ and}$$

$$r = \frac{\sigma_a^2 + \sigma_p^2}{\sigma_y^2}, \text{ respectively,}$$

where

$$\sigma_y^2 = \sigma_a^2 + \sigma_p^2 + \sigma_e^2.$$

Thus

$$\sigma_a^2 = \sigma_y^2 h^2,$$

$$\sigma_p^2 = \sigma_y^2 (r - h^2),$$

and

$$\sigma_e^2 = \sigma_y^2 (1 - r),$$

so that

$$\alpha = \frac{\sigma_y^2 (1-r)}{\sigma_y^2 h^2} = \frac{1-r}{h^2},$$

and

$$\kappa = \frac{\sigma_y^2 (1-r)}{\sigma_y^2 (r-h^2)} = \frac{1-r}{r-h^2}.$$

Given the assumed values of heritability and repeatability, $\alpha = 2$ and $\kappa = 2$.

The numerical representation of the model (3.16) using the data of Table 3.1., with records ordered so that parents precede progeny is

$$\begin{bmatrix} 98 \\ 132 \\ 100 \\ 124 \\ 69 \\ 75 \\ 88 \\ 138 \\ 116 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_4 \\ p_5 \\ p_6 \\ p_7 \\ p_8 \\ p_9 \end{bmatrix} + \begin{bmatrix} e_{141} \\ e_{242} \\ e_{151} \\ e_{161} \\ e_{171} \\ e_{272} \\ e_{181} \\ e_{282} \\ e_{291} \end{bmatrix}$$

The numerator relationship matrix, **A**, is created by using the recursive method

given by Quaas et al. (1984) (see section 3.4.). This matrix is

$$\mathbf{A} = \frac{1}{16}\begin{bmatrix} 16 & 0 & 8 & 8 & 8 & 0 & 12 & 0 & 10 \\ 0 & 16 & 0 & 8 & 0 & 8 & 4 & 0 & 2 \\ 8 & 0 & 16 & 4 & 4 & 0 & 6 & 0 & 11 \\ 8 & 8 & 4 & 16 & 4 & 4 & 12 & 0 & 8 \\ 8 & 0 & 4 & 4 & 16 & 0 & 6 & 0 & 5 \\ 0 & 8 & 0 & 4 & 0 & 16 & 2 & 0 & 1 \\ 12 & 4 & 6 & 12 & 6 & 2 & 20 & 0 & 13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 \\ 10 & 2 & 11 & 8 & 5 & 1 & 13 & 0 & 19 \end{bmatrix}$$

for example, $a_{54} = \frac{1}{2}(a_{51} + a_{52})$

$= \frac{1}{2}(\frac{1}{2} + 0)$

$= \frac{4}{16}$ because 1 and 2 are parents of 4.

A diagonal matrix, **D**, is required to construct the inverse of the numerator relationship matrix, $\mathbf{A}^{-1}$. Applying the rules given in section (3.4)

$$\mathbf{D} = \frac{1}{16}\begin{bmatrix} 16 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 7 \end{bmatrix}$$

for example the diagonal element 9 is

$$
\begin{aligned}
d_{99} &= \tfrac{1}{2} - \tfrac{1}{4} F_7 \\
&= \tfrac{1}{2} - \tfrac{1}{4}(0) - \tfrac{1}{4}(\tfrac{1}{4}) \\
&= \tfrac{1}{2} - \tfrac{1}{16} \\
&= \tfrac{7}{16} \text{ because 7 and 3 are parents of 1.}
\end{aligned}
$$

The inverse of **D** is

$$
\mathbf{D}^{-1} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{4}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{4}{3} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{4}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{16}{7}
\end{bmatrix}
$$

hence, the inverse of the numerator relationship matrix is

$$
\mathbf{A}^{-1} = \begin{bmatrix}
2.67 & 0.50 & -0.67 & -0.50 & -0.67 & 0.00 & -1.00 & 0.00 & 0.00 \\
0.50 & 1.83 & 0 & -1.00 & 0.00 & -0.67 & 0.00 & 0.00 & 0.00 \\
-0.67 & 1.00 & 1.90 & 0.00 & 0.00 & 0.00 & 0.57 & 0.00 & -1.14 \\
-0.50 & 0.00 & 0.00 & 2.50 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 \\
-0.67 & -0.67 & 0.00 & 0.00 & 1.33 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.33 & 0.00 & 0.00 & 0.00 \\
-1.00 & 0.00 & 0.57 & -1.00 & 0.00 & 0.00 & 2.57 & 0.00 & -1.14 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1 & 0.00 \\
0.00 & 0.00 & -1.14 & 0.00 & 0.00 & 0.00 & -1.14 & 0.00 & 2.29
\end{bmatrix}
$$

for example, the contributions to $\mathbf{A}^{-1}$ from animal 7 with parents 1 and 4 are

2 to a^{77},

$-\frac{1}{2}(2)=-1$ to a^{17}, a^{47}, a^{71} and a^{74},

$\frac{1}{4}(2)=0.5$ to a^{11}, a^{14}, a^{41} and a^{44}.

Taking the above results, the mixed model equations (3.17) are

$$\left[\begin{array}{cc:ccccccccc:cccccc}
5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 4 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\
\hdashline
0 & 0 & 5.33 & 1 & -1.33 & -1 & -1.33 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 3.67 & 0 & -2 & 0 & -1.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1.33 & 0 & 3.81 & 0 & 0 & 0 & 1.14 & 0 & -2.29 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & -1 & -2 & 0 & 7.00 & 0 & 0 & -2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -1.33 & 0 & 0 & 0 & 3.67 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & -1.33 & 0 & 0 & 0 & 3.67 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & -2 & 0 & 1.14 & -2 & 0 & 0 & 7.14 & 0 & -2.29 & 0 & 0 & 0 & 2 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.00 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\
0 & 1 & 0 & 0 & -2.29 & 0 & 0 & 0 & -2.29 & 0 & 5.57 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hdashline
1 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 3
\end{array}\right]
\left[\begin{array}{c} h_1^o \\ h_2^o \\ \hdashline \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \\ \hat{u}_8 \\ \hat{u}_9 \\ \hdashline \hat{p}_4 \\ \hat{p}_5 \\ \hat{p}_6 \\ \hat{p}_7 \\ \hat{p}_8 \\ \hat{p}_9 \end{array}\right]
=
\left[\begin{array}{c} 479 \\ 461 \\ \hdashline 0 \\ 0 \\ 0 \\ 230 \\ 100 \\ 124 \\ 144 \\ 226 \\ 116 \\ \hdashline 230 \\ 100 \\ 124 \\ 144 \\ 226 \\ 116 \end{array}\right]$$

Let represent the above equations as $\mathbf{Cb=r}$, where $\mathbf{C}$ is of full rank and, in turn, can be represented as

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\alpha\mathbf{A}^{-1} & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W}+\kappa\mathbf{I}^{-1} \end{bmatrix} = \mathbf{C} = \begin{bmatrix} \mathbf{C}_{\beta\beta} & \mathbf{C}_{\beta u} & \mathbf{C}_{\beta p} \\ \mathbf{C}_{u\beta} & \mathbf{C}_{uu} & \mathbf{C}_{up} \\ \mathbf{C}_{p\beta} & \mathbf{C}_{pu} & \mathbf{C}_{pp} \end{bmatrix}$$

The vector solution, $\hat{\mathbf{b}}$, is obtained as $\hat{\mathbf{b}} = \mathbf{C}^{-1}\mathbf{r}$, i.e.,

$$\hat{\mathbf{b}} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta u} & \mathbf{C}^{\beta p} \\ \mathbf{C}^{u\beta} & \mathbf{C}^{uu} & \mathbf{C}^{up} \\ \mathbf{C}^{p\beta} & \mathbf{C}^{pu} & \mathbf{C}^{pp} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

The vector solution is

$$\hat{\mathbf{b}} = \begin{bmatrix} \mathbf{h}^o \\ \hat{\mathbf{u}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} h_1^o \\ h_2^o \\ \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \\ \hat{u}_8 \\ \hat{u}_9 \\ \hat{p}_4 \\ \hat{p}_5 \\ \hat{p}_6 \\ \hat{p}_7 \\ \hat{p}_8 \\ \hat{p}_9 \end{bmatrix} = \begin{bmatrix} 98.09 \\ 122.83 \\ -7.02 \\ 2.08 \\ -3.42 \\ -3.70 \\ -2.42 \\ 6.02 \\ -11.94 \\ 0.84 \\ -7.58 \\ 4.11 \\ 1.45 \\ 6.63 \\ -13.27 \\ 0.84 \\ 0.24 \end{bmatrix}$$

Diagonal elements of the submatrix $\mathbf{C}^{uu}$ are required to estimate the prediction error variance and reliability of breeding values (section 3.3). These values are

Diag{$\mathbf{C}^{uu}$}={0.43 0.46 0.46 0.40 0.41 0.41 0.47 0.38 0.47}.

The estimated breeding values and their prediction error variances and reliabilities are summarised in Table 3.2.

For example, the diagonal element of C^{uu} for animal 9 is 0.47, therefore,

$PEV_{EBV_9} = 0.47\sigma_e^2$

$REL_{EBV_9} = 1\text{-}(0.47)2 = 0.06$

Table 3.2. Estimates of breeding values (EBV) for lactation milkfat yield and their prediction error variances (PEV_{EBV}) and reliabilities (REL_{EBV}) for animals in the example of genetic evaluation through BLUP procedure.

Animal no.	EBV	PEV_{EBV}*	REL_{EBV}
1	-7.02	0.43	0.14
2	2.08	0.46	0.08
3	-3.42	0.46	0.08
4	-3.70	0.40	0.20
5	-2.42	0.41	0.18
6	6.02	0.41	0.18
7	-11.94	0.47	0.06
8	0.84	0.38	0.24
9	-7.58	0.47	0.06

* these values need to be multiplied by σ_e^2.

The sum of $\hat{\mathbf{u}} + \hat{\mathbf{p}} = \hat{\mathbf{q}}$ are the EPA of the cows with milkfat yield records. Applying the general theory of estimability (see Searle, 1971) $\hat{\mathbf{u}} + \hat{\mathbf{p}}$ is an estimable function. Defining

$$\mathbf{a}' = \left[\begin{array}{cc:cccccccccc:cccccc}
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]$$

the vector of EPAs, $\hat{\mathbf{q}}$, is obtained as $\mathbf{a}'\hat{\mathbf{b}}$, i.e.,

$$\hat{\mathbf{q}} = \begin{bmatrix} 3.70 + 4.11 \\ -2.42 + 1.45 \\ 6.02 + 6.63 \\ -11.94 - 13.27 \\ 0.84 + 0.84 \\ -7.58 + 0.24 \end{bmatrix} = \begin{bmatrix} 0.41 \\ -0.97 \\ 12.65 \\ -25.21 \\ 1.68 \\ -7.34 \end{bmatrix}.$$

The prediction error variances of these estimates are calculated as

$$\text{Var}(\hat{\mathbf{q}}) = \text{Var}(\mathbf{a}'\hat{\mathbf{b}}) = \mathbf{a}'\mathbf{C}^{-1}\mathbf{a}\sigma_e^2,$$

where $\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}$ is a matrix of order 6 x 6 and its diagonal elements, required to calculate the PEV of EPA (PEV_{EPA}), are

$$\text{Diag}\{\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}\} = \{0.55 \;\; 0.64 \;\; 0.63 \;\; 0.58 \;\; 0.50 \;\; 0.72\}.$$

The reliabilities of estimated producing abilities for the six cow with records are calculated using the methods already explained in section (3.3).

$$\text{REL}_{\text{EPA}} = (r_{\hat{\mathbf{q}},\mathbf{q}})^2 = \frac{[\text{cov}(\hat{\mathbf{q}},\mathbf{q})]^2}{\text{var}(\hat{\mathbf{q}})\,\text{var}(\mathbf{q})}$$

where $\mathbf{q}$ are the true producing abilities estimated by $\hat{\mathbf{q}}$. Using selection theory, it can be shown that the above expression can be simplified to

$$\text{REL}_{\text{EPA}} = (r_{\hat{\mathbf{q}},\mathbf{q}})^2 = 1 - \frac{1-r}{r}\text{Diag}\{\mathbf{a}'\mathbf{C}\mathbf{a}\}$$

Results of cows with records are summarised in Table 3.3.

Table 3.3. Estimates of producing abilities (EPA) for lactation milkfat yield and their prediction error variances (PEV_{EPA}) and reliabilities (REL_{EPA}) for the cows with records in the example of genetic evaluation through BLUP procedure.

Cow no.	EPA	PEV_{EPA}*	REL_{EPA}
4	0.41	0.55	0.45
5	-0.97	0.64	0.36
6	12.65	0.63	0.37
7	-25.21	0.58	0.42
8	1.68	0.50	0.50
9	-7.34	0.72	0.28

* these values need to be multiplied by σ_e^2.

An estimable function of interest is the difference between the two herd-year groups, $h_1 - h_2$. Let $\mathbf{k}'=[1\ \ -1]$, then

$$\text{BLUE}(h_1 - h_2) = h_1^\circ - h_2^\circ$$

$$= \mathbf{k}'\beta^\circ$$

$$= [1 \quad -1] \begin{bmatrix} 98.09 \\ 122.83 \end{bmatrix}$$

$$= -24.74.$$

This means that cows which made a lactation in herd-year 2 averaged 24.74 kg milkfat more than cows which made a lactation in herd-year 1.

Using the general results explained in section (3.3), the variance of $\mathbf{k}'\beta^\circ$ is calculated as

$$\text{var}(\mathbf{k}'\beta^\circ) = \mathbf{k}'C^{\beta\beta}\mathbf{k}\sigma_e^2,$$

where the submatrix $C^{\beta\beta}$ for this example is

$$\begin{bmatrix} 0.4590 & 0.2467 \\ 0.2467 & 0.5702 \end{bmatrix},$$

so that

$$\text{var}(\mathbf{k}'\beta^{o}) = 0.54\sigma_e^2.$$

3.8 Heterogeneous Variance

In animal breeding practice, selection decisions often have to be made among animals in separate environmental groups, which may differ both in mean and in the variability within them. Homogeneity of variance is assumed in most currently applied models of genetic evaluation. However, numerous studies (Van Vleck, 1963; Everett et al. 1982; Powell et al. 1983; Hill et al. 1983; Mirande and Van Vleck, 1985; Brotherstone and Hill, 1986; De Veer and Vleck, 1987; Meinert et al. 1988; Dong and Mao, 1990; Boldman and Freeman, 1990; Wiggans and VanRaden, 1991; Visscher et al. 1991) have reported heterogeneous genetic, residual, permanent environmental, and phenotypic variances for production traits with respect to geographical region, production level, herd, and other factors. Several possibilities can be found about heterogeneity of variances across environments (e.g. herds). Garrick and Van Vleck (1987) based on Henderson (1984) identified the followings:

(1) Unit genetic correlation between genetic merit in each environment
   - (i) Equal additive genetic and residual variances in all environments. General assumption in BLUP procedures for genetic evaluation.
   - (ii) Equal additive genetic variances but residual variances with magnitudes dependent on the environment. Consequently, heritability will vary between environments.
   - (iii) Additive genetic variances differing according to the environment and residual variances constant. Heritability will vary with environment.
   - (iv) Additive genetic and residual variances changing proportionally such that heritability remains constant across environments.
   - (iv) Additive genetic and residual variances changing such that heritability is variable.

(2) Genetic correlation of less than one between performance in different environments.

3.8.1 Effects of Heterogeneous Variance

3.8.1.1 *Accuracy of Selection*

Failure to take account of the heterogeneity of variance may lead to inaccurate and biased predictions. If variances increase with mean yield but are assumed to be homogeneous, superior cows in herds with large variances or sires with a large percentage of their daughters in herds with large variances would tend to be overevaluated (Visscher and Hill, 1992). An apparent excess of elite cows has been reported for high mean herds (Powell et al. 1983) and for high variance herds (Everett et al. 1982) from analysis that assume equal genetic and residual variances for all records.

3.8.1.2 *Mass Selection*

When selection is made on individual performance, as in the case of selection of cows to breed bulls, expected genetic superiority of cows in herd i (ΔG_i) is

$$\Delta G_i = i_i h_i \sigma_a$$

where i_i is the standardised selection differential in herd i, h_i is the square root of the heritability for the trait in the herd i and σ_a is the additive genetic standard deviation. For example, if the proportion of selected cows is 0.10, then i_i =1.75 and

$$\Delta G_i = 1.75\, h_i \sigma_a .$$

One potential problem arising from heterogeneity of variance is that the proportions of animals selected from herds with different variances is a function of herd variation for the criterion of selection. Hill (1984), showed that if the phenotypic standard deviation, σ_{p_i}, of a herd i is 1.5 times greater than another herd, and the overall proportion selected is 0.10, 72% of the cows selected will come from the more variable herd. This yields proportion selected of 0.144 (i_1=1.57) and 0.052 (i_2=2.03) in the more and less variable herds, respectively. However, the effect on response to selection from choosing greater proportions of individuals from the more variable herds depends on the extend to which the greater variability is due to genetic as opposed to environmental

factors. Van Vleck (1987) and Vinson (1987) gave examples of losses in response to selection when heterogeneity of variance is not taken into account. For example, under the same situation as above given, if the additive genetic standard deviation is likewise 50% larger in the more variable herds such that heritabilities are equal in the two herds, weighting expected genetic superiority of selected cows (measured relative to some standard environment with additive genetic standard deviation, σ_a) in the two herds according to the proportion of cows selected from each, yields:

$$\Delta G = [0.72(1.57) + 0.28(2.03)]\, h\sigma_a = 1.69 h\sigma_a$$

This value compared to $1.75\, h_i\sigma_a$ shows a reduction in genetic superiority of selected cows due to heterogeneity of variance of 3.4%.

A further problem is that the proportion of individuals selected from the more variable herd increases with the intensity of selection (Hill, 1984). Altering this example to an overall proportion selected of 0.05 rather than 0.10, 82% of the cows selected will come from the more variable herd (Table 1 in Hill, 1984). This new proportions will reduce the expected genetic superiority of selected cows in 5.3% (Vinson, 1987).

3.8.1.3 *Sire Selection*

Consider the simple genetic model in Equation (3.4), $y_{ijk} = h_i + s_j + e_{ijk}$, where y_{ijk} is the observation in the k^{th} daughter of the j^{th} sire in the i^{th} herd, h_i is the fixed herd effect and s_j and e_{ijk} are random sire and residual effects respectively. If genetic and residual variances are known in each herd, the appropriate predictions would be obtained using selection index theory. Ignoring relationships between sires, in herd i, an estimate of the sire effect is

$$\hat{s}_j = b_{ij}\bar{y}_{ij\cdot}$$

where $\bar{y}_{ij\cdot}$ is the average of records of n_{ij} daughters of sire j in herd i adjusted for fixed non-genetic effects and b_{ij} is a weighting factor.

For selection of bulls from bulls proofs based on mixture of daughters from two herds, a weighted estimate of the breeding value of sire j (I_j) was derived by Van Vleck (1987) as follows:

$$I_j = \frac{b(n_{1j}\bar{y}_{1j} + n_{2j}\bar{y}_{2j})}{(n_{1j} + n_{2j})},$$

then, the expected genetic superiority of the selected bulls when used in environment i is:

$$\Delta G_i = \frac{Cov(u_i, I)}{\sigma_I} i$$

where u_i is the true breeding value of the bulls in environment i, σ_I is the standard deviation of the weighted index and i is the selection intensity factor. After some algebra, it can shown that

$$\Delta G_i = \frac{\left[\frac{b}{2(n_{1j} + n_{2j})}\right]\left[n_{1j}\sigma^2_{a_1} + n_{2j}\sigma_{a_1,a_2}\right]}{\sqrt{\left[\frac{b}{(n_{1j} + n_{2j})}\right]^2\left[n^2_{1j}Var(\bar{y}_{1j}) + n^2_{2j}Var(\bar{y}_{2j}) + 2n_{1j}n_{2j}Cov(\bar{y}_{1j}, \bar{y}_{2j})\right]}} i$$

where $\sigma^2_{a_i}$ is the additive genetic variance in herd i, σ_{a_1,a_2} is the additive genetic covariance for the two herds. The variance of the average of records of cows in herd i, daughters of sire j is

$$\sigma^2_{\bar{y}_{ij}} = \frac{\sigma^2_{y_i}\left[1 + \frac{(n_{ij}-1)h^2_i}{4}\right]}{n_{ij}}$$

and where $\sigma^2_{y_i}$ is the phenotypic variance of the trait in environment i, and h^2_i is the

heritability for the trait in herd i. The covariance between the average of observations of daughters of sire j in herd 1 and herd 2 is

$$\mathrm{Cov}(\bar{y}_{1j}, \bar{y}_{2j}) = \frac{\sigma_{a_1,a_2}}{4}.$$

Van Vleck (1987) estimated the expected genetic superiority of selected bulls based on evaluation in two environments when selection was based on various numbers of daughters from two environments having different heritabilities and different residual standard deviations for different values of genetic correlations between expressions of genotypes in two environments.

Evaluation of bulls exclusively in herds with larger heritability gave greatest genetic superiority in both environments if the genetic correlation was unity. If the genetic correlation was less than 1, then the expected genetic superiority of the selected bulls in the environment with larger heritability was still maximum when all daughters were in that environment but expected genetic superiority in the environment with smaller heritability also could be greater when all daughters were in the environment with larger heritability. If the genetic correlation was substantially less than unity, inappropriate allocation of daughters to test environment could reduce expected genetic superiority.

If heterogeneity of variance is ignored, therefore, the performance of daughters of a sire in effect will be weighted in proportion to the phenotypic standard deviations of the herds in which the daughters perform. The results is that performance of daughters in more variable herds will influence the eventual evaluation to a greater extend than does performance of daughters in less variable herds. To the extent that daughters of sires are distributed equally with respect to herd variation, no bias in evaluation should result (Vinson, 1987). However, if heritability also differs across herds, accuracy of evaluation will be reduced by failure to account for such differences when evaluations of sires are computed.

3.8.1.4 *BLUP*

For most genetic evaluations in practice using BLUP with a sire model or an animal model, different types of relatives contribute to the prediction of breeding values

so it is not obvious how to predict losses in accuracy and therefore in genetic gain if heterogeneity of variance is ignored.

In a simulation study, Garrick and Van Vleck (1987) investigated losses in response due to heterogeneous variances between three environments corresponding to herd mean production groups by deterministic simulation using two set of parameters. Responses were compared to the response using a multitrait model which was assumed to be correct model. For one parameter set phenotypic variances and heritabilities increased from the low to the high mean production group, and ignoring heterogeneity of variance led to a negligible reduction in asymptotic rate of response to selection. For a second parameter set, genetic (sire) variances were constant across environments and residual variances decreased with increased herd mean production levels, so that heritabilities decreased with increased phenotypic variances. This led to a reduction in the rate of response of approximately 3% when heterogeneous variances were ignored in the genetic evaluation. The authors thus concluded that, in practice, progeny testing schemes were robust to violations of assumptions regarding homogeneity of variances between environments.

Meuwissen and Van der Werf (1993) simulated data for a dairy population undergoing selection and investigated losses in response to selection when heterogenous variances between herds were ignored in the genetic evaluation. Data were simulated either with or without heterogeneous variances and heritabilities were constant across herds. Breeding values were predicted using an animal model and response to selection was computed assuming there was homogeneity of variance. For both progeny testing and open nucleus schemes there were not significant losses in response when variances heterogeneous rather than homogeneous.

3.8.1.5 *Correlations Between Breeding Values Estimated at Different Environments*

There are two important questions about the genetic evaluation of dairy sires. First question is if estimated breeding values for a production trait of sires will rank in the same order over a wide range of environmental conditions. And second question is if the difference between the estimated breeding values for the production trait is the same at all levels of production. The first question is answered by the correlation between breeding values estimated at different environments and the second question is solved for

a statistical proof of significantly deviating breeding values.

Manson and Robertson (1956) used data of Red Danish herds in Denmark. Herds were classified into three equal groups on the basis of their average production. The ranking of bulls for breeding values, which were estimated using the contemporary comparison method, was apparently the same at all levels of milk production.

Robertson et al. (1960) used milk records of the progeny of Friesian and Ayrshire bulls in Great Britain. Contemporary groups (herd-year) were ranked in order of the average heifer yield and were split into three groups, low-yielding, medium-yielding and high-yielding contemporary groups. For each bull three independent contemporary comparisons were calculated, each based upon the same amount of information (the same number of "effective daughters"). The three set of contemporary comparisons provided three independent measures of the breeding value of each bull. Correlations between estimated breeding values at each pair of levels were obtained. The results showed that the correlation between breeding values of bulls at different production levels was very high indicating that irrespective of level of production, breeding values remain the same.

Lytton and Legates (1966) used first records of artificially sired daughters of 46 Holstein sires used in the northern and southern regions of the United States. Correlations between the average breeding values of the sires (estimated by comparison contemporary method) in the two regions for milk production and milkfat production approached unity, indicating that the ranking of sires was essentially the same in both regions.

Progeny of American Holstein bulls were classified into one of four groups, based on milk yield level of herdmates (McDaniel and Corley, 1967). Daughter average rose with increases in herdmate level, while the magnitude of the breeding value (Predicted difference) of the sires decreased. Correlations among predicted differences estimated at different herdmate levels were high (>0.88). The magnitude of the correlations suggested that the breeding values of the bulls was almost the same from level to level.

In Canada, Burnside and Rennie (1968) determined the correlations among the contemporary comparison proofs of 19 Holstein sires, each evaluated on the production of 20 or more daughters at four levels of herd production. The correlations ranged from 0.73 to 1.01. The authors concluded that there were little change in ranking of sires

across herd level of production.

Danell (1982) analysed first lactation records of Swedish Red and White cows. The data were divided into three groups according to average level of production in the herd. Breeding values were estimated for each sire on each production level by best linear unbiased predictor (BLUP) method. Product-moment and rank correlations between breeding values estimated at the three production levels were high and were not significantly different to the expected correlations. Therefore, she concluded that ranking of sires on the estimated breeding values for milk yield was essentially the same across production levels.

Estimated correlations between breeding values for sires in different countries have turned out, in some reports, to approximate the expected correlations, thus indicating that the genetic correlation is unity. Petersen (1975) reported genetic correlations of 0.91 for milk yield, 0.97 for fat percentage, and 0.79 for fat yield between progeny tests of 93 bulls in Denmark with test in Bulgaria and Czechoslovakia. Powell and Dickinson (1977) estimated the predicted differences by the Modified Contemporary Comparison method of bulls in the US and in Mexico. This situation provided a unique opportunity to study the relationship between progeny tests of the same bulls in different countries without the confusion of differing methods of evaluation. The correlation between Mexican and US Modified Contemporary Deviation (MCD) was 0.66, and the expected value was 0.65. In a subsequent study, Powell and Wiggans (1991) using an animal model through BLUP procedures, estimated that the correlations between Mexican and US probable transmission abilities (PTA) were 0.91 compared with an expected correlation of 0.90. Thus, these results suggest no evidence of a difference in ranking (genotype-country interaction) exists.

Other studies, in contrast, have reported low correlations between breeding values estimates in different countries. The breeding value of a small group of US sires which had progeny in the US, Mexico and Colombia, were estimated through BLUP procedures. Rank correlation for sire value in Mexico and Colombia was low (0.26) and not significant, indicating considerable changes in rank due to a possible sire-environment interaction (Abubakar et al. 1987).

In New Zealand, an experiment was initiated in 1984 to study the differences between the offspring of Canadian and New Zealand sires under New Zealand conditions (Canadian/New Zealand Genotype-Environmental Interaction Trial; Peterson, 1988).

Breeding values of sires in the trial were calculated using BLUP methodology from the experimental data. Home proofs for Canadian sires were from Agriculture Canada and Holstein Association of Canada. New Zealand proofs were supplied by the New Zealand Dairy Board. Correlations between the trial and home proofs were calculated and tested against a theoretically expected value. For the New Zealand sires these correlations were not different from the expected value and were all greater than zero. The correlation between the trial and home proofs of the Canadian sires was significantly different from expectation and not different from zero. This means that sire proofs made under Canadian feeding and management conditions are not good predictors of a bull's relative merit in the New Zealand environment.

3.8.1.6 *Genetic Correlations*

Lush (1945) suggested that selection for a trait should occur in the same environment in which the selected animals would eventually perform. Hammond (1947) indicated that selection for a trait in the environment which will allow its maximum expression will result in the most rapid genetic improvement. Sires to be used in low producing dairy herds should initially be progeny tested in herds with low production if the thesis of Lush is correct, whereas if the thesis of Hammond is correct, sires to be used in low producing herds should be selected on the basis of their progeny test in high producing herds. The importance of the interaction between genotypes and environments, or more precisely, between sires and environments, is the major determining factor in considering these opposing theses.

A first approach to the problem was made by Falconer (1952). He proposed that the genetic correlation between the expression of the same genotype in two different environments could be used as a measure of the genotype-environment interaction. The estimation problem was later discussed by Robertson (1959) and Dickerson (1962), who pointed out that the interaction component has to be adjusted for the differences in genetic variation among environments used. Otherwise the non-additive genetic effects will cause a pseudo-interaction even if the actual genetic correlation were equal to one.

First and second lactation records of New York Holstein daughters sired artificially were used by Van Vleck (1963). The data were divided into four group depending on the level of the adjusted herd-mate averages relative to the season average. Genetic correlations between the genotypes in different environments were high (near to

unity) indicating that the genetic evaluation of the sires would be nearly the same in all herd levels.

In the study carried out by Danell (1982), the ratio between expected and calculated correlations of breeding values estimated at different levels of production was considered as an estimator of the genetic correlation between the expression of the same genotype in two environments. Such estimator of genetic correlation fluctuated between 0.90 and 0.98 indicating that estimated breeding values of dairy sires were not affected by an interaction between sire-herd.

Hill et al. (1983) estimated genetic correlations of sire performances for milk production traits in herd groups split according to mean, phenotypic variance, or coefficient of variation, using ANOVA type estimation methods. For all criteria of grouping herds and for all traits, genetic correlations did not differ significantly from unity.

De Veer and Van Vleck (1987) used a multivariate linear model to estimate the genetic correlation between true breeding of dairy sires at three levels of production defined by mean yield of all cows freshing in the same-year-season. Estimated correlations among sire value at three levels were large in all cases (all 0.85 or greater). The authors concluded that these results indicate that ranking of sires is not greatly affected by production level where daughters make their records.

In Cuba, Menendez-Buxadera et al. (1989) analysed a total of 16,622 first lactations of Holstein cows. The sample was divided into three production levels of milk yield. No changes in the order of merit of sires in the three levels were evidenced by the estimates of genetic correlations higher than 0.83.

Dong and Mao (1990) estimated genetic correlations between sire performances in three herd groups (low, medium, high) split according to mean herd milk production level or herd standard deviation, for three separate 4-year time periods. All estimates were larger than 0.95, and for the most recent period (1984 to 1987) estimates of genetic correlations across herd groups were essentially unity.

Estimates of genetic correlations for milk yield, milkfat yield and fat content between pairs of states in the USA were reported by Carabaño et al. (1990). The lowest correlation was 0.93 (for milkfat yield between California and Wisconsin), and estimates

for fat content exceeded 0.98.

Hence, although there are several studies that report heterogeneity of variance, the above results indicate that the prediction of breeding value for milk and milkfat yield of sires is not largely affected.

3.8.2 Methods to Account for Heterogeneous Variance

BLUP method can account for heterogenous variance if heritability and phenotypic variance were known in each environment (herd-year groups for example) (Henderson, 1984), but without such knowledge alternative methods have been proposed to account for heterogeneous variances. A brief discussion of these methods is given in this section.

3.8.2.1 *Logarithm Transformation*

When standard deviation is a simple linear function of the mean, that is, the correlation between them is a unity, logarithmic transformation is the appropriate transformation to correct for heterogenous variance (Bartlett, 1947; Everett and Keown, 1984; Garrick and Van Vleck, 1987; Visscher and Hill, 1992). However, the relationship between the mean and standard deviation is often far from the ideal and some times, logarithm transformation can "overadjust" the data for heterogeneity of variance making that the correlation between the mean and phenotypic variance change from a positive value to a negative value (Visscher et al. 1991; Kachman and Everett, 1993). In addition, a linear model in the transformed scale is not linear in the original scale (Kachman and Everett, 1993).

Everett and Keown (1984) reported a positive relationship between mean and variance for milk yield of 0.24 in Holstein herds in the United States. It was suggested that a log transformation of the data could remove a large part of the relationship between the mean and the variance. When log transformation was done, the results showed that the relationship between mean and variance was reduced.

Visscher et al. (1991) estimated genetic and environmental variances for fat yield for large Holstein Friesian pedigree herds in the UK, using an Animal Model. The

correlation between herd means and the estimated herd phenotypic standard deviation was 0.71, the correlation on the log scale changed to -0.28. The authors concluded that the log transformation slightly "overadjusted" the data for heterogeneity of variance.

A consequence of overadjusting by the log transformation would then be that more animals (e.g. bull dams) would be selected from the herds of low means reducing the efficiency of selection as shown in section (3.8.1.2). Hence, as pointed out by others (Garrick and Van Vleck, 1987; Boldman and Freeman, 1990), one should therefore be careful in applying a log transformation for genetic evaluation purposes in dairy cattle.

3.8.2.2 *Scaling*

When heritabilities are the same in all environments, scaling observations by the estimated phenotypic standard deviation is an appropriate method of handling heterogeneous variance (Hill, 1984). Hill (1984) suggested the following transformation

$$Z_{ij} = \frac{(y_{ij} - \bar{y}_i)}{\hat{\sigma}_{y_i}} \qquad (3.18)$$

where y_{ij} is the record made by the animal j in the herd-year group i; $\bar{y}_i$ is the average of the herd-year group i; $\hat{\sigma}_{y_i}$ is the phenotypic standard deviation in the herd-year group i; and Z_{ij} is the transformed record. This method has the advantage that no information outside the actual data is required and that the expected accuracy, that is the correlation between Z_{ij} and the true breeding value, does not depend on variation in σ_y.

The efficiency of this method to reduce the heterogeneity of variance can, presumably, be increased by using prior information on the standard deviations of the individual groups or on the variability amongst these standard deviations (Hill, 1984). Thus if prior values, $\sigma^*_{y_i}$, of each σ_{y_i} are available, phenotypic deviations can be expressed as

$$Z_{ij} = \frac{(y_{ij} - \bar{y}_i)}{\sigma^*_{y_i}}$$

without estimating the standard deviation.

Alternatively, Hill (1984) suggested that records would be standardised by a posterior standard deviation, $\tilde{\sigma}_{y_i}$, which is basically a weighting between the "within-herd" and "population" standard deviation. This is estimated as:

$$\tilde{\sigma}_{y_i} = \frac{\dfrac{\sigma_y}{\text{var}(\sigma_{y_i})} + \dfrac{\hat{\sigma}_{y_i}}{\text{var}(\hat{\sigma}_{y_i})}}{\dfrac{1}{\text{var}(\sigma_{y_i})} + \dfrac{1}{\text{var}(\hat{\sigma}_{y_i})}} \qquad (3.19)$$

where

σ_y = mean population standard deviation,

$\text{var}(\sigma_{y_i})$ = the variance of σ_{y_i} about σ_y,

$\hat{\sigma}_{yi}$ = the estimated standard deviation for the herd i,

$\text{var}(\hat{\sigma}_{y_i})$ = the variance of $\hat{\sigma}_{y_i}$ about σ_{y_i}.

Once obtained these posterior standard deviations, the records would be standardised as

$$Z_{ij} = \frac{(y_{ij} - \bar{y}_i)}{\tilde{\sigma}_{y_i}}$$

3.8.2.3 *Weighting*

Brotherstone and Hill (1986) showed that within-herd variances are often heterogeneous across years, so that, they suggested that heterogeneous variance can be accounted for by a weighted average of the estimated within herd-year phenotypic standard deviation and the estimated population standard deviation. Visscher et al. (1991) applied this correction in the following form

$$y_{ij}^{c} = y_{ij}(\sigma_{y_p} / \hat{\sigma}_{y_i})$$

where y_{ij} is the record made by the animal j in the herd-year group i, $\hat{\sigma}_{y_i}$ is the estimated phenotypic standard deviation in the herd-year group i; σ_{y_p} is the estimated population phenotypic standard deviation, and y_{ij}^{c} is the adjusted record.

These authors analysed data of 26 large Holstein Friesian pedigree herds in the UK using an animal model which included the fixed effect of herd-year-season (HYS). They found a correlation between the HYS means and estimated HYS phenotypic standard deviation of 0.71 and that estimates of heritability are relatively constant across the HYS groups. They also found that the correction for the heterogeneity of phenotypic variance, by adjusting data for the above transformation, reduced the heterogeneity substantially.

Wiggans and VanRaden (1991) proposed a method where lactation records are standardised for differing genetic and error variances across herds and over time based on phenotypic variance for each herd-year-parity group. Each herd-year-parity phenotypic variance estimate was combined with those of adjacent years and regressed toward a region-year-parity phenotypic variance in the following form.

For observations in region (R) i, herd (H) j within region i, year (Y) k, and parity (P) l,

$$\tilde{\sigma}^2_{ijkl} = \frac{20\hat{\sigma}^2_{ikl} + \frac{\nu_{ij(k-1)l}}{2}\hat{\sigma}^2_{ij(k-1)l} + \frac{\nu_{ij(k+1)l}}{2}\hat{\sigma}^2_{ij(k+1)l} + \nu_{ijkl}\hat{\sigma}^2_{ijkl}}{20 + \frac{\nu_{ij(k-1)l}}{2} + \frac{\nu_{ij(k+1)l}}{2} + \nu_{ijkl}}$$

where

$\tilde{\sigma}^2_{ijkl}$ = posterior estimate of the within-RHYP variance,

$\hat{\sigma}^2_{ikl}$ = estimated (across-herd) variance in the region-year-parity (RYP) (i, k, l),

$\hat{\sigma}^2_{ij(k-1)l}$ = estimated variance within RHYP (i, j, k-1, l),

$\hat{\sigma}^2_{ij(k+1)l}$ = estimated variance within RHYP (i, j, k+1, l),

$\hat{\sigma}^2_{ijkl}$ = estimated variance within RHYP (i, j, k, l),

$\nu_{ij(k-1)l}$ = degrees of freedom for RHYP (i, j, k-1, l),

$\nu_{ij(k+1)l}$ = degrees of freedom for RHYP (i, j, k+1, l),

ν_{ijkl} = degrees of freedom for RHYP (i, j, k, l).

From the relationship between heritability and phenotypic variance, estimates of genetic variance were obtained. Lactations records were deviated from management group mean, and the deviation was multiplied by the ration of a estimate of population standard deviation to RHYP genetic standard deviation.

This method of adjusting by heterogeneity of variance was implemented in July 1991 for the national USDA animal model genetic evaluation (Wiggans and VanRaden, 1991). Estimated genetic trend for milk increased by nearly 5 kg/yr for Holstein with this adjustment, which caused predicted breeding values of oldest animals to be lower by about 100 kg. Cows in high variance herds were most likely to have large reductions in their evaluations.

3.8.2.4 *Multiple Trait Analysis*

In general, homogeneity of variance is not a requirement in a best linear unbiased prediction analysis. Treatment of random effects as separate traits in different environments allows handing heterogeneous variances as a multiple trait analysis (Henderson, 1984; Gianola, 1986; Garrick and Van Vleck, 1987). However, obtaining separate estimates of sufficient accuracy for variance and covariance components in each environment may be difficult because large data set are required.

To account for sampling errors of parameter estimates, Hill (1984), Brotherstone and Hill (1986) and Visscher and Hill (1992) proposed a Bayesian procedure in which individual herd parameters are regressed to an overall a priori estimate. The regression coefficient depends on the sampling variances of individuals herd estimates and the variance of the parameters, assumed to be known *a priori*. The regression may be written (Visscher and Hill, 1992) as

$$\tilde{\theta}_i = \hat{\theta}_o + \beta_i(\hat{\theta}_i - \hat{\theta}_o) \qquad (3.20)$$

where $\hat{\theta}_i$ is parameter estimate and $\tilde{\theta}_i$ its regressed estimate for herd i, and $\hat{\theta}_o$ is the overall (prior) estimate. The regression coefficient is

$$\beta_i = \frac{1}{1+\gamma_i}$$

with

$$\gamma_i = \frac{\mathrm{var}(\hat{\theta}_i|\theta_i)}{\mathrm{var}(\theta_i)}$$

being the ratio of the sampling variance to the variance of the parameter, or less formally, the ratio of variance "within" and "between" parameters θ.

Gianola et al. (1992) and Weigel and Gianola (1993) gave a formal empirical Bayesian method for estimation of heterogeneous variance. Essentially, they proposed that individual herd (genetic, environmental and phenotypic) variances can be estimated by

$$\tilde{\sigma}_i^2 = \frac{d_i\hat{\sigma}_i^2 + d_o\sigma_o^2}{d_i + d_o} = \sigma_o^2 + \frac{d_i}{d_i + d_o}(\hat{\sigma}_i^2 - \sigma_o^2)$$

where

$\tilde{\sigma}_i^2$ = the posterior estimate of variance in herd i,

$\hat{\sigma}_i^2$ = the estimated variance in herd i,

σ_o^2 = the average population value,

d_o = the "degrees of believe" of the average population value,

d_i = degrees of freedom for herd i.

3.8.2.5 *Multiplicative Model*

A multiplicative mixed model that incorporates scaling factors was presented by Kachman and Everett (1993). This model differs from the usual mixed models in that the fixed and random effects are scaled by the scaling factor of each environment.

3.8.3 Example of the Brotherstone and Hill Method

In section (3.8.2.2) the method of scaling as suggested by Hill (1984) was explained (Equation 3.18). Alternatively Hill (1984) suggested that records would be standardised by a posterior standard deviation (Equation 3.19). In this example, the method to obtain a posterior standard deviation is illustrated following the principles derived by Brotherstone and Hill (1986).

Consider milkfat yields of first calving cows in nine herds (Table 3.4). A suitable model for these data may be

$$y_{ij} = h_i + e_{ij}$$

where y_{ij} is the milkfat yield of the j animal in the herd i, h_i is the fixed effect of the herd i and e_{ij} is the residual error corresponding to y_{ij}.

Table 3.4. Milkfat yield of first calving cows in nine herds for the example of the Brotherstone and Hill method to correct for heterogeneity of variance.

Herd number								
1	2	3	4	5	6	7	8	9
135	141	155	138	153	171	143	157	177
130	145	149	141	163	168	142	158	181
133	137	161	145	159	173	145	161	194
128	147	159	142	163	175	149	167	175
135		168		155	169		164	189
		153			165			193
					173			

The within-herd mean($\bar{y}_i$), variance ($\hat{\sigma}^2_{yi}$) and standard deviation ($\hat{\sigma}_i$) for each herd are calculated using the following formulas:

$$\bar{y}_i = \frac{\sum_{i=1}^{n_i} y_i}{n_i}, \tag{3.21}$$

$$\hat{\sigma}^2_{y_i} = \frac{\sum_{i=1}^{n_i} y_i^2 - \frac{\left(\sum_{i=1}^{n_i} y_i\right)^2}{n_i}}{n_i - 1}, \tag{3.22}$$

$$\hat{\sigma}_{y_i} = \sqrt{\hat{\sigma}^2_{y_i}}. \tag{3.23}$$

Numerical values for of Equations (3.21)-(3.23) for each herd are shown in Table

3.5. The terms involved to obtain the posterior standard deviation (Equation 3.19) are obtained as follows.

An estimator of the population standard deviation, σ_y, is approximated by the unweighted mean of within-herd standard deviations

$$\sigma_y = \bar{\hat{\sigma}}_{y_i} = \frac{\sum_{i=1}^{k} \hat{\sigma}_{y_i}}{k} \tag{3.24}$$

$$\sigma_y = \bar{\hat{\sigma}}_{y_i} = \frac{3.11+4.43+...+8.26}{9} = 4.52,$$

with estimated variance

$$\hat{var}(\sigma_{y_i}) = \frac{\sum_{i=1}^{k}[(\hat{\sigma}_{y_i} - \bar{\hat{\sigma}}_{y_i})^2 - \frac{\bar{\hat{\sigma}}^2_{y_i}}{2n_i}]}{k-1} \tag{3.25}$$

The term $\bar{\hat{\sigma}}^2_{y_i}$ being the unweighted average of within-herd variances, i.e.,

$$\bar{\hat{\sigma}}^2_{y_i} = \frac{9.70+19.67+...+68.17}{9} = 23.36$$

so that,

$$\hat{Var}(\sigma_{y_i}) = \frac{\left[(3.11-4.52)^2 - \frac{23.36}{2(5)}\right] + \left[(4.43-4.52)^2 - \frac{23.36}{2(4)}\right] + ... + \left[(8.26-4.52)^2 - \frac{23.36}{2(6)}\right]}{9-1}$$

$$= 5.27.$$

And the variance of each within-herd estimated standard deviation is obtained as:

$$\text{var}(\hat{\sigma}_{yi}) \approx \frac{\sigma_y^2}{2n_i} \approx \frac{\bar{\hat{\sigma}}_y^2}{2n_i}. \quad . \tag{3.26}$$

Table 3.5. Estimates of phenotypic standard deviations for herds considered in the example of the Brotherstone and Hill method to correct for heterogeneity of variance.

Statistic	Number of Equation*	Herd number								
		1	2	3	4	5	6	7	8	9
$\bar{y}_i$	(3.21)	132.2	142.5	157.5	141.5	158.6	170.6	144.8	161.4	184.8
n_i		5	4	6	4	5	7	4	5	6
$\hat{\sigma}^2_{y_i}$	(3.22)	9.70	19.67	44.70	8.33	20.80	11.95	9.58	17.30	68.17
$\hat{\sigma}_{y_i}$	(3.23)	3.11	4.43	6.69	2.89	4.56	3.46	3.10	4.16	8.26
$\text{var}(\hat{\sigma}_{yi})$	(3.26)	2.04	2.55	1.70	2.55	2.04	1.46	2.55	2.04	1.70
$\tilde{\sigma}_{y_i}$	(3.19)	3.50	4.46	6.16	3.42	4.55	3.69	3.56	4.26	7.34

* Number of equation in the text.

Thus, calculated values of posterior standard deviations, $\tilde{\sigma}_{y_i}$, would be used to standardise deviations of individual records, y_{ij}, from within-herd means, $\bar{y}_i$, as suggested by Hill (1984) in Equation (3.18).

CHAPTER 4
MATERIAL AND METHODS

4.1 Material

Lactation records of dairy cows calving between 1986 and 1989 inclusive, were obtained from the Livestock Improvement Corporation files of the New Zealand Dairy Board. In total there were 2,004,854 lactation milkfat yields in the file, which had originally been set up for the purpose of sire evaluation. A pedigree file was available containing information of ancestors as well as number of register and breed. Cows were herd located in different areas of the country and all lactations were considered. Year and month of calving, as well as age of the cow at calving were available for each production record. Additional information such as if the calving was induced, was also available.

4.2 Methods

4.2.1 Estimation of Breeding Values

An increasing proportion of the dairy cattle population in New Zealand has been crossbred since 1960, mainly with the aim of changing from straightbred Jerseys to Holstein-Friesian. Thus, some farmers are choosing sires across breeds, on the basis of breeding index irrespective of the breed. Further, some studies (Nejati-Javaremi, 1991) have given evidence that genetic evaluation of sires may be across breeds.

Breeding values of sires were estimated through BLUP procedure (Henderson, 1973) by using a single trait repeatability animal model across breeds, which is currently being developed for the future genetic evaluation of dairy cattle in New Zealand (Garrick et al. 1993; Harris et al. 1993). The statistical model for analysis of a cow with milkfat yield was as follows (Harris et al. 1993):

$$y_{ijklmn} = HYA_i + M_j + D_k + a_m + \sum_{r=1}^{g_g} q_{mr} g_r + p_m + w_{ml} h_l + \sum_{t=1}^{b_g} s_{mt} b_t + e_{ijklmn} \quad (4.1)$$

where

y_{ijklmn} is the n^{th} milkfat production record made by the m^{th} cow in the i^{th} herd-year-age contemporary group of the l^{th} heterosis class calving in the j^{th} calendar month and in the k^{th} induced lactation class,

HYA_i is the fixed effect of the i^{th} herd-year-age contemporary group,

M_j is the fixed effect of the j^{th} calendar month of calving,

D_k is the fixed effect for the k^{th} induced lactation class,

a_m is the random additive genetic effect for the m^{th} animal;

q_{mr} is the fractional contribution of the r^{th} genetic group to the genetic merit of the m^{th} animal,

g_r is the fixed effect for the r^{th} genetic group with g_g classes;

p_m is the random non-additive genetic and permanent environment effect associated with records of the m^{th} animal; and

w_{ml} is the contribution of the l^{th} heterosis class of the m^{th} cow,

h_l is the fixed effect of the l^{th} heterosis class,

s_{mt} is the contribution of the t^{th} maternal breed class of the m^{th} cow,

b_t is the fixed effect for the t^{th} maternal breed with b_g classes,

e_{ijklmn} is the random residual.

Contemporary groups were defined as cows of the same age calving in the same herd-year. An effective contemporary group (HYA) was defined as that with more than 5 animals.

Fixed effect of month of calving was included in the model because it is known that, for the new Zealand milk production system, month of calving has an important effect on milk yield. Cows calving earlier than the average calving date are likely to have

longer lactations than cows calving later than the average calving date. However, the later are likely to be well fed in early lactation and can overcome the shorter lactations (Holmes, 1986).

Dairy farmers use calving induction with the objective of maintaining a concentrated calving interval. Cows with induced calving are likely to be vulnerable to the effects of stress due to weather and infections. They are also likely to produce less milk and to have low fertility. Owing to these reasons, the fixed effect of induced lactation was considered in the model.

Genetic groups were considered as developed by Robinson (1986), Westell and Van Vleck (1987), Quaas (1988) and Westell et al. (1988). For each animal with unknown ancestors, phantom parents without records were created and were considered unrelated. An animal was assigned up to four genetic groups depending on the number of breeds present in the animal (two parents unknown) or the number of breeds present in the unknown parent.

The expected coefficients for heterosis were computed for specific crossbreeding designs (Dickerson, 1973; Ahlborn-Breier and Hohenboken, 1991).

Breed group effects were considered as defined by Arnold et al. (1992). Each animal could have up to four breed codes and the proportions of the genes from each of the four breed stored on database. Eight breeds have been assigned to the New Zealand population:

1. Holstein-Friesian
2. Jersey
3. Ayrshire
4. Guernsey
5. Milking Shorthorn
6. European Red breeds
7. Brown Swiss
8. Other breeds including beef breeds.

Each animal could have up to four maternal breed classes with the coefficient for each class being the percentage of genes belonging to that breed class for the dam of the animal.

In matrix notation, model equation (4.1) can be rewritten as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{ZQg} + \mathbf{Zp} + \mathbf{ZWh} + \mathbf{ZSm} + \mathbf{e} \qquad (4.2)$$

where

y	is	the vector of records;
b	is	the vector of herd-season-age, month of calving and induced lactation class fixed effects;
a	is	the vector of random additive genetic effects including animals without records;
g	is	the vector of genetic group effects;
p	is	the vector of random non-additive genetic and permanent environment effects;
h	is	the vector of heterosis fixed effects;
m	is	the vector of maternal breed fixed effects;
X, Z, Q W and **S**	are	incidence matrices associating records with the elements of **b**, **a**, **g**, **h**, and **m**, respectively; and
e	is	the vector of random residuals.

The rows of **Z** associated with animals with no records contains all zero elements but are incorporated in the inverse of relationship matrix. The i^{th} row elements of **Q** contain the contribution of each genetic group to the genetic merit of animal m. The i^{th} row elements of **W** contain the proportion of heterosis expected for animal m for the appropriate breed combination. And, the i^{th} row elements of **S** contain the proportions of each breed of the dam of animal m.

The expectations and variances are:

$$E\begin{bmatrix}\mathbf{y}\\ \mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{Xb} + \mathbf{ZWh} + \mathbf{ZSm} + \mathbf{ZQg}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

and

$$\text{Var}\begin{bmatrix}\mathbf{y}\\ \mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix} \mathbf{ZAZ}'\sigma_a^2 + \mathbf{ZZ}'\sigma_p^2 + \mathbf{R}\sigma_e^2 & \mathbf{ZA}\sigma_a^2 & \mathbf{Z}\sigma_p^2 & \mathbf{R}\sigma_e^2 \\ & \mathbf{ZAZ}'\sigma_a^2 & \mathbf{0} & \mathbf{0} \\ \text{symmetric} & & \mathbf{ZZ}'\sigma_p^2 & \mathbf{0} \\ & & & \mathbf{R}\sigma_e^2 \end{bmatrix}$$

where σ_a^2 is the additive genetic variance, σ_p^2 is the non-additive genetic plus permanent environment variance, σ_e^2 is the residual variance, **A** is the numerator relationship matrix and **R** is a diagonal matrix.

The mixed model equations for the equation model (4.2) are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{ZQ} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{ZW} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{ZS} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\lambda_a\mathbf{A}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZQ} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZW} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZS} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZQ} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZW} & \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZS} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZQ} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\lambda_p\mathbf{I} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZW} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZS} \\ \mathbf{W}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZQ} & \mathbf{W}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZW} & \mathbf{W}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZS} \\ \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZQ} & \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZW} & \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{ZS} \end{bmatrix} \begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{a}}\\ \hat{\mathbf{g}}\\ \hat{\mathbf{p}}\\ \hat{\mathbf{h}}\\ \hat{\mathbf{m}}\end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{4.3}$$

where $\lambda_a = \dfrac{\sigma_e^2}{\sigma_a^2}$ and $\lambda_p = \dfrac{\sigma_e^2}{\sigma_p^2}$

The mixed model equations augmented to include the phantom parents without records (Westall et al. 1988) are:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & 0 & X'R^{-1}ZQ & X'R^{-1}Z & X'R^{-1}ZW & X'R^{-1}ZS \\ Z'R^{-1}X & Z'R^{-1}Z+\lambda_a A^{11} & \lambda_a A^{10} & Z'R^{-1}ZQ & Z'R^{-1}Z & Z'R^{-1}ZW & Z'R^{-1}ZS \\ 0 & \lambda_a A^{01} & \lambda_a A^{00} & 0 & 0 & 0 & 0 \\ Q'Z'R^{-1}X & Q'Z'R^{-1}Z & 0 & Q'Z'R^{-1}ZQ & Q'Z'R^{-1}Z & Q'Z'R^{-1}ZW & Q'Z'R^{-1}ZS \\ Z'R^{-1}X & Z'R^{-1}Z & 0 & Z'R^{-1}ZQ & Z'R^{-1}Z+\lambda_p I & Z'R^{-1}ZW & Z'R^{-1}ZS \\ W'Z'R^{-1}X & W'Z'R^{-1}Z & 0 & W'Z'R^{-1}ZQ & W'Z'R^{-1}Z & W'Z'R^{-1}ZW & W'Z'R^{-1}ZS \\ S'Z'R^{-1}X & S'Z'R^{-1}Z & 0 & S'Z'R^{-1}ZQ & S'Z'R^{-1}Z & S'Z'R^{-1}ZW & S'Z'R^{-1}ZS \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a}_1 \\ \hat{a}_0 \\ \hat{g} \\ \hat{p} \\ \hat{h} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ 0 \\ Q'Z'R^{-1}y \\ Z'R^{-1}y \\ W'Z'R^{-1}y \\ S'Z'R^{-1}y \end{bmatrix} \tag{4.4}$$

where $\mathbf{A}^{11}$, $\mathbf{A}^{10}$, $\mathbf{A}^{01}$ and $\mathbf{A}^{00}$ are submatrices of the inverse of the numerator relationship matrix including the phantom parents, i.e.,

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{10} \\ \mathbf{A}^{01} & \mathbf{A}^{00} \end{bmatrix}.$$

Solving the mixed model equations (4.4), the vector of estimates of breeding values, $\hat{\mathbf{u}}$, would be derived as indicated by Quaas (1988)

$$\hat{\mathbf{u}} = \text{BLUP}(\mathbf{u}) = \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{a}}.$$

A solution for $\hat{\mathbf{a}}$ in this form, however, would be computationally difficult. Further, inclusion of phantom parents in the equations increases the number of equations to solve. The solution vector of estimates of breeding values would be obtained directly by using the QP transformation (Quaas and Pollak, 1981; Pollak and Quaas, 1983; Quaas, 1988) and because the solutions for the phantom parents are not needed, these would be absorbed into the QP transformed equations. This yields the following set of equations (Westall et al. 1988):

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & 0 & X'R^{-1}Z & X'R^{-1}ZW & X'R^{-1}ZS \\ Z'R^{-1}X & Z'R^{-1}Z+\lambda_a K_{11} & \lambda_a K_{12} & Z'R^{-1}Z & Z'R^{-1}ZW & Z'R^{-1}ZS \\ 0 & \lambda_a K_{12}' & \lambda_a K_{22} & 0 & 0 & 0 \\ Z'R^{-1}X & Z'R^{-1}Z & 0 & Z'R^{-1}Z+\lambda_p I & Z'R^{-1}ZW & Z'R^{-1}ZS \\ W'Z'R^{-1}X & W'Z'R^{-1}Z & 0 & W'Z'R^{-1}ZQ & W'Z'R^{-1}ZW & W'Z'R^{-1}ZS \\ S'Z'R^{-1}X & S'Z'R^{-1}Z & 0 & S'Z'R^{-1}ZQ & S'Z'R^{-1}ZW & S'Z'R^{-1}ZS \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \\ \hat{g} \\ \hat{p} \\ \hat{h} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ 0 \\ Z'R^{-1}y \\ W'Z'R^{-1}y \\ S'Z'R^{-1}y \end{bmatrix} \tag{4.5}$$

where

$$\mathbf{Q}_0 = \mathbf{A}_{10}\mathbf{Q}$$

$$\mathbf{K}_{11} = (\mathbf{A}^{11} - \mathbf{A}^{10}(\mathbf{A}^{00})^{-1}\mathbf{A}^{01}),$$

$$\mathbf{K}_{12} = \mathbf{A}^{10}(\mathbf{A}^{00})^{-1}\mathbf{Q}_0, \text{ and}$$

$$\mathbf{K}_{22} = \mathbf{Q}_0'\mathbf{Q}_0 - \mathbf{Q}_0'(\mathbf{A}^{00})^{-1}\mathbf{Q}_0 = \mathbf{Q}_0'[\mathbf{I} - (\mathbf{A}^{00})^{-1}]\mathbf{Q}_0$$

The blocks of the partitioned matrix **K**, i.e.,

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{12}' & \mathbf{K}_{22} \end{bmatrix}$$

are defined as a function of the blocks of the partitioned matrix $\mathbf{A}^{-1}$, the numerator relationship matrix inverse for phantom parents and animals. The rules for determining the individuals elements of **K** are given by Westell et al. (1988) and follow the rules for computing $\mathbf{A}^{-1}$ with both parents known as presented by Henderson (1976); except the magnitude of the coefficients depends on the number of phantom parents. The coefficient in $\mathbf{K}_{11}$ are equivalent to $\mathbf{A}^{-1}$.

Mixed model equations (4.5) were constructed and solved by Jacobi iteration for the fixed effects, **b**, and by Gauss-Seidel block iteration for the random effects.

The vector of estimates of producing ability, $\hat{\mathbf{q}}$, are estimated as the sum of the vector of estimated breeding values and the vector of estimated permanent effects (Van Vleck, 1979), i.e.,

$$\hat{\mathbf{q}} = \text{BLUE}(\mathbf{q}) = \hat{\mathbf{u}} + \hat{\mathbf{p}}.$$

4.2.2 Definition of Levels of Production

After editing, there were 83,805 effective contemporary groups. For each contemporary group (HYA), the mean ($\bar{y}_i$), the within-HYA variance ($\hat{\sigma}^2_{y_i}$) and the within-HYA standard deviation ($\hat{\sigma}_{y_i}$) for milkfat yield were calculated. Because no other information about management factors, the HYA mean was used as a grouping factor. Contemporary groups were sorted by the mean milkfat production and subsequently, they were divided into three equal sized groups which were identified as low, medium and high production levels.

For each producing level, breeding values of sires were estimated by the using the model equation (4.1). Thus, three independent estimates of breeding values of sires were obtained.

4.2.3 Methods to Reduce Heterogeneity of Variance

Methods to reduce heterogeneity of variance were investigated by using three different methods.

4.2.3.1 *Mean Correction*

Assuming that the relationship between the within-herd-year-age standard deviation and herd-year-age mean is directly proportional, the records were scaled by the corresponding herd-year-age mean as follows

$$y^{mu}_{ijklmn} = \frac{y_{ijklmn}}{\bar{y}_i} \qquad (4.6)$$

where y_{ijklmn} is a record of milkfat production as defined in equation (4.1), y^{mu}_{ijklmn} the scaled record, and $\bar{y}_i$ is the mean milkfat production for the i^{th} herd-year-age group.

4.2.3.2 *Standard Deviation Adjustment*

Milkfat yields were adjusted using a weighted combination of an estimate of the population standard deviation (σ_y) and a posterior herd-year-age standard deviation ($\tilde{\sigma}_{y_i}$) as follows

$$y^{sd}_{ijklmn} = y_{ijklmn}\left(\frac{\sigma_y}{\tilde{\sigma}_{y_i}}\right) \tag{4.8}$$

where y_{ijklmn} is a milkfat yield record as defined in the model equation (4.1) and y^{sd}_{ijklmn} is the adjusted record.

The population standard deviation was taken as the unweighted average of within-herd-year-age standard deviations

$$\sigma_y = \frac{\sum_{i=1}^{c_g} \hat{\sigma}_{y_i}}{c_g}$$

and $\tilde{\sigma}_{y_i}$ was obtained by using a Bayesian regression approach (see section 3.8.2.4 and example in section 3.8.3) (Hill, 1984; Brotherstone and Hill, 1986; Weigel and Gianola, 1993) as follows

$$\tilde{\sigma}_{y_i} = \frac{\frac{\sigma_y}{var(\sigma_{y_i})} + \frac{2\nu_i\hat{\sigma}_{y_i}}{\sigma_y^2}}{\frac{1}{var(\sigma_{y_i})} + \frac{2\nu_i\hat{\sigma}_{y_i}}{\sigma_y^2}}$$

where ν_i are the degree of freedom to estimate the within-HYA standard deviation, and $var(\sigma_{y_i})$ is the variance of σ_{y_i} about σ_y. $Var(\sigma_{y_i})$ define the extent of variability in within-herd variance. An estimate of this parameter was calculated as

$$var(\sigma_{y_i}) = \frac{\sum_{\substack{i=1 \\ i\neq j}}^{c_g} \hat{\sigma}^2_{y_i} - 2\bar{\hat{\sigma}}_{y_i} \sum_{\substack{i=1 \\ i\neq j}}^{c_g} \hat{\sigma}_{y_i} + (\bar{\hat{\sigma}}_{y_i})^2 [c_g - 1 - \sum_{\substack{i=1 \\ i\neq j}}^{c_g} \frac{1}{2\nu_i}]}{c_g - 2}$$

where $i \neq j$ means that the contemporary group i is not included, so that, $\bar{\hat{\sigma}}_{y_i}$ is the unweighted average excluding the i^{th} contemporary group.

An adjustment for herd-year-age standard deviation rather than for herd standard deviation was made because it is known that within-herd variances are often heterogeneous across years (Brotherstone and Hill, 1986), and because herd-year-age rather than herds were fitted as fixed effects in the genetic evaluation.

4.2.3.3 *Logarithmic Transformation*

The natural logarithm (ln) transformation of the observations were made

$$y^{ln}_{ijklmn} = \ln(y_{ijklmn}) \qquad (4.9)$$

where y^{ln}_{ijklmn} is the data after logarithmic transformation.

Estimated breeding values of sires were estimated after transformations (4.7), (4.8) and (4.9) by solving mixed model equations (4.5) for the entire data and for each independent subset based on level of production.

4.2.4 Measurement of the Effect of Heterogeneous Variance

4.2.4.1 *Relationship Between Mean and Standard Deviation*

Genetic evaluation is simplified when residual and genetic variances across environments are constant. This means a correlation of zero between standard deviation and mean. A significant correlation would indicate departure from this assumption (Everett and Keown, 1984) and as a result, estimation of breeding values may be biased. Correlation coefficients (SAS, 1985) between within-contemporary group means and standard deviations were investigated as evidence for heterogeneity of variance.

4.2.4.2 *Correlations Between Breeding Values*

Suppose a sire contributes genes affecting records in each of two distinct environments. Now, this sire could be evaluated separately in each environment from the records obtained in that environment. Bereskin and Lush (1965), Hickman et al. (1969), Garrick (1988) and Notter and Diaz (1993) have discussed the expected value of the correlation between the breeding values estimated in different environments for the same sire.

As discussed in section (3.8.1.2), but consider now the simple genetic in Equation (3.1), $y_{ijk} = h_i + u_j + e_{ijk}$, where, y_{ijk} is the milkfat yield of the k^{th} daughter of the j^{th} sire in the i^{th} environment, h_i is the fixed environmental effect and u_j and e_{ijk} are random genetic and residual effects respectively. Define $\hat{u}_{1j}$ as an estimate of the sire effect obtained from records on n_{1j} daughters in environment 1 and $\hat{u}_{2j}$ as an estimate of the sire effect from records on n_{2j} daughters in environment 2. If genetic and residual variances are known in each environment, the appropriate predictions would be obtained using selection index theory. Ignoring relationships between sires, $\hat{u}_{ij} = b_{ij}\bar{y}_{ij}$ for some

appropriate b_{ij} and where $\bar{y}_{ij}$ is the average of records of n_{ij} daughters of sire j in herd i adjusted for fixed non-genetic effects, so that, $\bar{y}_{ij} = u_{ij} + e_{ijk}$.

Now as shown by Garrick (1988) and Garrick et al. (1989), the correlation between two estimates of breeding values for the sire j is

$$r_{\hat{u}_{1j}\hat{u}_{2j}} = \frac{\mathrm{cov}(\hat{u}_{1j}, \hat{u}_{2j})}{\sqrt{\mathrm{var}(\hat{u}_{1j}) \cdot \mathrm{var}(\hat{u}_{1j})}}$$

$$= \frac{\mathrm{cov}(b_{1j}\bar{y}_{1j}, b_{2j}\bar{y}_{2j})}{\sqrt{\mathrm{var}(b_{1j}\bar{y}_{1j}) \cdot \mathrm{var}(b_{2j}\bar{y}_{2j})}}$$

$$= \frac{\mathrm{cov}(\bar{y}_{1j}, \bar{y}_{2j})}{\sqrt{\mathrm{var}(\bar{y}_{1j}) \cdot \mathrm{var}(\bar{y}_{2j})}}$$

$$= \frac{\mathrm{cov}(u_{1j}, u_{2j})}{\sqrt{\mathrm{var}(u_{1j}) \cdot \mathrm{var}(u_{2j})}} \cdot \frac{\sqrt{\mathrm{var}(u_{1j}) \cdot \mathrm{var}(u_{2j})}}{\sqrt{\mathrm{var}(\bar{y}_{1j}) \cdot \mathrm{var}(\bar{y}_{2j})}}$$

$$= r_{u_{1j}u_{2j}} \sqrt{\frac{\mathrm{var}(u_{1j}) \cdot \mathrm{var}(u_{2j})}{\mathrm{var}(y_{1j}) \cdot \mathrm{var}(y_{2j})}} \qquad (4.10)$$

Now consider the correlation between a true genetic effect and an estimate of the same effect based on the corresponding observation:

$$r_{u_{ij}\hat{u}_{ij}} = r_{u_{ij}(b_{ij}\bar{y}_{ij})}$$

$$= \frac{\mathrm{cov}(u_{ij}, b_{ij}u_{ij} + b_{ij}e_{ijk})}{\sqrt{\mathrm{var}(u_{ij}) \cdot \mathrm{var}(b_{ij}\bar{y}_{ij})}} = \frac{\mathrm{var}(u_{ij})}{\sqrt{\mathrm{var}(u_{ij}) \cdot \mathrm{var}(\bar{y}_{ij})}} = \sqrt{\frac{\mathrm{var}(u_{ij})}{\mathrm{var}(\bar{y}_{ij})}}$$

using this in (4.10) results in expression

$$r_{\hat{u}_{1j}\hat{u}_{2j}} = r_{u_{1j}u_{2j}} r_{u_{1j}\hat{u}_{1j}} r_{u_{2j}\hat{u}_{2j}} \tag{4.11}$$

The identity in (4.11) means that the expected correlation between breeding values of sires estimated in different environments ($r_{\hat{u}_{1j}\hat{u}_{2j}}$) depends on the true genetic correlation ($r_{u_{1j}u_{2j}}$) and the accuracy of breeding value prediction in those environments ($a_{1j} = r_{u_{1j}\hat{u}_{1j}}$ and $a_{2j} = r_{u_{2j}\hat{u}_{2j}}$). This identity is true under certain conditions. These include (Taylor, 1983 as cited by Notter and Diaz, 1993):

1. No environmental correlation between performance in the different environments.

2. No relationships among parents of measured animals.

3. No other covariances among predicted breeding values within either environment.

4. Sires are chosen at random.

For sire evaluation, with these assumptions,

$$a_{ij}^2 = \frac{n_{ij}}{n_{ij} + \lambda_i}$$

$$a_{ij} = \sqrt{\frac{n_{ij}}{n_{ij} + \lambda_i}} \tag{4.12}$$

where λ_i is the ratio of residual to sire variance for environment i. Under BLUP, accuracies of estimated breeding values for non-inbred animals are obtained as:

$$a_{ij} = \sqrt{\frac{1 - c^{ii}}{\sigma_a^2}}$$

where c^{ii} is the i^{th} diagonal element of $\mathbf{C}^{uu}$, the prediction error covariance matrix of $\hat{\mathbf{u}}$ (see section 3.3). If the model is complete and properly parameterised, accuracies are expected to equal correlations between actual and predicted breeding values.

Assuming that the true genetic correlation between breeding values is the unity (i.e., the genes for both environments are the same) then the correlation between estimated breeding values for a sire is

$$r_{\hat{u}_{1j}\hat{u}_{2j}} = 1 \cdot r_{u_{1j}\hat{u}_{1j}} r_{u_{2j}\hat{u}_{2j}}$$

$$r_{\hat{u}_{1j}\hat{u}_{2j}} = a_{1j} \cdot a_{2j}$$

$$r_{\hat{u}_{1j}\hat{u}_{2j}} = \sqrt{\frac{n_{1j}}{n_{1j} + \lambda_1}} \sqrt{\frac{n_{2j}}{n_{2j} + \lambda_2}}.$$

If heritability of the trait is 0.25 and is constant through the environments, then $\lambda_i = \sigma_e^2 / \sigma_s^2 = 15$. This means that residual and sire variance are constant or vary proportionally through environments. Using these results, the expected correlation between estimated breeding values of sample of m sires evaluated in different environments (Hickman et al. 1969) is:

$$r_{\hat{u}_{1j}\hat{u}_{2j}} = r_e = \frac{\left[\sum_{j=1}^{m}\left(\frac{n_{1j}}{n_{1j}+15}\right)\cdot\left(\frac{n_{2j}}{n_{2j}+15}\right)\right]}{\sqrt{\sum_{j=1}^{m}\left(\frac{n_{1j}}{n_{1j}+15}\right)\cdot\sum_{j=1}^{m}\left(\frac{n_{2j}}{n_{2j}+15}\right)}}.$$

Observed correlations (r_0) between breeding values estimated in two environment, would be compared to their expected correlations. Expected confidence limits for observed correlations would be used to evaluate if significant departures exist from their expectations and this would be used a measure of the effect of heterogeneous variance.

For correlation analysis, the statistic of Fisher (Snedecor and Cochran, 1980)

$$Z = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right]$$

has variance of $\approx (m-3)^{-1}$, where r is the observed correlation and m is the number of sires in the sample. By converting the observed correlation coefficient to Z using this transformation, a confidence interval for μ_z is given by

$$L = Z - Z_{\alpha/2} \cdot \frac{1}{\sqrt{m-3}} \text{ and}$$

$$U = Z + Z_{\alpha/2} \cdot \frac{1}{\sqrt{m-3}}$$

as the lower and upper limits. The value $Z_{\alpha/2}$ is obtained from the standard normal table. The limits on μ_z, L and U, may then be converted to limits on the parameter by using tables provided in Pfaffenberger and Patterson (1977).

Product-moment and rank correlations between the estimated breeding values at the three different levels of production for untransformed and transformed data for sires having 10, 20, 30, etc. up to 100 daughters at each level of production were calculated. Pearson's correlation coefficient (r_p) for product-moment correlation and Spearman's correlation coefficient (r_s) for rank correlation were obtained by using SAS (1985) and confidence intervals were calculated for these coefficients.

4.2.4.3 *Significant Deviations*

Estimates of breeding values in each level of production will never give exactly

the same results, unless the information for each sire is large enough and assumptions in the model are satisfied (homogeneity of variance). The appearance of significantly different results between estimates of breeding values for sire j ($\hat{u}_{1j} - \hat{u}_{2j}$) calculated on two different levels of production would in that case be an indication of heterogeneous variance effect. The standard deviation ($sd_{\hat{u}_{1j}-\hat{u}_{2j}}$) of the differences between estimates of breeding values for the same sire can be calculated (Danell, 1982) as:

$$sd_{\hat{u}_{1j}-\hat{u}_{2j}} = \sqrt{\left[a_{1j}^2(1-a_{2j}^2)\sigma_a^2 + a_{2j}^2(1-a_{1j}^2)\sigma_a^2\right]}$$

where σ_a^2 is the genetic additive variance and a_{1j} and $a_{2j}$is the accuracy of breeding value prediction in the two levels of production as defined in Equation (4.12). This expression assumes that the errors in the two estimates are uncorrelated.

It was assumed that heritability was the same across production levels but estimated of the genetic additive variance were obtained as $\sigma_a^2 = h^2\sigma_y^2$ by using estimated of phenotypic variances. The number of significant differences between the estimated breeding values for the same sires outside a 95% confidence interval were counted and compared to the expected number. The same procedure was made for each set of data (corrected by the mean, adjusted by the standard deviation and logarithmic transformation).

4.2.5 Genetic Correlation

An estimate ($\hat{r}_G$) of the correlation between true breeding values ($r_{u_{1j}u_{2j}}$) may be obtained from the ratio of the observed and the expected correlations between the expression of the same genotype in two environments (Mason and Robertson, 1956; Danell, 1982; Notter and Diaz, 1993).

CHAPTER 5
RESULTS

5.1 Means and Standard Deviations and their Correlation

There were 83,805 contemporary groups (cows of the same age calving in the same herd-year) which were split into three equal sized groups (27,935) according to the mean contemporary milkfat yield, namely low, medium, and high producing levels. The grouping of records produced differences in lactation milkfat yield as shown in Table 5.1. The differences in milkfat yield between medium-low, high-low and high-medium production levels were 12.79, 32.97 and 20.18 kg, respectively. The corresponding differences in averages of standard deviations were: 1.02, 3.33, and 2.31. Thus, standard deviation increased with milkfat production. However, the change in the standard deviation was less than the change in mean production, such that the coefficient of variation decreased. This relationship was more clearly shown by the correlation coefficient between HYA means and standard deviations estimated as 0.44 (Table 5.2).

Table 5.1. Contemporary group numbers, averages of HYA means and within-HYA standard deviations and coefficient of variation for milkfat yield overall and at three levels of production.

	Production level			
Variable	All	Low	Medium	High
Number of contemporary groups	83,805	27,935	27,935	27,935
Average of HYA means	154.29	139.32	152.11	172.29
Average of within-HYA standard deviations	26.49	24.99	26.01	28.32
Coefficient of variation	17.17	17.94	17.10	16.44

Correction of the data by the herd-year-age mean, as indicated in Equation (4.6), reduced the average of within-HYA standard deviations, from 26.49 to 17.56 (Table 5.1). This correction notably reduced the correlation between the HYA means and within-HYA standard deviations with a change from a positive to negative value (0.44 to -0.27; Table 5.2).

Adjusting the data for herd-year-age standard deviation, as indicated in Equation (4.7), slightly reduced the averages of within-HYA standard deviations (from 26.49 to 25.82; Table 5.1) and, as a result, this adjustment reduced the correlation between HYA mean and within-HYA standard deviation from 0.44 to 0.31 (Table 5.2).

Effect on average within-HYA standard deviations of log transformation was similar to the effect of the mean correction. Log transformation reduced the average of standard deviations from 26.49 to 18.87 kg milkfat (Table 5.1) but also tended to overadjust the correlation between the mean and standard deviation changing it from 0.44 to -0.24 (Table 5.2).

Table 5.2. Averages and standard deviation of herd-year-age (HYA) means and within-HYA standard deviations and correlations between them for milkfat production using untransformed and transformed data.

	Transformation			
Variable	None	MEAN[1]	SD[2]	LOG[3]
Average of HYA means	154.29	100.00	157.11	147.07
Average of within-HYA standard deviations	26.49	17.56	25.82	18.87
Correlation between HYA mean and within-HYA standard deviation	0.44	-0.27	0.31	-0.24

[1] Milkfat yield scaled by the corresponding HYA mean.

[2] Milkfat yield adjusted by the corresponding estimate of HYA standard deviation.

[3] Natural logarithm of milkfat yield.

5.2 Averages and Standard Deviations of Estimates of Breeding Values

The averages and standard deviations of breeding values for milkfat yield of sires evaluated at three levels of production and with different minimum number of daughters are shown in Table 5.3. The number of sires with more than 10 daughters in each production level, was 1151. The average of estimated breeding values for these sires was -0.3±10.5 kg milkfat when evaluated at all levels. Small departures from this average was observed when the sires were evaluated at three different levels of

production. As the number of minimum number of daughters increased, the average breeding of sires also increased. Sires having more than 100 daughters in each production level averaged 3.0±8.7 kg milkfat for the whole data and when evaluated at the three levels of production were similar, 3.1 (low), 3.0 (medium) and 3.0 (high). The standard deviation of estimated breeding values increased as the level of production increased ±7.6, ±8.5 and ±9.8 kg milkfat, respectively.

The averages and standard deviations of breeding values for lactation milkfat yield corrected by the herd-year-age mean are showed in Table 5.4. The averages of estimated breeding values of sires using the whole data were similar to the averages at each of the production levels. Compared with the averages using the data without transformation, this method to account for heterogeneous variance reduced the variability of estimated breeding values. For instance, the standard deviation of estimated breeding value for sires with more than 100 daughters at each production level when evaluated without considering level of production was 8.7 (see Table 5.3) compared with 5.4 kg milkfat (see Table 5.4).

The averages and standard deviations of estimated breeding values for lactation milkfat yield adjusted by the estimate of the herd-year-age standard deviation are shown in Table 5.5. Similar trends to the data without correction were observed. The averages of estimated breeding values increased as the number of daughters increased. The averages at each production level were similar between them but higher standard deviations of breeding values estimated at the high level of production were observed.

Effect of logarithmic transformation on averages and standard deviation of estimated breeding values at different producing levels are shown in Table 5.6. Compared with the averages using the data without transformation, this method reduced the standard deviation of estimates of breeding values.

Table 5.3. Averages and standard deviations of breeding values for milkfat yield of sires evaluated at three levels of production and with different minimum number of daughters. Untransformed data.

Minimum number of daughters at each production level	Number of sires	Level of production			
		All	Low	Medium	High
10	1151	-0.3±10.5	-0.6± 9.6	-0.3± 9.9	0.1±10.4
20	486	0.5± 9.9	0.2± 9.3	0.4± 9.6	0.7±10.4
30	351	0.9± 9.6	0.8± 8.5	0.8± 9.4	1.2±10.3
40	306	1.2± 9.6	1.2± 8.4	1.2± 9.5	1.5±10.2
50	272	1.5± 9.3	1.6± 8.2	1.5± 9.1	1.7±10.2
60	248	2.2± 9.0	2.2± 8.0	2.1± 8.9	2.4±10.0
70	230	2.3± 9.1	2.4± 8.0	2.2± 9.0	2.3±10.0
80	216	2.6± 8.9	2.7± 7.8	2.5± 8.7	2.7± 9.9
90	209	2.9± 8.7	2.9± 7.7	2.9± 8.4	3.0± 9.7
100	198	3.0± 8.7	3.1± 7.6	3.0± 8.5	3.0± 9.8

Table 5.4. Averages and standard deviations of breeding values for milkfat yield of sires evaluated at three levels of production and with different minimum number of daughters. Data corrected by the HYA mean.

Minimum number of daughters at each production level	Number of sires	Level of production			
		All	Low	Medium	High
10	1151	-0.3±6.6	-0.5±6.4	-0.2±6.3	-0.0±6.0
20	486	0.2±6.2	0.1±6.1	0.2±6.1	0.4±6.0
30	351	0.5±6.0	0.5±5.8	0.5±5.9	0.7±5.9
40	306	0.7±5.9	0.8±5.7	0.7±6.0	0.8±6.0
50	272	0.9±5.8	1.2±5.6	0.9±5.8	1.0±6.0
60	248	1.4±5.6	1.5±5.3	1.3±5.6	1.4±5.9
70	230	1.4±5.6	1.7±5.3	1.4±5.6	1.3±5.8
80	216	1.7±5.5	1.9±5.2	1.6±5.5	1.6±5.8
90	209	1.9±5.4	2.0±5.1	1.8±5.4	1.8±5.6
100	198	1.9±5.4	2.1±5.1	1.9±5.4	1.8±5.7

Table 5.5. Averages and standard deviations of breeding values for milkfat yield of sires evaluated at three levels of production and with different minimum number of daughters. Data adjusted by the HYA sd.

Minimum number of daughters at each production level	Number of sires	Level of production			
		All	Low	Medium	High
10	1151	-0.5±9.5	-0.7±8.8	-0.5±9.1	-0.0±9.2
20	486	0.2±9.0	0.0±8.6	0.2±8.9	0.4±9.3
30	351	0.5±8.8	0.5±8.2	0.5±8.8	0.9±9.2
40	306	0.8±8.8	0.9±8.1	0.9±8.9	1.2±9.3
50	272	1.2±8.6	1.4±8.0	1.2±8.6	1.4±9.4
60	248	1.8±8.4	2.0±7.7	1.7±8.5	2.0±9.2
70	230	1.9±8.5	2.2±7.7	1.8±8.5	1.9±9.2
80	216	2.2±8.4	2.4±7.6	2.2±8.3	2.3±9.1
90	209	2.5±8.2	2.6±7.5	2.6±8.1	2.6±8.9
100	198	2.6±8.2	2.8±7.5	2.6±8.1	2.6±9.0

Table 5.6. Averages and standard deviations of breeding values for milkfat yield of sires evaluated at three levels of production and with different minimum number of daughters. Data after logarithmic transformation.

Minimum number of daughters at each production level	Number of sires	Level of production			
		All	Low	Medium	High
10	1151	-0.2±6.9	-0.5±6.8	-0.2±6.7	0.1±6.3
20	486	0.4±6.4	0.2±6.5	0.2±6.3	0.5±6.2
30	351	0.6±6.2	0.6±6.2	0.5±6.1	0.8±6.2
40	306	0.8±6.2	0.9±6.1	0.8±6.2	0.9±6.2
50	272	1.1±6.0	1.3±5.8	1.0±5.9	1.1±6.2
60	248	1.5±5.7	1.7±5.6	1.4±5.7	1.5±6.0
70	230	1.6±5.8	1.9±5.5	1.5±5.7	1.5±6.0
80	216	1.8±5.7	2.0±5.4	1.7±5.6	1.7±6.0
90	209	2.1±5.5	2.1±5.3	2.0±5.4	2.0±5.8
100	198	2.1±5.6	2.3±5.3	2.0±5.5	2.0±5.8

5.3 Correlation Between Estimated Breeding Values

5.3.1 Product-Moment Correlations

Expected and product-moment correlation (Pearson's coefficients) between breeding values of sires for milkfat production estimated at the three level of production and with different minimum number of daughters at each production level have been summarised in Tables 5.7, 5.8, 5.9 and 5.10 for raw data and for the three methods to reduce heterogeneity of variance (mean correction, standard deviation adjustment and log transformation, respectively).

Confidence intervals for a probability of 95% were constructed for observed correlations. If the expected correlation was outside of the confidence limits, then it was concluded that the observed correlation was significantly different to the expected correlation.

For raw data (Table 5.7) the product-moment correlations between breeding values for the comparison low-high producing levels were lower than the product-moment correlations between breeding values for the comparisons low-high and medium-high producing levels.

For all the comparisons, observed correlations were significantly different to the expected correlations except for the comparison low-medium production levels with sires having at least 10 daughters at each level of production. As shown in the table, the product-moment and expected correlations increased as the number of daughters in the estimation of breeding values increased. The lowest observed correlation (0.67) corresponds to the comparison low-high producing levels with sires having at least 10 daughters in each production level and the highest observed correlation (0.93) corresponds to the comparison medium-high producing levels with sires having at least 80 to 100 daughters in each producing level.

Product-moment correlations between breeding values for mean-corrected data estimated from each independent subset are shown in Table 5.8. All the observed correlations were significantly different to the expected correlations except for sires having at least 10 daughters in each level of production in the low-medium and medium-high comparisons. Observed correlations in the low-high producing level comparison

were lower than those of the low-medium and medium-high comparisons. The lowest observed correlation (0.67) corresponds to the comparison low-high producing levels with sires having at least 10 daughters in each production level and the highest observed correlation (0.93) corresponds to the comparison medium-high producing levels with sires having at least 100 daughters in each producing level. These lowest and highest observed correlations were the same to those observed for untransformed data.

Standard deviation adjustment slightly increased the product-moment correlations between estimates of breeding values obtained from the three independent subsets (compare Tables 5.7 and 5.9). Observed correlations were significantly different to the expected correlations except for sires with at least 10 daughters in each producing levels. The lowest observed correlation (0.69) corresponds to the comparison low-high producing levels and was not significantly different to the expected correlation (0.71). The highest observed correlation (0.94) corresponds to the comparison medium-high for sires having at least 100 daughters in each level of production. Between comparisons, lower correlations were observed in the low-high comparison.

Log transformation (Table 5.10) caused similar effects on product moment-correlations to those of the mean correction. All observed correlations were significantly different to expected correlations except for the comparison low-medium producing levels for sires having at least 10 daughters in each subset. Observed correlations increased as the minimum number of daughters in each production level increased.

Table 5.7. Expected (r_e) and product-moment (r_p) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with different minimum number of daughters. Untransformed data.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_p	r_e	r_p	r_e	r_p
10	1151	0.71	0.72	0.71	0.67*	0.73	0.72*
20	486	0.85	0.76*	0.85	0.73*	0.87	0.78*
30	351	0.90	0.78*	0.91	0.74*	0.91	0.82*
40	306	0.92	0.81*	0.92	0.77*	0.93	0.86*
50	272	0.94	0.83*	0.94	0.80*	0.94	0.88*
60	248	0.95	0.85*	0.95	0.82*	0.95	0.89*
70	230	0.95	0.86*	0.95	0.85*	0.96	0.92*
80	216	0.96	0.88*	0.96	0.87*	0.96	0.93*
90	209	0.96	0.90*	0.96	0.88*	0.97	0.93*
100	198	0.97	0.92*	0.97	0.91*	0.97	0.93*

* P<.05

Table 5.8. Expected (r_e) and product-moment (r_p) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with different minimum number of daughters. Data corrected by the HYA mean.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_p	r_e	r_p	r_e	r_p
10	1151	0.71	0.72	0.71	0.67*	0.73	0.71
20	486	0.85	0.78*	0.85	0.74*	0.87	0.78*
30	351	0.90	0.79*	0.91	0.75*	0.91	0.83*
40	306	0.92	0.82*	0.92	0.78*	0.93	0.86*
50	272	0.94	0.83*	0.94	0.80*	0.94	0.88*
60	248	0.95	0.84*	0.95	0.81*	0.95	0.89*
70	230	0.95	0.86*	0.95	0.84*	0.96	0.92*
80	216	0.96	0.87*	0.96	0.86*	0.96	0.92*
90	209	0.96	0.88*	0.96	0.88*	0.97	0.92*
100	198	0.97	0.91*	0.97	0.90*	0.97	0.93*

* P<.05

Table 5.9. Expected (r_e) and product-moment (r_p) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with different minimum number of daughters. Data adjusted by the HYA sd.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_p	r_e	r_p	r_e	r_p
10	1151	0.71	0.73	0.71	0.69	0.73	0.73
20	486	0.85	0.78*	0.85	0.75*	0.87	0.80*
30	351	0.90	0.80*	0.91	0.77*	0.91	0.84*
40	306	0.92	0.82*	0.92	0.79*	0.93	0.87*
50	272	0.94	0.84*	0.94	0.82*	0.94	0.89*
60	248	0.95	0.85*	0.95	0.84*	0.95	0.90*
70	230	0.95	0.86*	0.95	0.86*	0.96	0.92*
80	216	0.96	0.88*	0.96	0.88*	0.96	0.93*
90	209	0.96	0.90*	0.96	0.90*	0.97	0.93*
100	198	0.97	0.92*	0.97	0.92*	0.97	0.94*

* P<.05

Table 5.10. Expected (r_e) and product-moment (r_p) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with different minimum number of daughters. Data after log transformation.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_p	r_e	r_p	r_e	r_p
10	1151	0.71	0.68	0.71	0.62*	0.73	0.70*
20	486	0.85	0.73*	0.85	0.69*	0.87	0.75*
30	351	0.90	0.74*	0.91	0.70*	0.91	0.79*
40	306	0.92	0.77*	0.92	0.72*	0.93	0.84*
50	272	0.94	0.79*	0.94	0.76*	0.94	0.87*
60	248	0.95	0.81*	0.95	0.78*	0.95	0.87*
70	230	0.95	0.83*	0.95	0.81*	0.96	0.90*
80	216	0.96	0.84*	0.96	0.83*	0.96	0.90*
90	209	0.96	0.87*	0.96	0.85*	0.97	0.91*
100	198	0.97	0.89*	0.97	0.89*	0.97	0.92*

* P<.05

5.3.2 Rank Correlations

Spearman's correlation coefficient was used as measure of rank correlation between estimates of breeding values. Significant departures of observed from expected rank correlations means that sires ranked on breeding values in one environment are not ranked in same order in another environment. Rank correlations of estimates of breeding values for lactation milkfat yield of sires used in three producing levels are shown in Tables 5.11 (raw data), 5.12 (data corrected by the HYA mean), 5.13 (data adjusted by the HYA standard deviation), and 5.14 (data after log transformation).

Expected correlations were the same to those derived for the product-moment correlations and were a function of the number of daughter in each production level and assumed parameters of heritability. Confidence intervals for a probability of 95% were constructed for observed rank correlation coefficients transformed to the Fisher's Z. Expected correlation outside of the confidence lower and upper limits meant that the observed correlation was significantly different to the expected correlation.

Rank correlations were slightly lower than product-moment correlations, but both product-moment and rank correlations exhibited the same trend for raw data and for the data transformed by the three methods of reducing heterogeneity of variance.

For raw, mean-corrected and log-transformed data all rank correlations were significantly different to the expected correlation except for sires having at least 10 daughters in each level of production in the comparison low-medium producing levels. For standard deviation-adjusted data all observed rank correlations were significantly different to the expected correlations except for sires having at least 10 of daughters at each level of production in the three comparisons. Rank correlations were lower in the comparison low-high producing levels than those of the other comparisons (low-medium and medium-high).

Table 5.11. Expected (r_e) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated estimated at different levels of production and with different minimum number of daughters. Untransformed data.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_s	r_e	r_s	r_e	r_s
10	1151	0.71	0.70	0.71	0.65*	0.73	0.69*
20	486	0.85	0.76*	0.85	0.72*	0.87	0.77*
30	351	0.90	0.77*	0.91	0.74*	0.91	0.82*
40	306	0.92	0.80*	0.92	0.77*	0.93	0.86*
50	272	0.94	0.83*	0.94	0.80*	0.94	0.87*
60	248	0.95	0.83*	0.95	0.80*	0.95	0.88*
70	230	0.95	0.85*	0.95	0.84*	0.96	0.91*
80	216	0.96	0.87*	0.96	0.85*	0.96	0.91*
90	209	0.96	0.90*	0.96	0.87*	0.97	0.91*
100	198	0.97	0.91*	0.97	0.90*	0.97	0.92*

* P<.05

Table 5.12. Expected (r_e) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated estimated at different levels of production and with different minimum number of daughters. Data corrected by the HYA mean.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_s	r_e	r_s	r_e	r_s
10	1151	0.71	0.70	0.71	0.65*	0.73	0.68*
20	486	0.85	0.76*	0.85	0.73*	0.87	0.77*
30	351	0.90	0.78*	0.91	0.76*	0.91	0.82*
40	306	0.92	0.81*	0.92	0.77*	0.93	0.86*
50	272	0.94	0.82*	0.94	0.81*	0.94	0.87*
60	248	0.95	0.83*	0.95	0.81*	0.95	0.88*
70	230	0.95	0.84*	0.95	0.83*	0.96	0.90*
80	216	0.96	0.87*	0.96	0.85*	0.96	0.90*
90	209	0.96	0.89*	0.96	0.87*	0.97	0.91*
100	198	0.97	0.90*	0.97	0.89*	0.97	0.91*

* P<.05

Table 5.13. Expected (r_e) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated estimated at different levels of production and with different minimum number of daughters. Data adjusted by the HYA sd.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_s	r_e	r_s	r_e	r_s
10	1151	0.71	0.71	0.71	0.70	0.73	0.71
20	486	0.85	0.76*	0.85	0.74*	0.87	0.78*
30	351	0.90	0.78*	0.91	0.77*	0.91	0.82*
40	306	0.92	0.81*	0.92	0.78*	0.93	0.86*
50	272	0.94	0.84*	0.94	0.82*	0.94	0.88*
60	248	0.95	0.84*	0.95	0.82*	0.95	0.88*
70	230	0.95	0.86*	0.95	0.85*	0.96	0.91*
80	216	0.96	0.88*	0.96	0.87*	0.96	0.91*
90	209	0.96	0.90*	0.96	0.89*	0.97	0.92*
100	198	0.97	0.91*	0.97	0.91*	0.97	0.92*

* P<.05

Table 5.14. Expected (r_e) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated estimated at different levels of production and with different minimum number of daughters. Data after log transformation.

Minimum number of daughters at each production level	Number of sires	Level of production compared					
		Low-Medium		Low-High		Medium-High	
		r_e	r_s	r_e	r_s	r_e	r_s
10	1151	0.71	0.66	0.71	0.60*	0.73	0.66*
20	486	0.85	0.73*	0.85	0.69*	0.87	0.73*
30	351	0.90	0.74*	0.91	0.72*	0.91	0.78*
40	306	0.92	0.77*	0.92	0.73*	0.93	0.83*
50	272	0.94	0.80*	0.94	0.78*	0.94	0.86*
60	248	0.95	0.80*	0.95	0.78*	0.95	0.86*
70	230	0.95	0.82*	0.95	0.82*	0.96	0.89*
80	216	0.96	0.84*	0.96	0.83*	0.96	0.88*
90	209	0.96	0.87*	0.96	0.85*	0.97	0.89*
100	198	0.97	0.88*	0.97	0.88*	0.97	0.90*

* P<.05

5.3.3 Product-moment *versus* Rank Correlations

Comparisons between product-moment and rank correlations of breeding values for lactation milkfat yield of sires evaluated at different levels of production and with variable number of daughters are summarised in Appendix I. Table I.1 contains the comparison between correlation coefficients for untransformed data, Table I.2 for HYA mean-corrected data, Table I.3 for HYA standard deviation-corrected data and Table I.4 for log transformed data.

Using the Fisher's Z statistic, confidence intervals for a probability of 95% were constructed for both product-moment and rank correlations. It was found no significant differences between them but rank correlations were lower than product-moment correlations.

5.4 Genetic Correlations

The ratio between the observed and the expected correlation was considered as an estimator of the genetic correlation of production in two environments (Mason and Robertson, 1956; Danell, 1982; Notter and Diaz, 1993). For raw data (Table 5.15), the estimates of the genetic correlation of the same genotype expressed in the low and medium producing levels fluctuated between 0.87 and 1.0. When the low and high levels were compared, the estimates of genetic correlations were reduced varying between 0.82 and 0.95. In the comparison medium-high producing levels, estimates of genetic correlations varied between 0.90 and 0.99.

When data were corrected by the herd-year-age mean (Table 5.16), the pattern of estimates of genetic correlations was the same to that of raw data. The lowest estimates were obtained in the comparison low-high producing levels varying between 0.83 and 0.95.

Adjustment by the HYA standard deviation slightly increased the estimates of genetic correlations (Table 5.17). In the comparison low-medium producing levels, estimates varied from 0.88 to 1.02, in the low-high comparison estimates varied from 0.88 to 0.97 and in the medium-high comparison estimates were the highest varying from 0.92 to 1.00.

Log transformation reduced the estimates of genetic correlations between the expression of the same genotype in two environment (Table 5.18). The lowest estimates correspond to the comparison low-high producing levels varying between 0.77 and 0.92. In the other two comparisons, estimates varied from 0.82 to 0.95 and were lower than those obtained by using raw data.

5.5 Significantly Deviating Breeding Values

Breeding values for lactation milkfat yield of sires evaluated at different producing levels and with different minimum number of daughters were compared. Differences outside a 95% confidence interval for the difference between two sire proofs were counted. Comparisons between observed and expected number of sires with significantly deviating breeding values are shown in Tables 5.19 (raw data), 5.20 (data corrected by the herd-year-age mean), 5.21 (data adjusted by the HYA standard deviations) and 5.22 (data after logarithmic transformation).

For raw data (Table 5.19) observed numbers of sires with significantly deviating breeding values were greater than expected numbers, except for sires having at least 10 daughters in each producing level in the comparison low-medium-producing level. The greatest number sires with significant deviation of breeding values was observed in the comparison low-high comparison.

Correction by the herd-year-age mean slightly reduced the number of sires with significantly deviating breeding values (Table 5.20) but still observed numbers were greater than those expected, except for sires having at least 10 and 20 daughters in each level of production in the comparison low-medium producing levels and for sires having at lest 10 daughters in the comparison medium-high producing levels.

Table 5.15. Estimates of correlations between true breeding values expressed in different production levels and with variable minimum number of daughters. Untransformed data.

Minimum number of daughters at each production level	Number of sires	Level of production compared		
		Low-Medium	Low-High	Medium-High
10	1151	1.01	.95	.99
20	486	.89	.86	.90
30	351	.87	.82	.90
40	306	.88	.84	.93
50	272	.89	.85	.93
60	248	.89	.87	.93
70	230	.90	.89	.96
80	216	.92	.90	.96
90	209	.94	.92	.96
100	198	.95	.95	.96

Table 5.16. Estimates of correlations between true breeding values expressed in different production levels and with variable minimum number of daughters. Data corrected by the HYA mean.

Minimum number of daughters at each production level	Number of sires	Level of production compared		
		Low-Medium	Low-High	Medium-High
10	1151	1.01	.95	.98
20	486	.92	.87	.90
30	351	.88	.83	.90
40	306	.88	.84	.93
50	272	.89	.86	.94
60	248	.89	.86	.93
70	230	.90	.88	.96
80	216	.90	.89	.95
90	209	.92	.91	.96
100	198	.94	.93	.96

Table 5.17. Estimates of correlations between true breeding values expressed in different production levels and with variable minimum number of daughters. Data adjusted by the HYA sd.

Minimum number of daughters at each production level	Number of sires	Level of production compared		
		Low-Medium	Low-High	Medium-High
10	1151	1.02	.97	1.00
20	486	.91	.88	.92
30	351	.88	.85	.92
40	306	.89	.86	.94
50	272	.90	.87	.94
60	248	.90	.88	.94
70	230	.91	.90	.96
80	216	.92	.92	.96
90	209	.94	.93	.96
100	198	.95	.95	.97

Table 5.18. Estimates of correlations between true breeding values expressed in different production levels and with variable minimum number of daughters. Data after log transformation.

Minimum number of daughters at each production level	Number of sires	Level of production compared		
		Low-Medium	Low-High	Medium-High
10	1151	.95	.87	.92
20	486	.86	.81	.86
30	351	.82	.77	.86
40	306	.83	.77	.91
50	272	.85	.81	.92
60	248	.86	.82	.91
70	230	.87	.85	.94
80	216	.88	.86	.94
90	209	.90	.89	.94
100	198	.92	.92	.95

Standard deviation adjustment reduced the number of sires with significantly deviating breeding values (Table 5.21). The effect was similar to that of the mean correction in such a way that still observed values were significantly greater than those expected except for sires having at least 10 and 20 daughters in each level of production in the comparison low-medium producing levels and for sires having at lest 10 daughters in the comparison medium-high producing levels. The number of sires with significant deviations were higher in the low-high comparison than in the low-medium and medium-high comparisons.

Log transformation also reduced the number of sires with significantly deviating breeding values (Table 5.22) but at less extension than mean correction and standard deviation adjustment. Observed values were greater than expected values except for sires having at least 10 daughters in each level of production in the low-medium and medium-high comparison where the observed values were equal to the expected values.

Averages and standard deviations of breeding values of sires which were found with significantly deviating breeding values when evaluated at three levels of production are given in Appendix II. Table II.1 shows statistics for raw data and Tables II.2, II.3, and II.3 hold statistics for the three methods of reducing heterogeneity of variance.

Table 5.19. Expected[1] and observed number of sires with significantly deviating breeding values for milkfat yield estimated at three levels of production and with variable minimum number of daughters. Untransformed data.

Minimum number of daughters at each production level	Number of sires	Expected number of sires	Level of production compared		
			Low-Medium	Low-High	Medium-High
10	1151	58	50	90*	77*
20	486	24	33*	61*	51*
30	351	18	29*	56*	43*
40	306	15	27*	51*	38*
50	272	14	22*	49*	36*
60	248	12	19*	46*	35*
70	230	12	19*	44*	32*
80	216	11	17*	42*	31*
90	209	10	15*	41*	30*
100	198	10	14*	39*	30*

[1] 5% of the sires are expected to have significantly deviating breeding values.

* significantly different (P<.05)

Table 5.20. Expected[1] and observed number of sires with significantly deviating breeding values for milkfat yield estimated at three levels of production and with variable minimum number of daughters. Data corrected by the HYA mean.

Minimum number of daughters at each production level	Number of sires	Expected number of sires	Level of production compared		
			Low-Medium	Low-High	Medium-High
10	1151	58	39	60*	51
20	486	24	22	39*	33*
30	351	18	21*	36*	26*
40	306	15	19*	31*	22*
50	272	14	18*	29*	20*
60	248	12	16*	27*	20*
70	230	12	16*	25*	17*
80	216	11	15*	23*	17*
90	209	10	13*	22*	16*
100	198	10	12*	21*	16*

[1] 5% of the sires are expected to have significantly deviating breeding values.

* significantly different (P<.05)

Table 5.21. Expected[1] and observed number of sires with significantly deviating breeding values for milkfat yield estimated at three levels of production and with variable minimum number of daughters. Data adjusted by the HYA sd.

Minimum number of daughters at each production level	Number of sires	Expected number of sires	Level of production compared		
			Low-Medium	Low-High	Medium-High
10	1151	58	32	58	49
20	486	24	24	42*	38*
30	351	18	21*	40*	33*
40	306	15	18*	36*	30*
50	272	14	17*	35*	28*
60	248	12	17*	32*	28*
70	230	12	17*	30*	25*
80	216	11	15*	28*	25*
90	209	10	13*	27*	24*
100	198	10	12*	26*	24*

[1] 5% of the sires are expected to have significantly deviating breeding values.

* significantly different (P<.05)

Table 5.22. Expected[1] and observed number of sires with significantly deviating breeding values for milkfat yield estimated at three levels of production and with variable minimum number of daughters. Data after log transformation.

Minimum number of daughters at each production level	Number of sires	Expected number of sires	Level of production compared		
			Low-Medium	Low-High	Medium-High
10	1151	58	58	73*	58
20	486	24	29*	45*	33*
30	351	18	26*	42*	27*
40	306	15	23*	37*	21*
50	272	14	22*	32*	19*
60	248	12	18*	28*	18*
70	230	12	17*	26*	15*
80	216	11	16*	24*	15*
90	209	10	14*	23*	14*
100	198	10	13*	22*	14*

[1] 5% of the sires are expected to have significantly deviating breeding values.

* significantly different (P<.05)

CHAPTER 6
DISCUSSION

6.1 Correlation Between Mean and Standard Deviation

Existing literature estimates of heterogeneity of variance are often contradictory. While some studies have found a positive relationship between herd mean and herd (phenotypic) variance for milk yield (Hill et al. 1983; Mirande and Van Vleck, 1985; Brotherstone and Hill, 1986; Dong and Mao, 1990), others have found no evidence of such relation (Lofgren et al. 1985; Winkelman and Schaeffer, 1988). Even for the studies that did find a (positive) correlation, the relationship was not strong. A typical value would be 0.4-0.5.

Visscher et al. (1991) found a correlation between herd means and estimated herd phenotypic standard deviation for lactation milkfat yield of 0.71. This value was higher than the value of 0.44 for the correlation between the herd-year-age mean and herd-year-age phenotypic standard deviation for milkfat yield found in this study. This study does confirm the presence of heterogeneity of variance for milkfat yield in New Zealand dairy cattle, i.e., herds with high levels of production tends to exhibit more variability than those with low levels of production. However, heterogeneity of variance can only be partially explained by the scale effect because the correlation was 0.44 rather than unity.

6.2 Correlation Between Estimated Breeding Values from Independent Datasets

An important question is if breeding values of dairy sires will rank in the same order over a wide range of environmental conditions. The solution to such a question is given by the correlation between breeding values estimated at different environments. If the observed correlations approximate to the expected correlations, sires can be selected in one environment and used in another environment. On the contrary, if the observed correlations are significantly different to the expected correlations, daughters must be milked in that environment where their sires' genetic evaluation was made.

Several studies carried out on within-country basis have calculated high values of correlations between estimated breeding values at different producing levels (Manson and Robertson, 1956; Robertson et al. 1960; Lytton and Legates, 1966; McDaniel and

Corley, 1967; Burnside and Rennie, 1968; Danell, 1982; Winkelman and Schaeffer, 1988; Dong and Mao, 1990) indicating that progeny groups tend to rank similarly under different environments.

Estimated correlations between breeding values for sires in different countries have been contradictory. Some studies have found that observed correlations have approximated to the expected correlations (Petersen, 1975; Powell and Dickinson, 1977; Powell and Wiggans, 1991). Thus, the rank of sires on estimates of breeding values in one country tends to stay the same when the same sires are evaluated in another country. Other studies, in contrast, have estimated low correlations between breeding values estimated in different countries (Abubakar et al. 1987; Peterson, 1988)

In the present study three independent estimates of breeding values were calculated, each based upon the performance of daughters assessed in each production level. The three sets of estimates of breeding values provided, in effect, three independent measures of the breeding worth of each sire. The product-moment and rank correlations between breeding values for milkfat yield of sires evaluated at different production levels were high but still significantly different to the expected correlations (see Tables 5.7 and 5.11). Thus, sires tended to rank differently across levels of production.

There are, however, some implications that restrict the validity of the comparisons between expected and observed correlations. First, expected correlations were derived as a function of the true number of daughters in each producing level. Ideally, expected correlations would be obtained using accuracies of estimated breeding values computed from elements of the inverse matrix of the mixed model equations by using iterative procedures (VanRaden and Freeman, 1985; Greenhalgh et al. 1986; Robinson and Jones, 1987; Misztal and Wiggans, 1988; Meyer, 1989; VanRaden and Wiggans, 1991). Accuracy of estimated breeding values from the elements of the inverse matrix takes into account all available information of the animal being evaluated and therefore accuracy would be increased.

Second, expected correlations were derived under the assumptions given by Taylor (1983) (as cited by Notter and Diaz, 1993) which are: no environmental correlation between performance in the different environments, no relationships among parents of measured animals, no other covariances among predicted breeding values within either environment, and sires are chosen at random. This may be not the case

since genetic evaluations of sires in this study were made by using an animal model which considered relationships between animals.

And last, values of heritability in calculating expected correlations were assumed constant across production levels. This may be true under two conditions (Henderson, 1984; Garrick and Van Vleck, 1987; Visscher and Hill, 1992): (i) equal additive genetic and residual variances in all environments (homogeneity of variance) or (ii) when the additive genetic and residual variances are changing proportionally such that heritability remains constant across environments (heterogeneity of variance). In this study, contemporary groups were defined as herd-year-age. Obtaining estimates of genetic and residual variance for each of them would be inaccurate because of the small number of animals in each group.

Differences between product-moment and rank correlations may indicate deviations from normality in the distribution of breeding values (Danell, 1982). Both rank and product-moment correlations were calculated between breeding values for milkfat yield of sires evaluated at different producing levels and with different minimum number of daughters. The results are presented in Table I.1 of Appendix I. The figures for both type of correlation were very similar and there were no significant differences. There was a trend for rank correlations to be slightly lower than the product-moment correlations. Thus, the ranking of the sires seemed to change more than the averages and the variation at the different levels of production. Similar results were reported by Danell (1982) for Swedish Red and White dairy cattle.

6.3 Genetic Correlations

A matter for consideration is whether dairy sires should be selected on the basis of daughter records made in the same environment where the sires are going to be used (Lush, 1945) or whether a particular type of environment may be more favourable for discriminating between sires to be used in another environment (Hammond, 1947). A solution to this question was given by Falconer (1952) who introduced the concept of a genetic correlation between performance in different environments and used the ratio of indirect to direct response to selection to determine the optimum environment for selection.

The formula given by Falconer (1952) refers to mass selection. In that case the

response in environment 1 when selecting in environment 1 is

$$\Delta G_1 = i \cdot h_1 \cdot \sigma_{a_1}$$

and when selection is based on environment 2, the magnitude of the correlated response is

$$\Delta_c G_1 = i \cdot h_2 \cdot r_G \cdot \sigma_{a_1}.$$

The ratio of the correlated to the direct response assuming no change in selection intensity or generation interval becomes

$$\frac{\text{correlated reponse}}{\text{direct response}} = \frac{\Delta_c G_1}{\Delta G_1} = \frac{h_2}{h_1} \cdot r_G. \qquad (6.1)$$

Thus it is possible to evaluate the correlated response relative to the direct response simply in terms of the two heritabilities and the genetic correlation. The important point to note in Equation (6.1) is that when the genetic correlation of the same trait expressed in two environments is the unity and the heritability across environments is the same, the correlated response will be equal to the direct response. In the case of genetic correlation less than the unity but still equal heritabilities, the proportion of correlated response to the direct response will be less than one.

The correlation between the true breeding values for milk yield (Robertson et al. 1960; Lytton and Legates, 1966; Danell, 1982; Hill et al. 1983; De Veer and Van Vleck, 1987; Menendez-Buxadera, et al. 1989; Boldman and Freeman, 1990; Carabaño et al. 1990) and milkfat yield (Lytton and Legates, 1966; Carabaño et al. 1990) of bulls evaluated at different levels of production has been reported very close to one, which seems to indicate little evidence of any important genotype by production level interactions for production traits, because if the genetic correlation is high, then performance in two different production levels represents very nearly the same characters, determined by very nearly the same genes.

In this analysis, the estimates of breeding values were obtained from a BLUP procedure using animal model. An estimate of the genetic correlation was obtained from the ratio between the observed and the expected correlation of estimated breeding values at the three levels of production as suggested by Mason and Robertson (1956), Calo et

al. (1973), Danell (1982) and Notter and Diaz (1993). Approximate genetic correlations between milkfat yield at the three production levels fluctuated from 0.87 to 1.01, from 0.85 to 0.95, and from 0.90 to 0.96, for the comparisons low-medium, low-high and medium-high, respectively (see Table 5.15). Estimates of genetic correlations were able to be greater than one because of the method used to generate approximate genetic correlations can yield estimates outside the parameter space.

Lack of knowledge concerning the distribution of the obtained estimates of genetic correlations allows neither accurate calculation of standard errors nor an objective statistical model to test whether or not the genetic correlations are significantly different from one. However, the standard errors associated with the estimates of genetic correlations would be small, given that the average number of daughter per sire were relatively large. This would suggest that in fact the correlation between the true breeding values is slightly lower than the unity indicating that if sires are genetically evaluated on the basis of daughters located in only one level of production, responses to selection may be lower than the expected.

6.4 Significantly Deviating Breeding Values

Averages of breeding values for milkfat yield of sires estimated at different levels of production were similar and they increased as the number of daughters increased (Table 5.3). The increase in the average as the number of daughters increased was a result of allowing the best sires to produce more daughters because they were selected as bulls to breed cows. However, this is not the point where attention must be focussed but in the number of sires with significantly deviating estimated breeding values (Table 5.19).

Greater number of sires with significantly deviating breeding values were observed in the comparison low-high producing levels (Table 5.19). This indicates that the estimates of breeding values of sires obtained from extreme production levels are affected in larger magnitude than when the sires are evaluated in low-medium or medium-high. This corresponds to the lower product-moment and rank correlations and genetic correlations observed in this comparison.

If a sire has large enough information and there is homogeneity of variance the estimated breeding values must be the same. For sires with more than 100 daughters in each production level the expected number of sires with significantly deviating breeding

values was 10. In the comparison low-high there were 39 sires observed, which is greater than the 5% expected. This indicates a clear effect of the heterogeneity of variance found in the data. Danell (1982) found in Swedish Red and White dairy cattle that, sires evaluated in different levels of production, the observed number of sires with significantly deviating were similar to the expected value.

6.5 Effect of Methods to Reduce Heterogeneity of Variance

6.5.1 Mean Correction

Scaling by the mean is used to reduce heterogeneity of variance when the relationship between the mean and standard deviation is directly proportional, thus, this method is equivalent to the logarithmic transformation. The present system of dairy sire genetic evaluation in New Zealand uses this method by expressing the contemporary comparison of a sire as the proportion of the performance of his daughters to the performance of the contemporaries.

The change from a positive to a negative correlation between mean and standard deviation caused by the herd-year-age mean correction indicates that the heterogeneity of variance of these data was overadjusted (Table 5.2). These results indicate that if this correction is applied in a BLUP analysis, assuming a constant variance among herds, the breeding values of superior cows from high yielding herds would be underpredicted relative to the breeding values of superior cows from low yielding herds. This would be very important when selection is on cows to breed bulls because more cows would be selected from the herds of low level of production.

The effect of this correction on sire evaluation can be assessed by the reduction of the correlation of estimated breeding values in different environments. The ratios between calculated and expected correlations were reduced, although, not significantly when compared to the untransformed data.

6.5.2 Standard Deviation Adjustment

Visscher et al. (1991) corrected first lactation milkfat yields for a ratio between the estimated population standard deviation and the estimated herd-year-season phenotypic standard deviation. They suggested the adjustment for herd-year-season standard deviation rather than for herd standard deviation because it is known that within-herd variances are often heterogeneous across years (Brotherstone and Hill, 1986), and because contemporary groups of herd-year-season rather than herds are usually fitted as fixed effects in the breeding value prediction. They found that this adjustment reduced the heterogeneity of variance substantially.

In this study, a similar adjustment was made considering a contemporary group as cows of the same age calving in the same herd and same year. But instead of using the within-herd-year-age standard deviation, an estimate for each herd-year-age standard deviation was used. This adjustment reduced the correlation between mean and standard deviations from 0.44 to 0.31 (see Table 5.2), i.e., the variances tended to be more stabilised than in the raw data.

The effects of this adjustment on the genetic evaluation of the sires with the animal model using BLUP is reflected in the correlation between breeding values estimated at different levels of production. Product-moment and rank correlations between estimated breeding values were slightly increased with respect to those obtained using raw data but observed correlations were still significantly lower than expected correlations (see Tables 5.9 and 5.13).

On the other hand, this method of correcting the scale effect increased the estimated genetic correlation between the expression of the same genotype in two environments. In the comparison between estimated breeding values at low and high levels the genetic correlation fluctuated between 0.87 to 1.01 for raw data (Table 5.15) and between 0.88 to 1.02 after this adjustment (Table 5.17)

6.5.3 Log Transformation

When the standard deviation is a simple linear function of the mean, that is, the correlation between them is a unity, logarithmic transformation is the appropriate transformation to correct for heterogenous variance (Everett and Keown, 1984; Garrick

and Van Vleck, 1987; Visscher and Hill, 1992). The logarithmic transformation is justified if the heterogeneity is just a scale effect (Visscher et al. 1991) resulting in the standard deviation being linearly related to the mean. If the mean-variance correlation has no genetic component, a logarithmic transformation will have the additional advantage of increasing the heritability. If the relationship is partly genetic, the heritability may be different on a logarithmic scale, depending on what proportion of the mean-variance correlation is genetically determined. Hill et al. (1983) used decimal logarithm transformation of milk, milkfat and protein yields and found that the phenotypic variances tended to be stabilised. That is, the correlation between the mean and phenotypic standard deviation was reduced.

The relationship between the mean and standard deviation is, however, often far from the ideal (correlation equal to zero) and sometimes, logarithm transformation can overadjust the data for heterogeneity of variance making that the correlation between the mean and phenotypic standard deviation changes from a positive value to a negative value. In the present study, logarithmic transformation changed the correlation between the herd-year-age means and the herd-year-age standard deviations from a positive value of 0.44 to a value of -0.24 (see Table 5.2). Similar effects of logarithmic transformation were found by Everett and Keown (1984) Mirande and Van Vleck (1985) and Visscher and Hill (1992). Superior cows in low producing herds would therefore be over-evaluated on the log scale.

The estimated genetic correlation of production in two environments using log-transformed data varied from 0.77 to 0.95 (Table 5.18) which was lower than that of untransformed data (Table 5.15). These results were similar to those reported by De Veer and Van Vleck (1987) who found that estimates of genetic correlations were large in all cases (>0.85) but were slightly smaller for logarithms than for their untransformed counterparts.

This latter result is consistent with the change of the sign in the correlation between mean and standard deviation caused by the logarithmic transformation (Table 5.2). The reduction in the estimated genetic correlation after logarithmic transformation indicates that the rank of the sires used across different environments would change as a consequence of overadjusting for this transformation. This was confirmed with the reduction in the product-moment and rank correlations of estimated breeding values after this transformation (Table 5.10 and Table 5.14).

Besides reranking bulls, log transformation produced significantly greater differences among estimated breeding values (Table 5.22) than occurred with untransformed data (Table 5.19).

6.6 Practical implications

If variances increase with mean yield but are assumed homogeneous, some animals will be misranked. Superior cows in herds with large variances will tend to be overevaluated (Vinson, 1987; Boldman and Freeman, 1990). An apparent excess of elite cows has been reported for herds with high means (Powell et al. 1983) and high variances (Everett et al. 1982) from analyses that assumed equal genetic and residual variances.

The effect on response to selection in choosing a greater proportion of cows from the more variable herds depends on the extend to which the greater variability is due to genetic as opposed to environmental factors. This study, however, does not clarify if differences between contemporary groups in the average milkfat production per cow are genetic or environmental in origin.

Similarly, the consequences of the scaling problem on sire selection was illustrated by Van Vleck (1987) who concluded that it has large effect on response to selection if there is an inappropriate allocation of the sire's daughters under the progeny test. Sires with a large percentage of their daughters in herds with large variances would tend to be overevaluated. However, simulation studies (Garrick and Van Vleck, 1987; Meuwissen and Van der Werf, 1993) support the contention that the genetic evaluation of sires through BLUP procedures are robust to the violations of assumptions regarding homogeneity of variance only to the extent that sires have daughters spread in several environments.

Results from field data (Winkelman and Schaeffer, 1988) have indicated that the modification of the mixed model equations to account for heterogeneous variances has little effect on the overall ranking of sires. The most important evaluations, however, are for sires and cows ranking in the top percentage of the population. These top sires are selected for extensive use throughout the cow population, and elite cows are candidates for producing young bulls for progeny testing. If the subsequent progeny test is random across production levels, sires out of overevaluated cows should be identified as inferior

and culled (Boldman and Freeman, 1990). Hence, compared with sire evaluation, cow evaluation is more likely to be adversely affected by heterogeneous variances (Van Vleck, 1987; Vinson, 1987).

Because the results give an indication of the scaling problem, it must be considered in the genetic evaluation of sires and dams in the national herd. The question therefore is what strategy should be used to deal with the problem of heterogeneity of variance between environments. Quaas et al. (1989) indicated that when there are heterogeneous residual and (or) genetic variances but the genetic correlation between genetic effects expressed in different classes is unity, a simple scaling procedure can be used to stabilise the heterogeneity of variance. Attempts to correct this problem were made in this study but the results were still not satisfactory. The mean correction as used in the present system of genetic evaluation tends to overadjust rather than equalise the standard deviations at different levels of production. A similar problem is found by applying log transformation. This overadjustment was reflected in reducing, instead of increasing, the correlations between breeding values estimated at the three levels of production. The standard deviation adjustment reduced the relationship between the standard deviation and mean but not completely (from 0.44 to 0.31; see Table 5.2) slightly improving the correlations between estimated breeding values.

These results suggest that heterogeneity of variances is not simply the result of a relationship between mean and variance. For a normally distributed variable y with mean μ_y and variance σ^2_y, the log of y has approximate variance σ^2_y/μ^2_y (Van Vleck, 1988). Therefore logarithmic transformation and correction by the HYA mean will stabilise variances only if the standard deviation on the original scale varies directly with the mean (Bartlett, 1947); i.e., if coefficient of variation are equal. For these data, coefficients of variation of untransformed milkfat yield for low, medium and high production levels were 17.94, 17.10, and 16.44 (Table 5.1).

Under the assumption that the standard deviation adjustment is best, the use of linear yields is recommended over the use of log transformation if homogeneous variances are assumed. These results argue for continued investigation of what is an appropriate transformation. This is in agreement with Falconer (1952), who cautioned that scaling procedures should be chosen carefully and only when there is enough justification.

6.7 Alternative Methods

This investigation indicates that there is heterogeneity of variance for milkfat yield in New Zealand dairy cattle population and that the methods used in this work to stabilise the phenotypic variance are unsatisfactory. A multiple-trait approach, considering the breeding value in each environment as independent trait, seems theoretically best (Schaeffer et al. 1978; Henderson, 1984; Gianola, 1986; Garrick and Van Vleck, 1987). However, this requires that in each contemporary group, the residual and genetic variance must be known. This latter requirement represents a difficult task due to the contemporary group having few records resulting in sampling problems.

Given the literature findings and the results from the present study, it seems most practical to pre-adjust data for some estimate of the herd or herd-year-age phenotypic standard deviation as indicated by Visscher and Hill (1992). Posterior estimates of herd parameters (heritability, phenotypic, genetic and residual variances, for example) can be obtained by Equation (3.20), which is $\tilde{\theta}_i = \hat{\theta}_0 + \beta_i(\hat{\theta}_i - \hat{\theta}_0)$ where $\hat{\theta}_i$ is a parameter estimate and $\tilde{\theta}_i$ its regressed estimate for herd i, and $\hat{\theta}_0$ is the overall (prior) estimate. The regression coefficient is $\beta_i = 1/1+\gamma_i$ with $\gamma_i = \mathrm{Var}(\hat{\theta}_i|\theta_i)/\mathrm{Var}(\theta_i)$ being the ratio of the sampling variance to the variance of the parameter.

Garrick (pers. comm.) suggests a method that encompasses the desirable properties of the Bayesian approach (Hill, 1984; Brotherstone and Hill, 1986; Visscher and Hill, 1992; Gianola et al. 1992; Weigel and Gianola, 1993) and accounts for any relationship between the mean and standard deviation. First it is necessary define the following notation already used in section (3.8) Equation (3.19):

σ_y is the parameter value for the phenotypic standard deviation in the population. In practice, it will be estimated from the data.

σ_{y_i} is the parameter value for the phenotypic standard deviation in herd i.

$\mathrm{var}(\sigma_{y_i})$ is the parameter that defines the extent of heterogeneous variance. This parameter is zero if variance is homogeneous across herds.

$\hat{\sigma}_{y_i}$ is a statistic, a strictly within-herd estimate of standard deviation in herd i.

$\tilde{\sigma}_{y_i}$ is a statistic, a posterior or pooled estimate of the standard deviation in herd i. This estimate will be used to standardise phenotypic observations prior to formation and solution of mixed model equations.

A sensible compromise is to pool "estimates" of the population "parameter" and the within-herd estimated standard deviation, weighted according to their respective "reliabilities", so that the sum of the weights equals unity (Brotherstone and Hill, 1986). Recalling the definition of $\tilde{\sigma}_{y_i}$ as the pooled (and therefore best) estimate of the within-herd standard deviation, the relevant formula is:

$$\tilde{\sigma}_{y_i} = w\sigma_y + (1-w)\hat{\sigma}_{y_i} \tag{6.2}$$

for

$$w = \frac{\text{var}(\hat{\sigma}_{y_i})}{[\text{var}(\sigma_{y_i}) + \text{var}(\hat{\sigma}_{y_i})]}.$$

Manipulation of Equation (6.2) gives Equation (3.19). Implementation of this approach requires specification of the parameter values σ_y and $\text{var}(\sigma_{y_i})$, the statistic $\hat{\sigma}_{y_i}$ and its sampling variance, $\text{var}(\hat{\sigma}_{y_i})$. Following Brotherstone and Hill (1986),

$$\text{var}(\hat{\sigma}_{yi}) \approx \frac{\sigma_y^2}{2n_i} \approx \frac{\overline{\hat{\sigma}}_y^2}{2n_i} \tag{6.3}$$

where n_i is the number of observations for herd i, and $\overline{\hat{\sigma}}_y^2$ is the unweighted average of within-herd variances similar to Equation (3.26). Now, a practical estimate (Brotherstone and Hill, 1986) of the variance of the within herd parameter values, i.e., the measure of the extent of heterogeneity of variance in k herds is:

$$\hat{\text{var}}(\sigma_{y_i}) = \frac{\sum_{i=1}^{k}[(\hat{\sigma}_{y_i} - \overline{\hat{\sigma}}_{y_i})^2 - \frac{\overline{\hat{\sigma}}^2_{y_i}}{2n_i}]}{k-1} \tag{6.4}$$

similar to Equation (3.25). That is, an estimate of the parameter value can be obtained by calculating the variance of the estimates and by reducing this quantity by an amount to account for the sampling variation in the estimated standard deviations.

Two other points are worth noting. Brotherstone and Hill recognise that (6.4) weights each within-herd estimate equally. Weighting within-herd estimates by their degrees of freedom might be statistically "better", but does not allow for biological increases in variance that may result from increasing herd size and therefore cow competition. They also suggest (p 302) that heterogeneous variations are "induced" by management and preferential treatment (rather than being a biological phenomenon, due to "inappropriate" choice of measurement scale, or inaccurate model specification).

Once defined the above notation, the method follows:

1. Calculate the within-herd standard deviation and mean.

2. Accumulate the sums of squares and cross-products to enable calculation of the regression coefficient (and intercept) for estimated standard deviation on estimated herd mean.

3. Accumulate the error sums of squares for regression at the same time.

4. Pass through the data again and adjust the error sums of squares as in (6.4) by subtracting a term for the sampling error of the estimated standard deviation. However, instead of using the term based on the average subclass variance, use the standard deviation predicted from the regression equation for the herd subclass mean.

5. Divide the adjusted error sum of squares by its degrees of freedom (i.e. k-2) to get an estimate of $\mathrm{var}(\sigma_{y_i})$.

6. Estimation of $\tilde{\sigma}_{y_i}$ then proceeds as in Equation (6.2) except that $\mathrm{var}(\sigma_{y_i})$ is calculated as described in the previous paragraph and $\hat{\sigma}_{y_i}$ is calculated separately for each herd, conditional on the herd mean, using the regression equation for regressing estimated subclass standard on estimated subclass mean.

In the suggested approach, the relationship between herd mean and standard

deviation is used. Philosophically, the method regresses the within-herd estimate of standard deviation to a population value, as before. However, the assumed population value is the population value appropriate for the given herd mean as calculated using the data-derived regression of standard deviation on herd mean. In some cases this regression might have an estimated coefficient near unity, therefore the standardisation will be not unlike the current use of ratios. This approach seems desirable to assuming, *a priori*, that the standard deviation and mean are directly proportional with a correlation coefficient of unity. Some evidence suggests that although the standard deviation increases with the mean, it does not double when the mean doubles. Accordingly, standardisation using the mean (or log) *overadjusts* the data.

In principle, regressing within-herd estimates to a population value that is conditional on the herd mean seems a minor modification of Brotherstone and Hill's (1986) approach. However, there are two minor complications. First, note that the estimated standard deviations are regressed on *estimated* herd mean, rather than the true herd mean. In theory we should account for variation in the dependent variable as it will not be measured without error, as required in the usual regression situation. The second problem is what do we use as a "parameter" value for the variance of the population standard deviation (i.e., $\mathrm{var}(\sigma_{y_i})$)? This value should now reflect the variance about the regression line relating standard deviation to herd mean.

It is not appropriate to use the variance of a predicted value based on a regression equation as this variance would increase for herd means that deviated from the overall average herd mean. That is, the "weight" in (6.2) placed on the subclass standard deviation would be greatest in herds with very low or very high means.

If the population variance was used, the value would be identical for all points along the regression line. This seems reasonable at first, however, it is more likely that the variation in within-herd standard deviation is greater in high mean herds than in low mean herds. Some further (data analysis) work should be carried out to check this relationship. Perhaps the coefficient of variation of herd standard deviation is relatively constant.

SUMMARY AND CONCLUSIONS

The correlation of 0.44 between mean and standard deviation confirms the presence of heterogeneity of variance for milkfat yield in New Zealand dairy cattle.

The effect of ignoring heterogenous variance was reflected on the genetic evaluation of sires. Product-moment and rank correlations between breeding values estimated from the three production levels were significantly lower than the expected correlations and the number of sires with significantly deviating breeding values were greater than the expected values. This indicates that sires with a large percentage of their daughters in contemporary groups with large variances would tend to be overevaluated.

Estimates of genetic correlations lower than one suggest that if sires are genetically evaluated on the basis of daughters located in only one level of production, responses to selection may be lower than the expected.

The methods investigated for the accounting of scaling were not appropriate. The mean correction and log transformation tended to overadjust rather than stabilise the standard deviations. This overadjustment was reflected in reducing, instead of increasing, the correlations between estimated and true breeding values at the three production levels. The adjustment by the standard deviation reduced the relationship between the standard deviation and mean but not completely (from 0.44 to 0.31) improving slightly the correlations between estimated and true breeding values.

Alternative methods, therefore, are required. A multiple-trait approach, considering the breeding value in each environment as an independent trait, seems theoretically best but it may require large amount of data and computer capacity. It seems most practical to pre-adjust data for some estimate of the herd or herd-year-age phenotypic standard deviation. Another alternative method is suggested in this study, which is derived the Bayesian approach and take into account any relationship between herd mean and standard deviation.

REFERENCES

Abubakar, B. Y., R. E. McDowell, and L. D. Van Vleck. 1987. Interaction of genotype and environment for breeding efficiency and milk production of Holstein in Mexico and Colombia. *Trop. Agric., Trinidad* **64**:17-22.

Ahlborn, G., and A. M. Bryant. 1992. Milk production and stocking rates: Production, economic performance and optimum stocking rates of Holstein-Friesian and Jersey cows. *Proc. N. Z. Soc. Anim. Prod.* **52**:7-9.

Ahlborn-Breier, G., and W. D. Hohenboken. 1991. Additive and nonadditive genetic effects on milk production in dairy cattle: evidence for major individual heterosis. *J. Dairy Sci.* **74**:592-602.

Ahlborn-Breier, G., B. W. Wickham, and J. W. Stichbury. 1987. Genetic progress in New Zealand Ayrshire, Friesian, and Jersey populations. *Proc. 4th. Anim. Sci. Congr. Asian-Australasian Assoc. Anim. Prod. Soc.*, pp 153.

Anderson, R. D. 1974. A study of biases in dairy sire evaluation. M. Agr. Sc. thesis, Massey University.

Arnold, J. W., J. K. Bertrand, and L. L. Benyshek. 1992. Animal model for genetic evaluation of multibreed data. *J. Anim. Sci.* **70**:3322-3332.

Bartlett, M. S. 1947. The use of transformations. *Biometrics* **3**:39-52.

Berger, P. J., G. R. Luecke, and J. A. Hoekstra. 1989. Iterative algorithms for solving mixed model equations. *J. Dairy Sci.* **72**:514-522.

Bereskin, B., and J. L. Lush. 1965. Genetic and environmental factors in dairy sire evaluation. III. Influence of environmental and other extraneous correlations among the daughters. *J. Dairy Sci.* **48**:356-360.

Blair, H. T., and E. J. Pollak. 1984. Comparison of an animal model and an equivalent reduced animal model for computational efficiency using mixed model methodology. *J. Anim. Sci.* **58**:1090-1096.

Boldman, K. G., and A. E. Freeman. 1990. Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. *J. Dairy Sci.* **73**:503-512.

Brotherstone, S., and W. G. Hill. 1986. Heterogeneity of variance amongst herds for milk production. *Anim. Prod.* **42**:297-303.

Brougham, R. W. 1959. The effects of season and weather on the growth rate of a ryegrass and clover pasture. *N. Z. J. Agric. Res.* **2**:283-296.

Brumby, P. J. 1961. The causes of differences in production between dairy herds. *Anim. Prod.* **3**:2777-294.

Bryant, A. M. 1980. Maximizing milk production from pasture. *Proc. N. Z. Grassland Assoc.* **42**:82-91.

Bryant, A. M. 1983. Does nitrogen fertilizer have a place in Dairying. *Proc. Ruakura Farmer's Conference,* pp 121-124.

Bryant, A. M. 1990. Optimum stocking and feed management practices. *Proc. Ruakura Farmer's Conference,* pp 55-59.

Bryant, A. M., R. A. MacDonald, and D. G. Clayton. 1982. Effects of nitrogen fertilizer on production of milk solids from grazed pastures. *Proc. N. Z. Grassland Assoc.* **43**:58-63.

Burnside, E. B., and J. C. Rennie. 1968. Sire evalaution and herd level of milk production. *Can. J. Anim. Sci.* **48**:97-102.

Campbell, A. G. 1966. Grazed pasture parameters III. Relationships of pasture and animal parameters in, and general discussion of, a stocking rate and grazing management experiment with dairy cows. *J. Agric. Sci., Cambridge* **67**:217-222.

Calo, L. L., R. E. McDowell, L. D. Van Vleck and P. D. Miller. 1973. Genetic aspects of beef production among Holstein-Friesian pedigree selected for milk production. *J. Anim. Sci.* **37**:676-682.

Carabaño, M. J., K. M. Wade, and L. D. Van Vleck. 1990. Genotype by environment interaction for milk and fat production across regions of the United States. *J. Dairy Sci.* **73**:173-180.

Coop, I. 1987. *Livestock in New Zealand.* N. Z. Soc. Anim. Prod. Christchurch.

Cooper, J. P. 1970. Potential production and energy conversion in temperate and tropical grasses. *Herbage Abstracts* **40**:1-15.

Conniffe, D., D. Browne, and M. J. Walshe. 1970. Experimental design for grazing trials. *J. Agric. Sci., Cambridge* **74**:339-342.

Cunningham, E. P. 1983. Structure of dairy cattle breeding in Western Europe and Comparisons with North America. *J. Dairy Sci.* 66:1579-1587.

Dalton, C., and C. Rumble. 1985. *50 years of artificial insemination and herd improvement in New Zealand.* Auckland Livestock Improvement Association, Hamilton.

Danell, B. 1982. Interaction between genotype and environment in sire evaluation for milk production. *Acta Agric. Scand.* **32**:33-46.

De Veer, J. C., and L. D. Van Vleck. 1987. Genetic parameters for first lactation milk yields at three levels of herd production. *J. Dairy Sci.* **70**:1434-1441.

Dempfle, L. 1984. Principles of estimation of breeding values. *Bulletin of International Dairy Federation* (183):24-32.

Dickerson, G. E. 1962. Implications of genetic-environmental interaction in animal breeding. *Anim. Prod.* **4**:47-63.

Dickerson, G. E. 1973. Inbreeding and heterosis in animals. In: *Proceedings of the Animal Breeding and Genetics Symposium held in Blacksburg, Virginia*, pp 54-71. American Society of Animal Sciences, Champaign, Illinois.

Dong, M. C., and I. L. Mao. 1990. Heterogeneity of (co)variance and heritability in different levels of intraherd milk-production variance and of average. *J. Dairy Sci.* **73**:843-851.

Eisenhart, C. 1947. The assumptions underlying the analysis of variance. *Biometrics* **3**:1-21.

Everett, R. W., and L. P. Jones. 1985. Breeding values - their calculation and potential use. In T. I. Phillips (Ed.) *Proc. of the conference on " The Challenge: Efficient Dairy Production"*, pp 406-419. Australian and New Zealand Soc. Anim. Prod. Albury-Wodonga.

Everett, R. W., and J. F. Keown. 1984. Mixed model sire evaluation with dairy cattle - Experience and genetic gain. *J. Anim. Sci.* **59**:529-541.

Everett, R. W., J. F. Keown, and J. F. Taylor. 1982. The problem of heterogeneous within herd error variances when identifying elite cows. *J. Dairy Sci. (Suppl. 1)* **65**:100 (Abstr.)

Falconer, D. S. 1952. The problem of environment and selection. *Amer. Nat.* **86**:293-298.

Falconer, D. S. 1960. *Introduction to Quantitative Genetics*. Oliver and Boyd. Edinburgh.

Garrick, D. J. 1988. Restricted maximum likelihood estimation of variance components for multiple traits with missing observations and an application to beef cattle. Ph. D. dissertation, Cornell University, Ithaca, New York.

Garrick, D. J. 1991. Best linear unbiased prediction for across-flock/year breeding values. *Proc. N. Z. Soc. Anim. Prod.* **51**:417-422.

Garrick, D. J., B. L. Harris, P. Shannon, and C. Sosa-Ferreyra. 1993. A prototype sire evaluation for New Zealand dairy cattle. *Proc. N. Z. Soc. Anim. Prod.* **53**:91-94.

Garrick, D. J., E. J. Pollak, R. L. Quaas, and L. D. Van Vleck. 1989. Variance heterogeneity in direct and maternal weight traits by sex and percent purebred for Simmental-sired calves. *J. Anim. Sci.* **67**:2515-2528.

Garrick, D. J., and D. L. Van Vleck. 1987. Aspects of selection for performance in several environments with heterogenous variances. *J. Anim. Sci.* **65**:409-421.

Gianola, D. 1986. On selection criteria and estimation of parameters when the variance is heterogeneous. *Theor. Appl. Genet.* **72**:671-677.

Gianola, D., J. L. Foulley, R. L. Fernando, C. R. Henderson, and K. A. Weigel. 1992. Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. *J. Dairy Sci.* **75**:2805-2823.

Greenhalgh, S. A., R. L. Quaas, and L. D. Van Vleck. 1986. Approximating prediction error variances for multiple trait sire evaluations. *J. Dairy Sci.* **69**:2877-2883.

Hammond, J. 1947. Animal breeding in relation to nutrition and environmental conditions. *Biol. Rev.* **22**:195-213.

Harris, B. L., R. G.Jackson, and D.J. Garrick. 1993. Animal Evaluation System for New Zealand Dairy Cattle. *Livestock Improvement, Hamilton.*

Hazel, L. N. 1943. The genetic basis for constructing selection indexes. *Genetics.* **28:**476-490.

Henderson, C. R. 1950. Estimation of genetic parameters. *Ann. Math. Stat.* **21**:309. (Abstr.)

Henderson, C. R. 1963. Selection index and expected genetic advance. In W. D. Hanson and H. F. Robinson (Ed.) *Statistical Genetics and Plant Breeding*. Pub. 982, pp 141-163. National Academy of Sciences-National Research Council, Washington, D. C.

Henderson, C. R. 1973. Sire evaluation and genetic trends. In: *Proceedings of the Animal Breeding and Genetics Symposium held in Blacksburg, Virginia*, pp 10-41. American Society of Animal Sciences, Champaign, Illinois.

Henderson, C. R. 1974. General flexibility of linear model techniques for sire evaluation. *J. Dairy Sci.* **57**:963-972.

Henderson, C. R. 1975a. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**:423-447.

Henderson, 1975b. Comparison of alternative sire evaluation methods. *J. Anim. Sci.* **41**:760-770.

Henderson, C. R. 1975c. Rapid method for computing the inverse of a relationship matrix. *J. Dairy Sci.* **58**:1727-1730.

Henderson, 1975d. Use of all relatives in intraherd prediction of breeding values and producing abilities. *J. Dairy Sci.* **58**:1910-1916.

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**:69-83.

Henderson, C. R. 1984. *Applications of linear models in animal breeding*. Univ. Guelph Press, Guelph, Ont., Canada.

Henderson, C. R. 1990a. Statistical methods in animal improvement: historical review. In D. Gianola and K. Hammond (Eds) *Advances in Statistical Methods for Genetic Improvement of Livestock.* Advanced Series In Agricultural Sciences **18**:2-14. Springer-Verlag, New York.

Henderson, C. R. 1990b. Accounting for selection and mating biases in genetic evaluations. In D. Gianola and K. Hammond (Eds) *Advances in Statistical Methods for Genetic Improvement of Livestock.* Advanced Series In Agricultural Sciences **18**:413-436. Springer-Verlag, New York.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. Von Krosigk. 1959. The estimation of environmental and genetic trends from records subjected to culling. *Biometrics* **15**:192-218.

Hickman, C. G., A. J. Lee and K. Gravir. 1969. Genotype x season x method interaction in evaluating dairy sires from progeny records. *Can. J. Anim. Sci.* **49**:151-155.

Hill, W. G. 1984. On selection among groups with heterogeneous variance. *Anim. Prod.* **39**:473-477.

Hill, W. G., M. R. Edwards, M. -K. A. Ahmed, and R. Thompson. 1983. Heritability of milk yield and composition at different levels and variability of production. *Anim. Prod.* **36**:59:68.

Hodgson, J. 1990. *Grazing Management, Science into Practice.* Longman Scientific and Technical, Essex.

Holmes, C. W. 1982. The effect of fertilizer nitrogen on the production of pasture and milk on dairy farmlets:1971-1974. *Proc. N. Z. Grassland Assoc.* **43**:53-57.

Holmes, C. W. 1986. Calving dates and their importance to the dairy farmer. *Dairyfarming Annual, Massey University*, pp 55-67.

Holmes, C. W. 1987. Milk production from managed grasslands. In R. W. Snaydon (Ed.) *Managed Grasslands, B. Analytical Studies*, pp 101-112. Elsevier, Amsterdam.

Holmes, C. W., and K. L. MacMillan. 1982. Nutritional management of the dairy herd grazing on pasture. In K. Mcmillan and V. Taufa. (Eds.) *Dairy Production from Pastures,* pp 244-274. N. Z. Soc. Anim. Prod., Hamilton.

Holmes, C. W., and W. Parker. 1992. Stocking rate and its effects on dairy farm productivity. *Dairyfarming Annual, Massey University*, pp 4-15.

Holmes, C. W., and G. F. Wilson. 1984. *Milk production from pasture.* Butterworth Agricultural books, Wellington.

Holmes, W. 1980. Grazing management. In W. Holmes (Ed.) *Grass; Its production and utilization,* pp 125-173. Blackwell, London.

Hudson, G. F. S. 1984. Extension of a reduced animal model to recursive prediction of breeding values. *J. Anim. Sci.* **59**:1164-1175.

Hutton, J. B. 1978. *Dairy cattle management, irrigation potential in South Auckland.* Ministry of Agriculture, N. Z. Ag Link No. FPP 167.

Jones, R. L., and R. L. Sandland. 1974. The relation between animal gain and stocking rate: Derivation of the relation from results of grazing trials. *J. Agric. Sci., Cambridge* **83**:335-342.

Kachman, S. D., and R. W. Everett. 1993. A multiplicative mixed model when the variances are heterogenous. *J. Dairy Sci.* **76**:859-867.

King, K. R., and C. R. Stockdale. 1980. The effects of stocking rate and nitrogen fertilizer on the productivity of irrigated perennial pasture grazed by dairy cows. 2. Animal production. *Aust. J. Exp. Agric. Anim. Husb.* **20**:537-542.

Livestock Improvement. 1991a. *Dairy Statistics 1990/91.*

Livestock Improvement. 1991b. *Services and Sires 1991.*

Lofgren, D. L., W. E. Vinson, R. E. Pearson, and R. Thompson. 1985. Heritability of milk yield at different herd means and variance for production. *J. Dairy Sci.* **68**:2737-2739.

Lush, J. L. 1945. *Animal Breeding Plans.* Iowa State College Press, Ames.

Lytton, V. H., and J. E. Legates. 1966. Sire by region interaction for production traits in dairy cattle. *J. Dairy Sci.* **49**:874-878.

Macdonald Committee Report. 1992. *Herd Testing and Related Services.* New Zealand Dairy Board.

MacMillan, K. L., V. K. Taufa, and M. C. Pearce. 1984. Calving patterns and their effects on herd production. *Proc. Ruakura Farmer's Conference,* pp 25-28.

Mason, I. L., and A. Robertson. 1956. The progeny testing of dairy bulls at different levels of production. *J. agric. Sci.* **47**:367-375.

McDaniel, B. T., and E. L. Corley. 1967. Relationships between sire evaluations at different herdmate levels. *J. Dairy Sci.* **50**:735-741.

McMeekan, C. P. 1956. Grazing management and animal production. *Proc. 7th Int. Grassl. Congr.*, pp 146-156.

McMeekan, C. P., and M. J. Walshe. 1963. The inter-relationships of grazing method and stocking rate in efficiency of pasture utilization by dairy cattle. *J. Agric. Sci., Cambridge* **61**:147-166.

Meinert, T. R., R. E. Pearson, W. E. Vinson, and B. G. Casell. 1988. Prediction of daughter's performance from dam's cow index adjusted for within herd variance. *J. Dairy Sci.* **71**:2220-2231.

Menendez-Buxadera, A., A. de los Reyes, D. Guerra, and J. Cordovi. 1989. Genetic variability of milk production from Holstein cows according to the level of milk yield of the herd. *Cuban J. Agric. Sci.* **23**:9-15.

Meuwissen, T. H. E., and J. H. J. Van der Werf. 1993. Impact of heterogeneous within herd variance on dairy cattle breeding schemes: a simulation study. *Livest. Prod. Sci.* **33**:31-41.

Meyer, K. 1989. Approximate accuracy of genetic evaluation under an animal model. *Livest. Prod. Sci.* **21**:87-100.

Mirande, S. L., and L. D. Van Vleck. 1985. Trends in genetic and phenotypic variances for milk production. *J. Dairy Sci.* **68**:2278-2286.

Misztal, I., and D. Gianola. 1987. Indirect solution of mixed model equations. *J. Dairy Sci.* **70**:716-723.

Misztal, I., and G. R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* **71** (Suppl. 2):27-32.

Mirande, S. L., and L. D. Van Vleck. 1985. Trends in genetic and phenotypic variances for milk production. *J. Dairy Sci.* **68**:2278-2286.

Mott, G. O. 1960. Grazing pressure and the measurement of pasture production. *Proc. 8th Int. Grassl. Congr.*, pp 606-611.

Nejati-Javaremi, A. 1991. Genetic evaluation of Holstein-Friesian and Jersey sires using records from pure- and cross-bred progeny in New Zealand. M. Agr. Sc. thesis, Massey University.

New Zealand Jersey Cattle Breed Association. 1932. *The Jersey In New Zealand.*

New Zealand Dairy Board. 1970. *46th Farm Production Report, 1969-70 season.*

Notter, D. R., and C. Diaz. 1993. Use of covariances between predicted breeding values to assess the genetic correlation between expressions of a trait in 2 environments. *Genet. Sel. Evol.* **25**:353-372.

Ortega, J. M. 1987. *Matrix Theory: A Second Course.* Plenum Press, New York and London.

Petersen, P. H. 1975. Genotype-environment interaction in milk production under Danish and Bulgarian-Czechoslovakian conditions. *Anim. Prod.* **21**:101-108.

Peterson, R. 1988. *Comparison of Canadian and New Zealand sires in New Zealand for production, weight and conformation traits.* Res. Bull. No. 5. Livestock Improvement Corporation Ltd. New Zealand Dairy Board, Hamilton.

Peterson, R. G., H. L. Lucas, and G. O. Mott. 1965. Relationship between rate of stocking and per animal and per acre performance on pasture. *Agronomy Journal* **57**:27-30.

Pfaffenberger, R. C., and J. H. Patterson. 1977. *Statistical methods for business and economics.* Richard D. Irwin, Inc. Homewood, Illinois.

Pollak, E. J., and R. L. Quaas. 1983. Definition of groups in sire evaluation models *J. Dairy Sci.* **66**:1503-1503.

Powell, R. L., and F. N. Dickinson. 1977. Progeny tests of sires in the United States and in Mexico. *J. Dairy Sci.* **60**:1768-1772.

Powell, R. L., H. D. Norman, and B. T. Weinland. 1983. Cow evaluation at different milk yields of herds. *J. Dairy Sci.* **66**:148-154.

Powell, R. L., and G. R. Wiggans. 1991. Animal model evaluations for Mexican Holsteins. *J. Dairy Sci.* **74**:1420-1427.

Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* **32**:949-953.

Quass, R. L. 1988. Additive genetic model with groups and relationships *J. Dairy Sci.* **71**:1338-1345.

Quaas, R. L., R. D. Anderson, and A. R. Gilmour. 1984. *BLUP School Handbook: Use of Mixed Model for Prediction and for Estimation of (Co)Variance Components.* Animal Genetics and Breeding Unit, University of Armidale, Australia.

Quaas, R. L., D. J. Garrick, and W. H. McElhenney. 1989. Multiple trait prediction for a type of model with heterogeneous genetic and residual covariance structures. *J. Anim. Sci.* **67**:2529-2535.

Quaas, R. L., and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* **51**:1277-1287.

Quaas. R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. *J. Dairy Sci.* **64**:1868-1872.

Radcliffe J. E., and J. A. Baars. 1987. The productivity of temperate grasslands. In R. W. Snaydon (Ed.) *Managed Grasslands, B. Analytical studies*, pp 7-17. Elsevier, Amsterdam.

Rendel, J. M., and A. Robertson. 1950. Estimation of genetic gain in milk by selection in a closed herd of dairy cattle. *J. Genetics* **50**:1-8.

Roberts, A. H. C., and N. Thompson. 1989. Use of nitrogen fertilizer for intensive dairy production. *Proc. of workshop on Nitrogen in New Zealand,* pp 42-56. Agric. and Hort. Fert. and Lime Res. Centre, Massey University.

Robertson, A. 1959. The sampling variance of the genetic correlation coefficient. *Biometrics* **15**:469-485.

Robertson, A., L. K. O'Connor, and J. Edwards. 1960. Progeny testing dairy bulls at different management levels. *Anim. Prod.* **2**:141-152.

Robinson, G. K. 1986. Groups effects and computing strategies for models for estimating breeding values. *J. Dairy Sci.* **69**:3106-3111.

Robinson, G. K., and L. P. Jones. 1987. Approximations for prediction error variances. *J. Dairy Sci.* **70**:1623-1632.

SAS. 1985. *SAS User's Guide: Statistics*. SAS Institute Inc, Cary, NC.

Schaeffer, L. M., and B. W. Kennedy. 1986a. Computing solutions to mixed model equations. *Proc. 3rd World Congress on Genetics Applied to Livestock Production*. **XII**:382-393. Lincoln, Nebraska.

Schaeffer, L. M., and B. W. Kennedy. 1986b. Computing strategies for solving mixed model equations. *J. Dairy Sci.* **69**:575-579.

Schaeffer, L. R., J. W. Wilton, and R. Thompson. 1978. Simultaneous estimation of variance and covariance components from multitrait mixed model equation. *Biometrics* **34**:199-208.

Schwarz, H. R., H. Rutishauser, and E. Stiefel. 1973. *Numerical Analysis of Symmetric Matrices*. Prentice-Hall, Inc. Eaglewood Cliffs, N. J.

Searle, S. R. 1964. Review of sire-proving methods in New Zealand, Great Britain, and New York State. *J. Dairy Sci.* **47**:402-413.

Searle, S. R. 1971. *Linear Models*. John Wiley and Sons, New York.

Searle, S. R. 1982. *Matrix Algebra Useful for Statistics*. John Wiley and Sons. New York.

Shannon, P. 1974. The evaluation of dairy sire progeny tests in New Zealand. *Proc. of 1st World Congress on Genetics Applied to Livestock Production* **2**:105:110.

Snedecor, G. W., and W. G. Cochran. 1980. *Statistical Methods*. The Iowa State University Press, Ames, IA.

Stichbury, J. M. 1957. New developments in herd improvement. *Proc. Ruakura Farmer's Conference*, pp 196-205.

Stockdale, C. R., and K. R. King. 1980. The effects of stocking rate and nitrogen fertilizer on the productivity of irrigated perennial pasture grazed by dairy cows. 1. Pasture production, utilization and composition. *Aust. J. Exp. Agric. Anim. Husb.* **20**:529-536.

Thompson, N. A., A. H. C. Roberts, T. A. Judd, and J. S. Clough. 1991. Maximising dairy production by using nitrogen fertilizer and calving early. *Proc. N. Z. Grassland Assoc.* **53**:85-90.

Thompson, R. 1979. Sire evaluation. *Biometrics* **35**:339-353.

Ufford, G. R., C. R. Henderson, and L. D. Van Vleck. 1979. Computing algorithms for sire evaluation with all lactation records and natural service sires. *J. Dairy Sci.* **62**:511-519.

Van der Werf, J., L. Schaeffer, J. Van Arendonk, and A. Groen. 1991. *Breeding Value Estimation: Lecture Notes*. Department of Animal Breeding, Wageningen Agricultural University, Wageningen.

VanRaden, P. M., and A. E. Freeman. 1985. Rapid method to obtain bounds on accuracies and prediction error variances in mixed models. *J. Dairy Sci.* **68**:2123-2133.

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of animal model information. *J. Dairy Sci.* **74**:2737-2746.

Van Vleck, L. D. 1963. Genotype and environment in sire evaluation. *J. Dairy Sci.* **46**:983-987.

Van Vleck, L. D. 1979. *Notes on the theory and application of selection principles for the genetic improvement of animals*. Cornell University Ithaca. N. Y.

Van Vleck, L. D. 1987. Selection when traits have different genetic and phenotypic variances in different environments. *J. Dairy Sci.* **70**:337-344.

Van Vleck, L. D. 1988. Alternatives for evaluations with heterogenous genetic and environmental variances. *J. Dairy Sci. (Suppl. 2)* **71**:83.

Van Vleck, L. D. 1992. The revolution in statistical computing: from Least Squares to DFREML. *Proc. 41th Annual National Breeders Roundtable,* pp 1-35. St. Louis, Missouri.

Van Vleck, L. D., and D. J. Dwyer. 1985. Comparison of iterative procedures for solving equations for sire evaluation. *J. Dairy Sci.* **68**:1006-1014.

Van Vleck, L. D., E. J. Pollak, and E. A. B. Oltenacu. 1987. *Genetics for Animal Science.* W. H. Freeman and Company, New York.

Vinson, W. E. 1987. Potential bias in genetic evaluations from differences in variation within herds. *J. Dairy Sci.* **70**:2450-2455.

Visscher, P. M., and W. G. Hill. 1992. Heterogeneity of variance and dairy cattle breeding. *Anim. Prod.* **55**:321-329.

Visscher, P. M., R. Thompson, and W. G. Hill. 1991. Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livest. Prod. Sci.* **28**:273-290.

Vishwanath, R. 1992. Recent developments in semen technology. *Dairyfarming Annual, Massey University*, pp 155.

Weigel, K. A., and D. Gianola. 1993. Computationally simple Bayesian method for estimation of heterogenous within-herd phenotypic variances. *J. Dairy Sci.* **76**:1455-1465.

Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. *J. Dairy Sci.* **71**:1310-1318.

Westell, R. A., and L. D. Van Vleck. 1987. Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *J. Dairy Sci.* **70**:1006-1017.

White, D. H. 1987. Stocking rate. In R. W. Snaydon (Ed.) *Managed Grasslands*, pp 227-238. Elsevier, Amsterdam.

Wickham, B. W. 1984. Progeny testing of dairy bulls in New Zealand. *Bulletin of International Dairy Federation* (183):133-138.

Wickham, B. W. 1993. Size and efficiency in Holstein-Friesian animals. *N. Z. Holstein-Friesian Journal.* February, pp 68-71.

Wickham, B. W., and J. W. Stichbury. 1980. Cow and sire evaluation in New Zealand. *Proceedings of the symposium on Dairy Cattle Sire Evaluation.* Warsaw, Poland.

Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988a. Animal model evaluation of Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. *J. Dairy Sci.* **71**:1319-1329.

Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988b. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. *J. Dairy Sci.* **71** (Suppl. 2):54-69.

Wiggans, G. R., and P. M. VanRaden. 1991. Method and effect of adjustment for heterogeneous variance. *J. Dairy Sci.* **74**:4350-4357.

Winkelman, A., and L. R. Schaeffer. 1988. Effect of heterogeneity of variance on dairy sire evaluation. *J. Dairy Sci.* **71**:3033-3039.

Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* **56**:330-338.

Wright, D. F., and R. M. Pringle. 1983. Stocking rate effects in dairying. *Proc. N. Z. Soc. Anim. Prod.* **43**:97-100.

APPENDIX I

This Appendix I contains the comparison between product-moment and rank correlations of breeding values for lactation milkfat yield of sires evaluated at different levels of production and with variable number of daughters. Table I.1 contains the comparison between correlation coefficients for raw data. Table I.2 shows the comparison between correlation coefficient for data corrected by the herd-year-age mean. Table I.3 contains the comparison between correlation coefficients for data adjusted by the weighted combination of the population standard deviation and the posterior estimate of herd-year-age standard deviation. And, Table 1.4 shows the comparison between correlation coefficients for data after logarithmic transformation.

Table I.1. Product-moment (r_p) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with variable minimum number of daughters. Untransformed data.

Minimum number of daughters at each production level	Number of sires	Levels of production compared					
		Low-Medium		Low-High		Medium-High	
		r_p	r_s	r_p	r_s	r_p	r_s
10	1151	0.72	0.70	0.67	0.65	0.72	0.69
20	486	0.76	0.76	0.73	0.72	0.78	0.77
30	351	0.78	0.77	0.74	0.74	0.82	0.82
40	306	0.81	0.80	0.77	0.77	0.86	0.86
50	272	0.83	0.83	0.80	0.80	0.88	0.87
60	248	0.85	0.83	0.82	0.80	0.89	0.88
70	230	0.86	0.85	0.85	0.84	0.92	0.91
80	216	0.88	0.87	0.87	0.85	0.93	0.91
90	209	0.90	0.90	0.88	0.87	0.93	0.91
100	198	0.92	0.91	0.91	0.90	0.93	0.92

* P<.05

Table I.2. Product-moment (r_p) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with variable minimum number of daughters. Data corrected by the HYA mean.

Minimum number of daughters at each production level	Number of sires	Levels of production compared					
		Low-Medium		Low-High		Medium-High	
		r_p	r_s	r_p	r_s	r_p	r_s
10	1151	0.72	0.70	0.67	0.65	0.71	0.68
20	486	0.78	0.76	0.74	0.73	0.78	0.77
30	351	0.79	0.78	0.75	0.76	0.83	0.82
40	306	0.82	0.81	0.78	0.77	0.86	0.86
50	272	0.83	0.82	0.80	0.81	0.88	0.87
60	248	0.84	0.83	0.81	0.81	0.89	0.88
70	230	0.86	0.84	0.84	0.83	0.92	0.90
80	216	0.87	0.87	0.86	0.85	0.92	0.90
90	209	0.88	0.89	0.88	0.87	0.92	0.91
100	198	0.91	0.90	0.90	0.89	0.93	0.91

* P<.05

Table I.3. Product-moment (r_p) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with variable minimum number of daughters. Data adjusted by the HYA sd.

Minimum number of daughters at each production level	Number of sires	Levels of production compared					
		Low-Medium		Low-High		Medium-High	
		r_p	r_s	r_p	r_s	r_p	r_s
10	1151	0.73	0.71	0.69	0.70	0.73	0.71
20	486	0.78	0.76	0.75	0.74	0.80	0.78
30	351	0.80	0.78	0.77	0.77	0.84	0.82
40	306	0.82	0.81	0.79	0.78	0.87	0.86
50	272	0.84	0.84	0.82	0.82	0.89	0.88
60	248	0.85	0.84	0.84	0.82	0.90	0.88
70	230	0.86	0.86	0.86	0.85	0.92	0.91
80	216	0.88	0.88	0.88	0.87	0.93	0.91
90	209	0.90	0.90	0.90	0.89	0.93	0.92
100	198	0.92	0.91	0.92	0.91	0.94	0.92

* P<.05

Table I.4. Product-moment (r_p) and rank (r_s) correlations between breeding values for milkfat yield of sires evaluated at different levels of production and with variable minimum number of daughters. Data after log transformation.

Minimum number of daughters at each production level	Number of sires	Levels of production compared					
		Low-Medium		Low-High		Medium-High	
		r_p	r_s	r_p	r_s	r_p	r_s
10	1151	0.68	0.66	0.62	0.60	0.70	0.66
20	486	0.73	0.73	0.69	0.69	0.75	0.73
30	351	0.74	0.74	0.70	0.72	0.79	0.78
40	306	0.77	0.77	0.72	0.73	0.84	0.83
50	272	0.79	0.80	0.76	0.78	0.87	0.86
60	248	0.81	0.80	0.78	0.78	0.87	0.86
70	230	0.83	0.82	0.81	0.82	0.90	0.89
80	216	0.84	0.84	0.83	0.83	0.90	0.88
90	209	0.87	0.87	0.85	0.85	0.91	0.89
100	198	0.89	0.88	0.89	0.88	0.92	0.90

* P<.05

APPENDIX II

This Appendix II contains the number of sires and averages and standard deviations of breeding values for lactation milkfat yield of sires evaluated at different levels of production and with different minimum number of daughters at each production level for sires with significantly deviating breeding values. Table II.1 holds statistics for raw data and Tables II.2, II.3, and II.3 hold statistics for the three methods of reducing heterogeneity of variance, namely, herd-year-age mean correction, herd-year-age standard deviation adjustment, and log transformation, respectively.

Table II.1. Number of sires (m_o) and averages ($\bar{\hat{u}}$) and standard deviations ($sd_{\hat{u}}$) of breeding values for milkfat yield estimated at different levels of production and with different minimum number of daughters for sires with significantly deviating* breeding values. Untransformed data.

Minimum number of daughters at each production level	Number of sires	Levels of production compared: Low - Medium			Low - High			Medium - High		
		m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$	m_o	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$
10	1151	50	-3.2±13.3	-2.7±13.1	90	-2.4±12.4	0.3±15.5	77	-2.6±13.0	1.8±16.2
20	486	33	-2.5±12.6	-0.3±12.5	61	-0.4±10.7	1.6±14.9	51	-0.9±12.7	4.5±15.5
30	351	29	-0.7±11.1	-0.3±13.0	56	0.4±10.0	1.3±14.7	43	-0.4±13.2	3.2±15.9
40	306	27	-0.5±11.2	-0.3±13.2	51	1.0± 9.5	1.7±14.3	38	0.5±13.3	4.2±15.0
50	272	22	-0.2±11.4	-1.0±11.6	49	1.4± 9.4	1.2±14.4	36	1.5±12.6	4.3±15.3
60	248	19	1.7±10.0	0.2±11.9	46	2.1± 8.8	2.5±13.6	35	2.1±12.3	5.4±14.0
70	230	19	1.7±10.0	0.2±11.9	44	2.2± 9.0	1.6±13.1	32	2.6±12.6	4.5±14.0
80	216	17	2.0±10.1	0.2±12.3	42	2.4± 8.8	1.9±13.4	31	3.5±11.8	4.8±13.8
90	209	15	1.1±10.5	1.6±12.5	41	2.3± 8.9	2.3±13.2	30	4.0±11.7	4.9±14.1
100	198	14	2.4± 9.5	1.6±13.0	39	2.9± 8.4	2.1±13.5	30	4.0±11.7	4.9±14.1

* $P<.05$

Table II.2. Number of sires (m_o) and averages ($\bar{\hat{u}}$) and standard deviations ($sd_{\hat{u}}$) of breeding values for milkfat yield estimated at different levels of production and with different minimum number of daughters for sires with significantly deviating* breeding values. Data corrected by the HYA mean.

Minimum number of daughters at each production level	Number of sires	Levels of production compared								
		Low - Medium			Low - High			Medium - High		
		m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$	m_o	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$
10	1151	39	-1.3±9.4	-1.4±8.0	60	-3.7±9.3	0.3±8.2	51	-2.4±8.7	2.7±8.0
20	486	22	1.0±8.1	0.0±6.6	39	-0.8±8.1	-0.4±6.8	33	-0.4±7.9	4.0±7.1
30	351	21	2.0±7.0	0.0±6.8	36	-0.6±7.1	0.6±8.3	26	-0.6±8.1	3.6±7.0
40	306	19	2.1±6.7	-0.2±7.1	31	0.5±7.4	-0.1±7.0	22	-0.0±8.2	4.1±7.2
50	272	18	2.3±6.9	-0.7±6.9	29	1.0±7.4	0.4±8.3	20	1.1±7.4	4.3±7.3
60	248	16	3.3±6.1	-0.1±7.1	27	1.8±6.6	1.2±7.9	20	1.1±7.4	4.3±7.3
70	230	16	3.3±6.1	-0.1±7.1	25	2.0±6.8	0.6±6.8	17	1.8±7.7	3.4±7.3
80	216	15	3.1±6.3	0.2±7.2	23	2.3±6.7	1.1±6.9	17	1.8±7.7	3.4±7.3
90	209	13	2.7±6.6	1.3±7.2	22	2.2±6.8	1.2±7.3	16	2.3±7.6	3.5±7.6
100	198	12	3.8±5.5	1.3±7.5	21	2.8±6.3	1.1±7.4	16	2.3±7.6	3.5±7.6

* $P<.05$

Table II.3. Number of sires (m_o) and averages ($\bar{\hat{u}}$) and standard deviations ($sd_{\hat{u}}$) of breeding values for milkfat yield estimated at different levels of production and with different minimum number of daughters for sires with significantly deviating* breeding values. Data adjusted by the HYA sd.

Minimum number of daughters at each production level	Number of sires	Levels of production compared: Low - Medium			Low - High			Medium - High		
		m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$	m_o	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$
10	1151	32	-3.7±11.9	-0.7±11.5	58	-2.3±11.8	0.4±14.1	49	-0.0±11.5	4.5±13.0
20	486	24	-1.6±11.0	0.5±10.5	42	-0.0±10.9	1.4±13.7	38	1.3±11.6	5.4±12.0
30	351	21	0.8± 9.3	0.4±11.1	40	0.2±11.1	0.6±13.5	33	1.2±11.8	5.2±12.2
40	306	18	1.4± 9.4	-0.1±11.0	36	1.0±10.5	1.3±13.4	30	1.3±12.2	5.5±12.4
50	272	17	2.2± 9.1	-0.5±11.2	35	1.4±10.3	0.2±10.6	28	2.5±11.6	5.7±12.6
60	248	17	2.2± 9.1	-0.5±11.2	32	2.4± 9.7	2.5±12.8	28	2.5±11.6	5.7±12.6
70	230	17	2.2± 9.1	-0.5±11.2	30	2.6± 9.9	1.3±10.7	25	3.4±11.9	4.7±12.8
80	216	15	2.5± 9.2	-0.4±11.7	28	1.9±10.9	1.9±12.8	25	3.4±11.9	4.7±12.8
90	209	13	1.7± 9.7	1.3±11.7	27	2.9± 9.9	2.6±12.4	24	4.0±11.7	4.7±13.1
100	198	12	3.1± 8.6	1.4±12.2	26	3.6± 9.4	2.7±12.7	24	4.0±11.7	4.7±13.1

* $P<.05$

Table II.4. Number of sires (m_o) and averages ($\bar{\hat{u}}$) and standard deviations ($sd_{\hat{u}}$) of breeding values for milkfat yield estimated at different levels of production and with different minimum number of daughters for sires with significantly deviating* breeding values. Data after log transformation.

Minimum number of daughters at each production level	Number of sires	Levels of production compared: Low - Medium			Low - High			Medium - High		
		m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	m_o	$\bar{\hat{u}}_L \pm sd_{\hat{u}_L}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$	m_o	$\bar{\hat{u}}_M \pm sd_{\hat{u}_M}$	$\bar{\hat{u}}_H \pm sd_{\hat{u}_H}$
10	1151	58	-4.8±10.4	-2.9±8.3	73	-5.0±11.1	-2.3±7.3	58	-4.6±9.3	-0.1±10.1
20	486	29	-2.5± 9.4	-2.1±6.7	45	-1.1± 9.4	-0.5±8.7	33	-1.9±7.9	3.1± 8.5
30	351	26	-1.6± 8.3	-2.1±7.0	42	-0.7± 8.9	-0.8±7.0	27	-2.5±7.5	2.6± 8.6
40	306	23	-1.1± 8.1	-2.3±7.3	37	0.1± 8.5	-0.3±8.6	21	-1.7±7.7	3.6± 8.1
50	272	22	-0.9± 8.2	-2.9±6.9	32	0.5± 8.0	-0.8±8.9	19	-0.7±7.1	3.8± 8.3
60	248	18	1.2± 6.3	-1.7±6.7	28	1.9± 6.5	0.3±8.8	18	-0.2±7.0	4.0± 8.5
70	230	17	1.6± 6.2	-2.0±6.8	26	2.0± 6.7	-0.8±8.2	15	0.2±7.3	2.7± 8.4
80	216	16	1.4± 6.3	-1.9±7.0	24	1.7± 6.9	-0.4±8.4	15	0.2±7.3	2.7± 8.4
90	209	14	0.6± 6.3	-1.1±7.1	23	1.5± 7.0	0.2±7.9	14	0.8±7.3	2.7± 8.8
100	198	13	1.6± 5.1	-1.3±7.4	22	2.2± 6.3	0.1±8.1	14	0.8±7.3	2.7± 8.8

* $P<.05$