GENETIC EVALUATION OF HOLSTEIN-FRIESIAN AND JERSEY SIRES USING RECORDS FROM PURE- AND CROSS-BRED PROGENY IN NEW ZEALAND.

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

ARDESHIR NEJATI-JAVAREMI

1991

In the name of God the most compassionate the most merciful.

ACKNOWLEDGMENT

First of all thanks to God for giving me the ability of learning.

I am especially indebted to my supervisors Dr. D.J. Garrick and Dr. H.T. Blair for their appropriate guidance, comments, suggestions and encouragements. Any time I needed their help I was provided friendly answer.

I would also like to thank other staff members of Animal Science Department of Massey University especially Professor S.N.McCutcheon, head of the department and Associate Professors D.D.S. Mackenzie and M.F. McDonald.

Grateful acknowledgment is made to Professor R.D. Anderson, Dean of Faculty of Agricultural and Horticultural Sciences for his many kindnesses.

Full scholarship given by Ministry of Jehad-e-Sazandagi and Ministry of Culture and Higher Education of Iran is gratefully appreciated.

Friendly environment provided by my fellow graduate students in Animal Science Department is much appreciated. Especially friendship of Mr. James Maas will never be forgotten.

Presence of other Iranian graduate students at Massey University made me feel living home. It is my pleasure to thank all of them.

Livestock Improvement Corporation of New Zealand Dairy Board has kindly permitted me to have access to their data base which was the main source of data for this study. My gratitude is extended to Mrs. Gisela Ahlborn-Breier for helping me in this matter.

My family and myself have enjoyed hospitality of friendly people of New Zealand during more than two years of staying in this country. I would like to thank you all very much.

I am thankful to my parents, brothers and sister for their help and encouragement over the years especially when I was away from home. Without their help I could not achieve this.

My sincere thanks is due to my wife for her support, patience, encouragement and undertaking my own duty to educate our children. Patience of my son, Alireza and my daughter Mahdieh over the nights I was working late is appreciated. Their smile encouraged me to cope with difficulties.

ABSTRACT

Milk Yield, Milk Fat Yield, Milk Protein Yield and Days in Milk of 72,480 dairy cows of Holstein-Friesian, Jersey and their reciprocal crossbred were analysed. The main objective of this study was to investigate possible use of crossbred progeny records to genetically evaluate sires.

Bulls of each breed were evaluated separately using their purebred, crossbred or both purebred and crossbred progeny. First, crossbred progeny were used without including genetic groups. Secondly, crossbred progeny were used with genetic groups included. Rank correlations for different types of evaluations were calculated. In total, 10 different comparisons, 5 for each breed, were performed. The expected correlation of ranks of sires obtained using different data sets were estimated where applicable.

Reliability and Prediction Error Variance of sire proofs were estimated both by an approximate method and direct calculation. Over-estimation of reliability and under-estimation of Prediction Error Variance by the approximate method was given.

High correlations between ranks of Holstein-Friesian sires evaluated using different data sets were observed, while, the correlation between ranks of Jersey sires evaluated using purebred progeny with ranks of the same sires evaluated using only crossbred progeny was less than expected. Correlations of ranks of Jersey sires evaluated using all progeny with ranks of the same sires evaluated using only crossbred progeny were also low.

After plotting EBVs of sires of each breed obtained using only purebred against EBVs of the same sires obtained using both purebred and crossbred progeny, two lines with slightly different slopes were observed. The reasons for the formation of these two lines were investigated. It was found that the number of effective crossbred progeny of sires was affecting the regression of EBVs of sires obtained using all progeny on EBVs of same sires obtained using only purebred progeny.

It was concluded that crossbred progeny of Holstein-Friesian sires may be used to assist in their evaluation under New Zealand conditions, but, further research is recommended before using crossbred progeny of Jersey sires in sire evaluation.

TABLE OF CONTENTS

Page

Acknowledgments ii
Abstracts iv
List of Tables ix
List of Figures xii

CHAPTER I: DAIRY FARMING IN NEW ZEALAND 2

1.1 Introduction 2

1.2 Dairy Cattle in New Zealand and their breeds 3

1.3 Production System 4
1.3.1 Calving Date 6
1.3.2 Stocking Rate 7

1.4 Genetic Improvement 8
1.4.1 Data Collection 10
1.4.2 Herd Testing 11
1.4.3 Sire Evaluation 11
1.4.4 Artificial Insemination 13

CHAPTER II: SIRE EVALUATION 16

2.1 Sire Evaluation 16

2.2 Best Linear Unbiased Prediction 19
2.2.1 Contemporary Groups 23
2.2.2 Effective Progeny Number 24
2.2.3 Prediction Error Variance 25
2.2.4 Using Relationship Between Bulls 27
2.2.5 Reliability 28
2.2.6 Accounting for Genetic Groups in the Model 31

2.3 Solving Linear Equations 34

CHAPTER III: MATERIALS AND METHOD 41

3.1 Data Description 41

3.2 Model of Sire Evaluation 42
3.2.1 Sire Evaluation without inclusion of Genetic Groups 42
3.2.1.1 Calculating Residual Variances 44
3.2.1.2 Adding $\mathbf{G}^{-1}$ to the LHS Matrix 48

3.2.2 Sire Evaluation with Inclusion of Genetic Groups 49

3.3 Correlation Between Sire Evaluations from Different Methods 50

3.4 Calculating Effective Progeny Number, Variance of Prediction Error and Reliability 53

CHAPTER IV: RESULTS 58

4.1 General Statistics for Breeds 59
4.2 Milk Volume 62
4.3 Milk Fat 70
4.4 Milk Protein 73
4.5 Days in Milk 76

CHAPTER V: DISCUSSION 80

REFERENCES 89

APPENDICES 97

Appendix I 98
Appendix II 103

Appendix III 108
Appendix IV 119

LIST OF TABLES

Page

Table 4.1.1: Mean and standard deviation of various traits for different breeds in the sire evaluation 60

Table 4.1.2: Mean and standard deviation of various traits for different breeds in the sire evaluation(Adjusted to equal days in milk) 61

Table 4.1.3: Standard deviation of various traits for different breeds (Pooled within HYS) 61

Table 4.2.1: Distribution of different statistics in evaluation of sires (for milk volume) 64

Table 4.2.2: Mean over-estimation(%) of r_{TI}^2 and other statistics with different number of EPN (for milk volume) 65

Table 4.2.3: Mean under-estimation(%) of PEV and other statistics with different number of EPN (for milk volume) 66

Table 4.2.4: Correlation of true and estimated r_{TI}^2 and, true and estimated PEV with different EPN 67

Table 4.2.5: Rank correlations of sire evaluations using different data sets (for milk volume) 67

Table 4.3.1: Rank correlations of sire evaluations using different data sets (for milk fat) 71

Table 4.3.2: Distribution of different statistics in evaluating sires (for milk fat) 72

Table 4.4.1: Rank correlations of sire evaluations using different data sets (for milk protein) 74

Table 4.4.2: Distribution of different statistics in evaluating sires (for milk protein) 75

Table 4.5.1: Rank correlation of sire evaluations using different data sets (for days in milk) 77

Table 4.5.2: Distribution of different statistics in evaluating sires (for days in milk) 78

Table 5.1: Coefficients of regression of EBVs of Holstein sires evaluated using both pure- and crossbred progeny on the EBVs of the sires evaluated using only purebred progeny and with different effective crossbred progeny numbers (for milk volume) 86

Table 5.2: Coefficients of regression of EBVs of Holstein sires evaluated using both pure- and crossbred progeny on the EBVs of the sires evaluated using only purebred progeny and with different percentage of effective crossbred progeny numbers (for milk volume) 87

Table A.2.1: Comparison of correlations of sire evaluations using different data sets for milk volume when different lower limits for progeny plus ½ grand progeny numbers were used 104

Table A.2.2: Comparison of correlations of sire evaluations using different data sets for milk fat when different lower limits for progeny plus ½ grand progeny numbers were used 105

Table A.2.3: Comparison of correlations of sire evaluations using different data sets for milk protein when different lower limits for progeny plus ½ grand progeny numbers were used 106

Table A.2.4: Comparison of correlations of sire evaluations using different data sets for days in milk when different lower limits for progeny plus ½ grand progeny numbers were used 107

Table A.4.1: Different estimates of heritabilities of dairy performance traits 120

LIST OF FIGURES

Figure A.3.1: Comparing two data sets (purebred progeny vs. all progeny) in ranking Holstein sires for different traits 109

Figure A.3.2: Comparing two data sets (crossbred progeny without including genetic groups vs. all progeny) in ranking Holstein sires for different traits 110

Figure A.3.3: Comparing two data sets (purebred progeny vs. all progeny) in ranking Jersey sires for different traits 111

Figure A.3.4: Comparing two data sets (crossbred progeny without including genetic groups vs. all progeny) in ranking Jersey sires for different traits 112

Figure A.3.5: Comparing two data sets (crossbred progeny with including genetic groups vs. all progeny) in ranking Holstein sires for different traits 113

Figure A.3.6: Comparing two data sets (crossbred progeny with including genetic grops vs. all progeny) in ranking Jersey sires for different traits 114

Figure A.3.7: Comparing two data sets (purebred progeny vs. crossbred progeny without including genetic groups) in ranking Holstein sires for different traits 115

Figure A.3.8: Comparing two data sets (purebred progeny vs. crossbred progeny without including genetic groups) in ranking Jersey sires for different traits 116

Figure A.3.9: Comparing two data sets (purebred progeny vs. crossbred

progeny with including genetic groups) in ranking Holstein sires for different traits 117

Figure A.3.10: Comparing two data sets (purebred progeny vs. crossbred progeny with including genetic groups) in ranking Jersey sires for different traits 118

CHAPTER ONE

CHAPTER ONE

DAIRY FARMING IN NEW ZEALAND

1.1- Introduction:

The trend in the dairy cattle population of New Zealand is an increasing proportion of Holstein Friesian cattle and a decreasing proportion of Jersey cattle. Different types of crossbreeding dairy cows in different parts of the world are being practiced, for example, increased exchange of semen to upgrade native low producing cows of some countries with mating to other genetically improved exotic breeds. Therefore, the observed genetic effects in the resulting population might not be solely additive and may contain non additive genetic effects.

Evaluating sires of each breed is usually done using their purebred progeny. Possible use of crossbred progeny of young bulls of progeny testing program can reduce the costs of progeny testing and increase the reliability of estimates.

Reciprocal crossbreeding of two or more breeds in places where farmers intend to keep a combination of these breeds needs the sires of different breeds to be jointly evaluated. This type of evaluation makes direct comparison of sires of different breeds possible. When two sires have equal breeding values for specific traits, farmers would be able to select the sire of that breed which helps them to keep a desirable breed composition of their herds.

Existence of significant number of herds containing both Holstein-Friesian and Jersey cows as well as their reciprocal crossbred progeny in New Zealand provides the opportunity of evaluating sires of these two breeds using different combinations of their progeny and comparing the proofs obtained by different evaluations.

The purpose of this research is to investigate the possible use of crossbred progeny to evaluate bulls of Holstein and Jersey breeds and compare the ranks of sires obtained using different data sets.

However, extending the results of this study to other breeds and/or other countries needs further investigation.

1.2- Dairy Cattle in New Zealand and Their Breeds:

The New Zealand dairy industry started in 1814 by landing a bull and two heifers in the Bay of Islands. Further small importations of cattle took place and in the early decades of the 19th century pioneering efforts resulted in the first signs of trading in dairy produce (Dalton and Rumble, 1985).

In 1932, about two thirds of New Zealand dairy cows in milk (out of total of 1,292,873) were Jerseys or Jersey grades (NZJCBA, 1932). Over the next sixty years, the breed composition of the New Zealand dairy herd has substantially changed (MacMillan et.al., 1981). A predominantly Jersey herd has become a "Friesian x Jersey" crossbred herd, with inseminations using semen from Friesian sires exceeding 50% of all inseminations since 1970.

Over the 20th century the number of dairy herds in New Zealand has decreased

but the number of cows per herd has increased. In 1920-25 the number of herds supplying dairy factories was about 53,000, however, this number decreased to 15,581 in 1979-80. The average number of cows per herd was 56 in 1900-1901 but this increased to 156 per herd in 1979-80 (Holmes, 1984).

In 1953, 100% of artificial insemination (AI) was based on semen from Jersey sires. This percentage decreased over 30 years to 40%, due to substitution of primarily Friesian semen. In 1979, 36% of all dairy herds were mainly Friesian, 37% were mainly Jersey and 28% were mixed of two breeds. The majority of town-supply herds are Friesian because of this breed's ability to produce larger volumes of milk (Holmes, 1984). It was expected that 46% of herds would be mainly Friesian, 32% mainly Jersey and 23% mainly mixed in 1985.

An additional reason that New Zealand dairy farmers prefer Holstein-Friesian to Jersey cows is the ability of this breed to produce more valuable male calves for beef production (Holmes, 1990).

1.3- Production System

The New Zealand dairy industry is currently based on more than two million dairy cows in about 15000 herds, with majority of cows producing 130 to 150 kg of milk-fat per year. The majority of cows calve in springtime (July to September) (Holmes and Wilson, 1987) and the majority of milk (more than 90%) is manufactured into dairy products (Dairy Statistics, 1988-89). About three-quarters of all herds are managed by the owner, with most of the rest being farmed under a share-milking agreement (Holmes and Wilson, 1987).

Share-milking, the practice where someone other than the owner milks the cows

for a predetermined share of returns from the property, plays an important part in the New Zealand dairy industry (Newell, 1973).

The average herd size in 1988-89 was 157 cows. However, herds ranged in size from ten cows to more than 1000 cows. The number of herds with more than 300 cows was 644 (or 5% of total herds). This compares with 1.5% of herds in 1980-81 and 0.4% in 1970-71 (Dairy Statistics 1988-89).

In New Zealand the seasons change regularly and are distinct with less sunshine and lower temperature in winter than in summer. Soil is cooler and wetter in winter than in summer. There is seasonal variation in pasture growth rate due to rainfall and temperature changes. Because of the climatic variation and little use of supplements other than hay or silage, dairying in New Zealand is highly dependent on weather conditions.

Two systems of dairy farming are practiced in New Zealand:

i- Town Supply Dairy Farms, in which the cows may calve in spring and in autumn, or throughout the year, and the farm supplies a specified amount of milk throughout the year (Holmes and Wilson, 1987). The farmer is paid per litre of milk supplied (provided that the milk satisfies certain minimum compositional standards, namely 3.25 percent milk-fat, 8.50 percent solids-non-fat). In 1988-89 eight percent of dairy herds (an estimate of 1,151 herds) supplied milk to town milk industry, for domestic liquid milk consumption (Dairy Statistics, 1988-89). Milk is used for consumption without significant processing.

ii- Seasonal or Factory Supply Dairy Farms, in which the cows calve in springtime, do not produce milk in winter and all milk is manufactured into dairy

products. The farmer receives payment based on quantity of milk-fat and protein with a penalty for milk volume (Holmes and Wilson, 1987). In 1988-89 ninety two percent of New Zealand dairy herds (an estimate of 13,593 herds) had their milk processed into dairy products by one of 22 factories owned by dairy company cooperatives (Statistics, 1988-89).

1.3.1- Calving Date:

One characteristic of seasonal supply dairy farming in New Zealand is the short calving period. The winter is an important time for maintenance in N.Z. dairy factories, and there is no milk collection for four to six weeks in mid winter. While all seasonal supply herds calve during the period June to October, individual herds vary greatly in timing and concentration of calving within this period (Simmonds, 1985).

A seasonal supply dairy farmer in New Zealand receives the same payment for each unit of milk-fat irrespective of whether it is produced in winter, spring, summer or autumn. Dairy companies pay on a complex formula involving milk solids and a volume deduction. On seasonal supply farms the cows should calve as early as possible in spring provided that they can be fed well from the day that they calve.

If calving is concentrated, but occurs before the spring growth of grass has begun, then the majority of the herd will be underfed in early lactation. It is essential, therefore, that all other aspects of management must be well organized in order to take full advantage of the potential benefits of a concentrated calving (Holmes and Wilson, 1987).

In most seasonal herds, a cow's lactation length is largely dictated by its calving date relative to the planned start of the concentrated calving date. This is because

most cows remaining in the herd for another production season will be dried off within a week of each other from two to three months before the planned date of calving. The availability of fresh pasture during late lactation and the volume of conserved pasture as hay or silage are the major factors influencing a herd owner's decision to stop milking remaining cows in milk at that time (MacMillan, 1985).

1.3.2- Stocking Rate:

Daily intake of fresh herbage and dry matter of the grazing animal varies with the species and live weight of the animal (Hafez, 1946).

The grazing time required per unit intake of pasture was appreciably greater in the Jersey than Friesian, whereas the feed intake required per kg of fat-corrected milk produced was greater in the Friesian than Jersey. When the fat corrected milk production of the two breeds was the same, differences between them in grazing times were negligible. The Friesian animals, however, tended to have a somewhat greater rumination time (Brumby, 1959).

The single most important factor controlling efficiency of grazing is stocking rate (Bryant and Holmes, 1985). Stocking rate is conventionally measured in cows per ha; this implies that "cows" and "ha" are both unvarying items which is not the case (Holmes, 1984). Differences between size and breed of cows and also differences in productivity of lands will affect the stocking rate. Because of high dependency of New Zealand dairy farming on pasture, it is important to use pasture as efficiently as possible. This can only be achieved with stocking rate close to the optimum. Although stocking rate can influence pasture growth, this effect is variable and it probably depends on other factors, for example, the actual intensity of grazing at different times of the year (Holmes, 1984). The economic importance of stocking rates

is centered on the raising of existing farm income and profitability by increasing the number of animals per unit of land and on the efficiency of pasture utilization.

Low stocking rate can cause a decrease in pasture growth because of the presence of increased amount of reproductive and older plant material which causes death and decay of the lower leaves and suppression of legume species. On the other hand, a very high stocking rate will also cause a decrease in pasture growth because of excessively intense grazing or over-grazing (Sung Ho, 1986). Stockdale and King (1980) found that at a stocking rate of 8.6 cows/ha pasture grew at 62 kgDM/ha daily in comparison with 73 kgDM/ha daily at a stocking rate of 4.4 cows/ha. In another experiment the mean daily intake of herbage DM was equivalent to 13.4 kg/cow at 3 cows/ha and 10.4 kg/cow at 4.9 cows/ha (Freer, 1960). These figures show that increasing stocking rate more than a specific level would cause a decrease in the amount of pasture eaten by each cow.

Stocking rates change between farms mainly because of variability in soil types, number of dry cattle carried or in the pasture conservation and fertilizer policy (Holmes, 1984).

1.4- Genetic Improvement:

A dairy cattle population can be genetically changed along any one or more of four separate pathways. Desirable genes are identified and then passed along one of these pathways to benefit the next generation. The four pathways are:

1. Male parents to male progeny.
2. Female parents to male progeny.
3. Male parents to female progeny.
4. Female parents to female progeny.

The idea in developing a plan for genetic improvement is to use all four pathways in the best possible combination. The best combination is going to be the one which provides the largest net income per cow under commercial dairying conditions (Wickham et.al., 1978). Searle (1961) and Evans (1969) found that the male to female and male to male pathways can make the greatest contribution to rates of genetic gain in New Zealand dairy cattle. The superior contribution of these male pathways rely on using AI to increase selection intensity relative to natural mating. This is being successfully achieved in New Zealand partly as a result of AI using fresh semen which can be further diluted than when semen is frozen. Also using the best females for contract matings to produce bull calves for the progeny test program utilises the female to male pathway to increase genetic gain. With advances in applying multiple ovulation and embryo transfer and using cloning and sexing embryos, the future may allow better use of the female to female pathway. However, a relatively long generation interval due to time required for progeny testing, reduces the amount of genetic progress per year and research is being done to overcome this problem.

An estimate of the genetic merit of New Zealand dairy cows is provided by the Breeding Index (BI). The Breeding Index represents the genetic merit relative to a base value of 100, the average genetic merit of dairy cows across the country in 1960 to 1964. Comparing the average BIs of New Zealand dairy cows by birth year shows a constant increase in BI.

Improvements in BI have been recorded in commercial herds which have consistently used the Premier Sire Service provided by the NZ Dairy Board in collaboration with Livestock Improvement Associations to produce all their heifer replacements (Macmillan, 1982). Premier Sire Service is a service designed to make maximum use of the Corporation's best bulls and to provide farmers the maximum

gain at the least cost (Dairy Statistics, 1988-89). A comparison of BIs of bulls from 1975 through 1990, shows linear increase in BI (Wickham, 1989).

Genetic merit of New Zealand dairy cows compete with the animals of the same breed from other parts of the world. Although production performance of dairy cows in New Zealand is not as high as cows of some other countries; running them under similar nutritional environment to overseas dairy cattle allows a comparison of the genetic merit of N.Z. and foreign dairy cows. Results of a comprehensive comparison of different 'strains' of cattle from different countries obtained from FAO-Polish trial with 'black and white' cattle stated in Jasiorowski et. al. (1983) shows the position of Holstein-Friesian of New Zealand in the world. Records of progeny from young sires from N.Z. were lighter than average of all countries but produced more milk than average. Their fat percentage was highest and equal to those from Netherlands. Protein percentage of milk of daughters of bulls from New Zealand was second, with a very small difference, after bulls from USA. Their total butter fat and protein yield per lactation and per 100 kg of live weight was the highest among all other strains.

1.4.1- Data Collection:

Several organizations and individuals are involved in breeding dairy cattle in New Zealand including private breeding companies and private herd owners. The six Livestock Improvement Associations in addition to their artificial insemination services, provide the official herd recording services for dairy farmers and data collected by the herd testing service from sire proving herds are used in the progeny testing of bulls (Holmes and Wilson, 1987). The information from herd testing can also be used by the owner of the cows to assist in the management of their herds and in the selection of cows for the breeding of replacement heifers.

1.4.2- 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. 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 1939 different herd testing associations were amalgamated into six regional associations (NZDB, 1987-88).

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, milk-fat, and somatic cell counts. Sampling interval can be 4, 6 or 8 weeks or 2 or 3 times per year. Over 1 million cows, or 50% of the national herd, are tested. This testing forms the basis of production recording, progeny testing and some herd management decisions (Coop, 1987).

Presently record services are provided by Livestock Improvement Corporation which enables farmers to establish and maintain complete herd records.

1.4.3- Sire Evaluation:

The objective of sire evaluation is to allow comparisons of the genetic merit of sires. Sire proofs are based on a comparison of the production of a bull's daughters with that of other cows (herdmates) of the same age and by the same breed of sire, being milked in the same herd at the same time. In addition, information on the bull's ancestry and the breeding value of the herdmates' sires is taken into account (NZDSSR, 1988-89). The result of sire evaluation is referred to as the Breeding Index

(BI) of sires. In New Zealand, BIs are calculated for fat yield (FAT BI), protein yield (PROTEIN BI), milk yield (MILK BI), and for a combination of these three BIs (PAYMENT BI). The reliability for each BI is also calculated.

The PAYMENT BI has been designed for the payment system "a+b-c" in N.Z. This system is applied because protein yield is valuable to the factory supply farmer along with fat yield, but large volume of milk will be penalized. This BI reflects the true value of the milk of proven sires' daughters to both farmer and the manufacturing sector (NZDAAR, 1988-89). Each BI compares a sire with the average progeny tested bull in the base year (1960).

The Sire Proving Scheme is a progeny testing program designed to provide a continuous supply of proven bulls of high genetic merit. Each year about 125 bulls are progeny tested from which five or six best bulls being kept and used in Premier Sires and Nominated Service (Dairy Statistics, 1988-89). Based on the information about the ancestors and through contract matings between the best bulls and cows in the country, the necessary young bulls for progeny testing are produced. Semen is collected from these young bulls when they are about 15 months of age and used on cows in "Sire Proving Scheme" herds. These are normal commercial herds where the owner has agreed to let his cows be mated to unproven bulls in return for certain financial considerations. The aim is to provide about 50 daughters from each young bull distributed at one daughter per herd. The cows for these matings are selected randomly to prevent any bias through mating of some bulls to genetically superior cows.

To calculate BI, the genetic value of the bull is first estimated from his ancestry. The average milk, fat and protein yields of the bull's daughters milking in the current season are compared with the average yields of herdmates. After computing BIs for

fat, protein and milk, these BIs are combined to calculate PAYMENT BI.

In addition to production traits, dairy sires in N.Z. are ranked based on about seventeen traits other than production traits. These traits are considered because of their indirect contribution to the farmer's income. Using two year old progeny of sires their breeding value for each of these traits are calculated. Applying respective economic values for each score of each trait, the Economic Breeding Value of each trait other than production is calculated. Finally by summing Economic Breeding Value of the traits other than production and PAYMENT BI, The " total breeding index" of each sire will be calculated. The total breeding index ranks all sires within a breed according to their genetic and economic merit for production, management, efficiency and conformation (NZSER, 1988-89).

1.4.4- Artificial Insemination:

Early investigations and experiments concerning AI in New Zealand started about 1939. In 1943-44 the experimental work to inseminate 1000 cows was done of which 80 percent proved in-calf after a service of about 3 months (Dalton and Rumble, 1985).

According to the 1988-89 Cow Census made by New Zealand Dairy Board (Dairy Statistics, 1988-89) the number of cows which used AI summed to 1,652,409 and the number of yearlings which used AI was 57,942. Milk-fat production per cow for farms using AI was 143.9 kg while it was 130.3 kg for farms which didn't use AI. Milk-fat production per hectare was 358.6 kg and 279.5 kg for AI users and non-AI users respectively.

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 utilization of bulls. Farmers have the option of inseminating cows themselves or having a AB technician 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.

CHAPTER TWO

CHAPTER TWO

SIRE EVALUATION

2.1- Sire Evaluation:

Breeders have recognized, long before Mendel's experiments, the importance of the transmitting ability of sires for the simple but important reason that males can have many more offspring than females (Pirchner, 1984). Progeny testing of bulls has been made possible by recording of cows' production and the increased numbers of progeny per bull by use of artificial insemination.

The main method for sire evaluation in dairy cattle is progeny testing. Under idealized conditions, twice the deviation of the daughter average from the general mean is very close to Breeding Value. Therefore, the daughter average should usually enable the estimatiom 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 the sire (Dempfle, 1984).

One way to determine an animal's genotype is to obtain several progeny from it, that is, progeny testing. Progeny testing can be applied to both male and female individuals, but because of limitation of the number of progeny a female animal can produce, progeny testing is usually only useful for the genetic evaluation of males.

This procedure makes it possible to evaluate an animal for traits which are not observable, for example, milk production in dairy bulls (Van Vleck et. al., 1987).

The principles of a progeny test come from the sampling nature of inheritance. Each offspring receives from the parent a sample half of that parent's genotype. Each additional offspring receives another independent sample from the same source. If one can find out what was in several such samples, he will be fairly sure of what was in the parent (Lush, 1984). As the phenotype of an animal is not completely due to its genotype, in practice, there are some difficulties in using a progeny test because we don't know exactly what genes the progeny have. Another problem is that, each offspring receives half of its genes from each parent and we can not determine which part of its genotype is from the parent under test.

Overcoming these difficulties for qualitative traits can be easier than in quantitative traits, provided small numbers of genes control the expression of each qualitative trait. For quantitative traits, however, this is not the case. Each quantitative trait is under the control of a group of genes with different modes of expression. The genetic portion of a quantitative trait is affected by dominance, epistatic and additive effects. For dominance and epistatic effects the genotype of individuals affect the phenotype; and animals don't pass their genotype to their progeny, rather they pass a sample half of their genes. Averaging the performance of the progeny of a sire for a quantitative trait provides an estimation of the additive genetic value of that sire for the trait. Calculating this average needs specific procedures which will be discussed in succeeding sections. Although the progeny test is also useful to test for deleterious or other simply inherited traits such as red factor for coat colour, the main objective here is its application to quantitative traits.

It is possible to reduce errors in a progeny test caused by sampling variation of inheritance by increasing the number of progeny. Also the higher number of progeny will reduce the error coming from environmental conditions. Despite these advantages of increased progeny number, it has some disadvantages, such as, increasing the cost of progeny test and increasing the generation interval (Lush, 1984; Falconer, 1981). However, because of the time necessary to observe the results of a progeny test, it increases the generation interval.

Twice the average deviation of one animal's offspring from the population mean is called the Breeding Value (BV) of that animal assuming it was randomly mated with other individuals of the population. One step in estimation of the Breeding Value is to correct for the factors other than additive genes which influence the performance of progeny. Many factors influence the appearance of progeny. Some of these factors are visible and others are not. Some factors which affect the performance of dairy progeny are:

* Breeding value of the parents.

* Deviation of the breeding value of the progeny from mid-parent average (the effect of mendelian sampling).

* Previous culling policy.

* Age at calving.

* Number of days in milking.

* Number of times milked per day.

* Production system.

* Herd management.

* Nutrition during lactation.

In the model,

$$p_i = \mu + g_i + e_i$$

The μ term can be thought of as representing major and identifiable non-random environmental effects such as herd management level, the age of the animal when the record was made, the year, and the season of the year when the record was made (Van Vleck et. al., 1987). In that sense, μ might be different for each animal and each record of each animal. The model for corrected record is

$$x_i = p_i - \mu = g_i + e_i$$

Henderson (1973) has described development of methods of estimating BVs. Currently the method of choice for estimating BV is Best Linear Unbiased Prediction (BLUP). In this review, different aspects of BLUP will be discussed.

2.2-Best Linear Unbiased Prediction

The basic assumption in using performance records to calculate a selection index is that the records have been adjusted previously for all known sources of non-genetic bias. When there is no previous estimation of adjustment factors, these must be estimated using data under evaluation. For example, in ranking dairy sires using production records of animals of 3,000 herds obtained over a period of three years, there are 36,000 different Herd-Year-Seasons (HYS) (3,000 herds x 3 years x 4 seasons). Since each of these herd, year and season combinations usually represents a unique environmental effect, adjustment factors are required for 36,000 different HYSs. Because of uniqueness of each HYS the adjustment factor for each HYS must be estimated only from the data that is to be adjusted. Correct ranking of bulls requires applying proper adjustment factors to each record and calculating the estimated

breeding value (EBV) of each sire using $I = bC$, where C is the average adjusted records of all progeny of a given sire, and b is the appropriate index weight.

Best Linear Unbiased Prediction (BLUP) (Henderson, 1972) is a set of equations for predicting random effects without assuming parameter values of the fixed effects to be known (Garrick, 1991). Estimated Breeding Values calculated using the BLUP method have all the desirable properties of EBVs obtained from a conventional selection index. However, the BLUP method is more powerful than the selection index method. For example, the BLUP method can be used to compare estimates of the average breeding value of groups of animals born in different years, therefore providing an estimate of genetic trend. Also, BLUP can account for factors such as assortative mating, sires coming from different populations, differences between herds in average breeding value of cows and environmental trend over time.

BLUP is used to obtain estimates of the breeding values of various animals, taking account of fixed effects such as herd-year-season.

Consider the following equation of a mixed linear model:

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

Where:

$\mathbf{y}$ is a vector of observations.

β is a vector that contains fixed effects of various factors such as contemporary groups.

$\mathbf{X}$ is the incidence matrix associating effects in β to $\mathbf{y}$.

$\mathbf{u}$ is a vector of random genetic values.

Z is the incidence matrix associating effects in **u** to **y**.

and,

e is a vector of random residuals.

The expectation of **y**, **u**, and **e** are assumed to be:

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

When selection exists, then it is unlikely that E[**u**] = [**0**] because animals are progeny of selected parents which are assumed to have a mean better than the mean of the population.

The properties of best linear unbiased prediction as shown by Henderson (1975c) include unbiased prediction of **u** even when selection has occurred provided that records on which selection was based are included in the analysis and that all relevant genetic relationships are specified correctly.

It is commonly, although not necessary, assumed that there is no covariance between genetic and residual effects:

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

In the single trait example it is often assumed that:

$$\mathbf{G} = \mathbf{A}\,\sigma_g^2$$

Where:

A is the numerator relationship matrix, and σ_g^2 is the additive genetic variance.

This genetic variance structure assumes:

1) Only additive genetic effects are important.

2) Genetic values are all from the same distribution and have a common genetic variance.

3) Genetic differences are the same when expressed in the presence of any fixed factors (e.g. HYS) i.e. no genotype-environment interaction.

Again, in the simplest model, it is assumed that:

$$\mathbf{R} = \mathbf{I}\,\sigma_e^2$$

Where:

I is the identity matrix of order the number of records, and

σ_e^2 is the residual variance.

Residual variance structure assumes:

1) There are no covariances among the residual effects.

2) Residual variance is not influenced by genetic value, and

3) Residual variances are expressed to the same extent for all fixed or management factors.

In this example, a form of the mixed model equations for the general linear model are:

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

Where $\alpha = \sigma_e^2/\sigma_g^2$, is ratio of residual variance to additive genetic variace.

Henderson (1963) proved that BLUP(**u**) can be obtained from solving these mixed model equations.

2.2.1- Contemporary Groups

Contemporary groups are used to remove biases from genetic evaluations due to differential effects such as management associated with the grouping. Numerous groups, however, can result in small numbers of records per subclass with associated loss of effective progeny number and increased prediction error variance (Van Vleck, 1987). Contemporary groups can be defined as herd-year-season, lactation number, genetic groups, registered or non-registered, treated by hormones or not treated, or various combinations of these groups, etc.

The assignment of calendar day of freshening to a seasonal group is arbitrary. If a season is defined as a short period of time, e.g., a single calendar day or week, another problem arises due to the limited number of animals of the same lactation freshening on a particular calendar day or week. If no other animal freshens in that season, then that record has no contemporary records available for comparison. Thus, bias due to failure to adjust properly for seasonal conditions must be balanced against the increase in prediction error variance due to the loss of effective number of progeny (Van Vleck, 1987).

2.2.2- Effective Progeny Number

The ordinary least square (OLS) equation is:

$$\mathbf{X'Xb = X'y}$$

In sire evaluation the coefficient matrix is partitioned to $\mathbf{X} = [\mathbf{X}\ \mathbf{Z}]$, in which $\mathbf{X}$ is coefficient matrix for fixed effects and $\mathbf{Z}$ is coefficient matrix random effects. In this case with the conformable partitioning of **b** and **y** the OLS equations become:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

The diagonal coefficient of the least squares matrix corresponding to a sire after absorption of fixed effects ($\mathbf{Z'Z-Z'X(X'X)^{-}X'Z}$) is defined as the effective progeny number. The effective number of progeny of a sire can be affected by number of contemporaries of that sire's progeny. When a bull has n daughters in a contemporary group along with m daughters of other bulls, the contribution of the daughters of that bull to the diagonal of the bull equation after absorption of the contemporary equation is equal to (Van Vleck, 1987):

$$n^{*} = n - n^{2}/(n+m) = nm/(n+m)$$

The equation shows that regardless of the number of progeny of a sire, the effective progeny number is dependent on number of progeny of other sires in the corresponding

group. When all records in one contemporary group belong to the daughters of one sire, e.g. m is equal to zero, that contemporary group adds nothing to effective progeny number of the proposed sire. Increasing the number of contemporaries with different sires, increases the effective progeny number.

In a single trait example, the diagonal of the sire equation prior to absorption of any effects is the number of total lactation records plus the variance ratio. During absorption, this element is modified to reflect the number of records per sire and distribution of records across contemporary groups.

2.2.3- Prediction Error Variance

Sire evaluations are generally published with a repeatability or a confidence interval. Variance of prediction error of sire evaluations is required for computing repeatabilities or confidence intervals.

Prediction error variance (PEV) is the variance of differences between estimated breeding value ($\hat{\mathbf{u}}$) and true breeding value ($\mathbf{u}$) and is equal to:

$$\text{Var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{G} - \text{Var}(\hat{\mathbf{u}})$$

Where, $\mathbf{G} = \text{Var}(\mathbf{u})$ (Henderson, 1972).

The most commonly used estimate of PEV is the reciprocal of the diagonal element of the coefficient matrix after absorption of fixed effects (Greenhalgh and Quaas, 1986). Prediction error variances are usually difficult to calculate directly, requiring inversion of a large matrix (Ufford et. al., 1979). Therefore, the variance of prediction error is estimated. In some cases where the number of animals to be evaluated is small

the inverse can be obtained relatively easily (Szkotnicki et. al., 1978). In cases with large numbers of sires to be evaluated, it becomes desirable to determine a procedure which will give a close approximation with much less computation cost.

Because off-diagonal elements also affect the inverse, and also are generated by absorption, the inverse of the effective number of progeny can be considered only as an approximation to prediction error variance (Van Vleck, 1987). Ufford et. al. (1979) suggested the diagonal element of each sire equation in the coefficient matrix after absorption of fixed effects as a likely candidate for a good predictor of error variance.

After developing a simple method to compute the inverse of the relationship matrix from a list of parents and progeny (Henderson, 1976 and Quaas, 1976) and also after developing models to, simultaneously, evaluate sires for several correlated traits and therefore increasing the number of off-diagonal elements in the coefficient matrix, the reciprocal of the diagonal of the coefficient matrix is no longer suitable for estimating prediction error variance.

Weller et.al. (1985) analysed three sire evaluation models. Reciprocal of the diagonal elements was an accurate estimation of variance of prediction error only for the multiparity model for single traits without relationships. For the multiparity model for a single trait with relationships included, the variance of prediction error was estimated best by a function with four terms: the reciprocal of the diagonal element minus relationship contributions, the square and cube of the term, and the reciprocal of the diagonal element of the matrix of sire variance. For the multitrait model, all functions investigated resulted in less accuracy than desired for estimating variance of prediction error. However, Weller et.al.(1985) suggested that the function they developed could not be considered generally applicable, because the data sets they

used were relatively small and few bulls had large numbers of progeny and therefore low PEV.

2.2.4- Using Relationships Between Bulls

Some increase in accuracy of evaluation can be effected by taking into account the relationships among sires (Henderson, 1972). When two sires are related, progeny of one sire provide additional clues to the EBV of other sire and vice versa. The inverse of the relationship matrix is used in sire evaluation. When this matrix is large, inverting it using conventional methods is costly and impractical.

Henderson(1975a and 1975b) and Quaas (1976) presented a simple method of calculating the inverse of the relationship matrix which is designated as A^{-1}. The use of A^{-1} increases the accuracy of sire evaluation, particularly for young sires with few daughters. A^{-1} creates more ties between sires, which is equivalent to each sire having more daughters (Everett and Jones, 1985).

Potential advantages in using relationships among sires and, in addition, some female ancestors are that it:

1) Increases accuracy of prediction, particularly for sires with few or no progeny.

2) Requires fewer groups to account for genetic trend and for genetic differences among populations.

3) Evaluates sires earlier through use of records on dam, paternal sisters of dam, and their own paternal sisters (Henderson, 1975b).

4) Reduces the variance of prediction error (Kennedy and Moxley, 1975).

2.2.5- Reliability

Sire evaluations generally are published with a Repeatability (USA) or Reliability (Europe, NZ) or a confidence interval. The Reliability of estimates of breeding values is often expressed as the squared correlation between estimated and true breeding values $(r_{(\hat{u},u)})^2$. In best linear unbiased prediction procedures, which account for relationships between sires, reliability may be approximated by $(n_e + ka^{-1} - k)/(n_e + ka^{-1})$, where n_e is the effective progeny number, i.e. the diagonal of the sire coefficient matrix after absorption of fixed effects; k is the ratio of residual and sire variance; and a^{-1} is the diagonal element of the inverse of the numerator relationship matrix ($\mathbf{A}^{-1}$).

To calculate the true reliability of the estimate of breeding value, consider the following model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{e}$$

Where:

$\mathbf{y}$ = vector of observations,

β = vector of fixed effects,

$\mathbf{u}$ = vector of random effects of sires,

$\mathbf{e}$ = vector of random residuals,

$\mathbf{X,Z}$ = known incidence matrices.

Let $E(\mathbf{u})=\mathbf{0}$, $E(\mathbf{e})=\mathbf{0}$, $Var(\mathbf{u})=\mathbf{A}\sigma_g^2$, $Var(\mathbf{e})=\mathbf{I}\sigma_e^2$, where $\mathbf{A}$ is a matrix of additive genetic relationships among sires, $\mathbf{I}$ the identity matrix, and σ_g^2, σ_e^2 are genetic and residual variances, respectively.

Solutions for β and $\mathbf{u}$ could be (Henderson, 1972):

$$\begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

Where:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\alpha\mathbf{A}^{-1} \end{bmatrix}^{-}$$

a generalized inverse of the coefficient matrix, and,

$\alpha = \sigma_e^2/\sigma_g^2$,

$\mathbf{C_{22}} = [\mathbf{Z'Z} + \alpha\mathbf{A}^{-1} - (\mathbf{Z'X}(\mathbf{X'X})^{-}\mathbf{X'Z}]^{-1}$.

Let c_{ii} represent the i^{th} diagonal element of $\mathbf{C_{22}}$, then:

$$Var(\hat{\mathbf{u}}_i) = \sigma_g^2 - c_{ii}\sigma_e^2$$

$$Cov(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}) = Var(\hat{\mathbf{u}}_i)$$

True reliability $(r_{(\hat{u},u)}(T))^2$ of the estimate of breeding value for sire i equals:

$$(r_{(\hat{u},u)}(T))^2 = \frac{[Cov(\hat{\mathbf{u}}_i, \mathbf{u}_i)]^2}{Var(\hat{\mathbf{u}}_i).Var(\mathbf{u}_i)} = 1 - c_{ii}.k$$

This formula holds only if $\hat{\mathbf{u}}$ is the estimated breeding value.

One way to approximate the reliability for sire i is (Wilmink and Dommerholt, 1985):

$$(r_{(\hat{u},u)}(T))^2 = 1 - k/N_i = (N_i - k)/N_i$$

Where:

$$N_i = n_{e_i} + ka^{-1}_{ii}$$

n_{e_i} = effective number of progeny of sire i, found as the diagonal element for sire i in $\mathbf{Z'Z - Z'X(X'X)^{-}X'Z}$, and,

a^{-1}_{ii} = i_{th} diagonal element of the inverse of the numerator relationship matrix **A**.

Wilmink and Dommerholt (1985) concluded that approximation of reliability only from effective number of progeny and heritability can be considerably biased and the bias increases with increasing deviation of sire coefficient matrix from diagonality after absorption of fixed effects.

The more the sire coefficient matrix after absorption of fixed effects equations and addition of $\alpha\mathbf{A}^{-1}$ is diagonally dominant, the closer $1/N_i$ and c_{ii} are. For a given number of effective progeny, the variance of the off-diagonal element in $\mathbf{Z'Z - Z'X(X'X)^{-}X'Z}$ depends on the number of sires of contemporary groups, because for any row or column, elements sum to zero. Increasing the number of sires of contemporary groups, reduces the variance of off-diagonal elements and increases the number of nonzero elements. Consequently, matrix C_{22}^{-1} will be more diagonally dominant because $\alpha\mathbf{A}^{-1}$ remains unaltered.

Wilmink and Dommerholt (1985) in simulated data compared approximated reliability with true reliability and found that approximated reliability was biased upward for sire evaluation based on many effective progeny and few direct sire comparisons or few effective progeny and relationship to its sire. They suggested an alternative approximation procedure using the effective progeny number for a bull and his sire which reduces the bias and is as follows:

$$(r_{(\hat{u},u)(SA)})^2 = \frac{0.75\ m_e.n_e + k.n_e + 0.25\ .k.m_e}{0.75\ m_e.n_e + k.m_e + k.n_e + k^2}$$

Where:

$(r_{(\hat{u},u)}(SA))^2$ = reliability calculated by selection index approach.
m_e = effective number of progeny of the bull's sire, and
n_e = effective number of progeny of the bull.

2.2.6- Accounting for Genetic Groups in the Model

Genetic groups are used to represent populations of sires for which the mean breeding value may differ from population to population (Famula et.al., 1983). A population, here, may be defined as all individuals born in the same year or geographic area or all individuals from the same breed.

Referring to the question of whether a relationship matrix should be applied to the models including sire groups or not, Henderson (1975b) only stated that fewer groups are necessary. Swake and Bruns (1986) from their simulation study concluded that models considering relationships among sires seem to be superior to models which include grouping of sires. Kennedy and Moxley (1975) compared sire evaluations

with including groups to the method using relationships ignoring groups, and suggested that although, in general, accounting for relationships appears preferable to grouping, some grouping applied to unrelated sires may be desirable. Accounting for additive genetic relationships between sires in the sire evaluation procedure can lessen the need for grouping, and theoretically, if all possible relationships back to the unselected base population are accounted for, no grouping is necessary (Kennedy, 1981). Pollak and Quaas (1983) showed that the estimated group effect was dependent on the kind of relationship matrix used.

Of course, if there is a true difference between genetic groups, sire evaluations from an operational model that ignored groups are biased. However, when sires of two breeds are jointly evaluated using their purebred and crossbred progeny, grouping may be applied because there is no recent relationship between sires of two breeds.

Quaas and Pollak (1981) presented a modification of mixed model equations as follows:

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

Where:

$\mathbf{h}$ = fixed herd-year-season effects.

$\mathbf{X}$ = incidence matrix relating HYS to observations.

$\mathbf{s}$ = random vector of sire effects.

$\mathbf{Z}$ = incidence matrix relating sires to daughters' records.

$\mathbf{g}$ = vector of fixed group effects.

$\mathbf{ZQ}$ = incidence matrix relating group of sire to daughters' records.

$\mathbf{e}$ = a vector of random residuals.

The **s** and **e** have null means and

$$\text{Var}\begin{bmatrix}\mathbf{s}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}/\alpha & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\sigma_e^2$$

Where α is the ratio of residual variance to sire variance, σ_e^2 is residual variance, and **A** is the matrix of numerator relationships among sires. In some applications **A=I**, or some relationships are missing because of incomplete pedigree information; this is a main reason for groups in the model. The number of rows in matrix **Q** is equal to number of sires and the number of columns is equal to number of groups. Each row, say the i^{th} has a single nonzero element, one, in the column corresponding to the group of the i^{th} sire.

After absorbing **h** equations, the mixed model equations become:

$$\begin{bmatrix}\mathbf{Q'Z'MZQ} & \mathbf{Q'Z'MZ}\\ \mathbf{Z'MZQ} & \mathbf{Z'MZ}+\alpha\mathbf{A}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{g}}\\ \hat{\mathbf{s}}\end{bmatrix} = \begin{bmatrix}\mathbf{Q'Z'My}\\ \mathbf{Z'My}\end{bmatrix}$$

Where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$.

To obtain the equations that can be solved directly to obtain a vector of proofs, say $\hat{\mathbf{u}}$ where $\hat{\mathbf{u}}=\mathbf{Qg}+\mathbf{s}$, define the matrix **P** as:

$$\mathbf{P} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{Q} & \mathbf{I}\end{bmatrix}$$

Where:

$$\mathbf{P}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{Q} & \mathbf{I} \end{bmatrix}$$

and premultiplying both sides of absorbed mixed model equations by $(\mathbf{P}^{-1})'$ and inserting $\mathbf{P}^{-1}\mathbf{P}$ between coefficient matrix and solution vector gives:

$$\begin{bmatrix} \alpha\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} & -\alpha\mathbf{Q}'\mathbf{A}^{-1} \\ -\alpha\mathbf{A}^{-1}\mathbf{Q} & \mathbf{Z}'\mathbf{M}\mathbf{Z}+\alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{g}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{Z}'\mathbf{M}\mathbf{y} \end{bmatrix}$$

Which can be solved for $\hat{\mathbf{u}}$, but because the rank of the system of equations is the order of the system minus one (Quaas and Pollak, 1981), one constraint must be imposed before solving the equations.

2.3- Solving Linear Equations

Dairy sire evaluation (sire model) by Best Linear Unbiased Prediction ordinarily involves processing hundreds of thousands of records to construct mixed model equations of the order of tens of thousands of sires and other effects. The computing procedures have often been discussed and means of reducing the costs of computing have received attention. Solution of a large set of equations is usually by iteration techniques because they are too large to store in the memory of the computer for direct inversion.

Schaeffer and Kennedy (1986) showed how solutions can be obtained without setting up the system of equations. Their strategy, named indirect approach, performs

Gauss-Seidel (G-S) or Successive Over Relaxation (SOR) iteration while reading the data files rather than the matrix of coefficients.

Solving equations by iteration does not give an exact answer. The method of choice depends on the properties of the equations and how the method makes use of these properties. In some cases, the model and amount of data will influence the method of choice.

Linear equations with several unknowns can be represented in matrix form as $\mathbf{A}\,\mathbf{x} = \mathbf{y}$ where $\mathbf{x}$ is the vector of unknowns, $\mathbf{y}$ is a vector of known values and $\mathbf{A}$ is the matrix of coefficients. Providing $\mathbf{A}$ has an inverse the solution is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$, a situation demanding that $\mathbf{A}$ be square and that its rank equal its order (Searle, 1982). The more general case is when there are p equations in q unknowns.

$$\mathbf{A}_{(pxq)}\ \mathbf{x}_{(qx1)} = \mathbf{y}_{(px1)}$$

with the rank of $\mathbf{A}$ being $r_A = r$.

A linear set of equations can be solved when they are consistent (for further detail of consistency see for example Searle, 1982). Inconsistent equations can only be solved approximately. One approach is to find $\mathbf{x}$ such that $\mathbf{Ax}$ is as close to $\mathbf{y}$ as possible. This method is known as least squares and leads to Least Squares Equations:

$$\mathbf{X'X}\hat{\mathbf{b}} = \mathbf{X'y}$$

with $\mathbf{x}$ being the incidence matrix, and $\mathbf{b}$ and $\mathbf{y}$ the vectors of the unknowns parameters and data, respectively. Premultiplication of $\mathbf{X'y}$ by the inverse of the left-

hand side coefficient matrix (**X'X**) produces a solution. In general, this procedure is limited to problems with equation systems of a few thousands. Solving for **b** gives $\mathbf{b}^{\circ} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$. For larger systems, frequently arising in sire evaluation, a considerable amount of data is involved, and the number of multiplications required is also very large. Rounding errors therefore have a considerable effect. In these situations, iterative methods for approximating the solution of a large sysetm are useful. The traditional method is Gauss-Seidel (G-S) iteration (Van Vleck and Dwyer, 1985a), however, other iterative methods also can be used to solve large systems of equations (for further detail see for example Gastinel, 1970).

Mixed Model Equations:

The mixed model set of equations is :

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

with **b** and **u** being solution vectors for the fixed and random parts, respectively. The difference from the Generalized Least Square (GLS) set of equations is the inclusion of $\mathbf{G}^{-1}$, the inverse of the covariance matrix of **u**. When there is no relationship between animals, $\mathbf{G}^{-1}$ is diagonal with the additive genetic variance as elements. $\mathbf{G}^{-1}$ can be added after the complete system of equations is set up. Thus, all effects (random or fixed) can be treated alike. Therefore, it is convenient to use only one incidence matrix **W** as [**XZ**] (Groeneveld and Kovak, 1990). In the first step, the coefficient matrix $\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}$ and the RHS are set up from the data. In mixed models, covariance matrices have to be added in step 2 for all random effects (Groeneveld and Kovak, 1990).

Let the mixed model equations be represented as:

$$\mathbf{A}\,\mathbf{b} = \mathbf{r}$$

where **A** is a NxN matrix of coefficients, **b** is a vector of unknowns, and **r** is a vector of right-hand sides (RHS). For Gauss-Seidel iteration, the i^{th} solution in round n can be written:

$$\mathbf{b}_i^{n} = \mathbf{b}_i^{n-1} + \left(\mathbf{r}_i - \sum_{j=1}^{i-1} \mathbf{a}_{ij}\,\mathbf{b}_j^{n} - \mathbf{a}_{ii}\,\mathbf{b}_i^{n-1} - \sum_{j=i+1}^{N} \mathbf{a}_{ij}\,\mathbf{b}_j^{n-1} \right)$$

Where N is the number of equations, b_i and r_i are elements of **b** and **r** respectively, a_{ii} and a_{ij} are elements of **A**, and n is iterate number (Van Vleck and Dwyer, 1985a; Misztal and Gianola, 1987).

A difficulty with iterative methods is determining when to stop and accept solutions (Hornbeck, 1975). Most iterative methods are terminated when an error estimate is small enough (Berger et.al., 1989). For ranking animals, a less precise stopping point may be required than when genetic evaluations are used to predict genetic trend (Van Vleck and Dwyer, 1985b). Blair and Pollak (1984) suggested that sufficiently accurate ranking of animals for selection purposes is achieved long before random effect solutions converge. They found 100 rounds to be enough for ranking sires using an animal model but more rounds of iteration to be needed to obtain an accurate indication of the genetic trend.

The nature of Mixed Model Equations as applied to animal breeding problems results in sparse matrices. The proportion of nonzero elements rarely exceeds 1% (Groeneveld and Kovac, 1990). One possible method is to use **IA, JA, A** storage scheme as a general device to store and access the nonzero elements of sparse

coefficient matrix. The vector **A** contains the nonzero coefficients, **JA** the column number of each nonzero element in the coefficient matrix, and **IA** contains the starting address of each row of the system of equations in **JA** and **A**. To utilize memory most efficiently, Gauss-Seidel iteration should be performed on half-stored coefficient based on the **IA, JA,** and **A** vectors.

Consider a full-stored system of equations as follows (Groeneveld and Kovac, 1990):

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ & a_{22} & a_{23} \\ \text{Symmetric} & & a_{33} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} RHS_1 \\ RHS_2 \\ RHS_3 \end{bmatrix} \qquad [1]$$

In half-stored format the lower diagonal of [1] would not be available. Rewriting [1] accordingly gives [2]:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} RHS_1 \\ RHS_2 - a_{21}b_1 \\ RHS_3 - a_{31}b_1 - a_{32}b_2 \end{bmatrix} \qquad [2]$$

Note that the zeros are not stored in sparse format.

When performing Gauss-Seidel iteration, the first equation is solved for the first unknown, the second for the second, and so on. Because the coefficient matrix of the MME is symmetric, a_{12} is equal to a_{21}. After b_1 has been calculated, each coefficient of the current row has to be multiplied by the current solution and subtracted from the next RHS elements, as indicated in [2]. Because the solutions change from one round

of iteration to the next, this process has to be repeated for each round. Therefore, a temporary work space has to be provided to store the adjusted RHS.

CHAPTER THREE

CHAPTER THREE

MATERIALS AND METHOD

3.1- Data Description

First lactation records of 72,480 dairy cows of Holstein-Friesian, Jersey and crossbreds of these two breeds from herds throughout New Zealand were obtained from the Livestock Improvement Corporation of New Zealand Dairy Board database. Records were from herds comprising both breeds and their crossbred for lactations in 1987 and 1988. Four traits, milk volume, milk fat, milk protein and days in milk were used in analysis. For all production records, sire, dam and maternal grand-sire information were available. Identification of animals producing the records along with calving date and breed of animals and their pedigree were given. Bulls included in the data as maternal grand-sires of cows producing the records did not have their own pedigree in the records, unless, they had at least one progeny having a record in the data set.

In total 4256 sires, 3571 maternal grand-sires, 1176 paternal grand-sires and 1701 maternal grand-sires of sires (paternal great-grand sires) were included in the data set. Of the total number of sires, 2161 were of Holstein breed and the remaining 2095 were of Jersey breed. 40212 cows were sired by Holstein breed and the other 32268 were sired by Jersey breed. 36620 cows were purebred Holstein (HH), 3592 crossbred sired by Holstein (HJ), 29847 purebred Jersey (JJ) and 2421 crossbred sired by Jersey (JH). 248 sires had at least 20 progeny and the number of sires having more than 50 progeny

was 108. 40 sires had at least 20 progeny from both Holstein and Jersey cows while the number of sires having more than 50 progeny from each breed was 27.

3.2- Model of Sire Evaluation

Two kinds of analysis were applied. In the first analysis sire evaluations were undertaken without inclusion of genetic groups. In second analysis sire evaluation using groups for breed effects was practiced. Including genetic groups was based on the assumption that the base populations of two breeds were different.

3.2.1- Sire Evaluation without Inclusion of Genetic Groups

In this analysis, for bulls of each breed, two separate evaluations were performed. At first, sires of each breed were evaluated only on their purebred progeny. Then sires were evaluated only on their crossbred progeny. In these two evaluations, sires possessing both purebred and crossbred progeny were used to compare the result of their evaluations by two data sets. Rank correlation was used to compare the position of each sire in the rankings using two methods. Records for cows with sires or maternal grand-sires with low progeny numbers were included for estimation of fixed effects but ancestral information was treated as unknown. The criteria used for including a bull in the vector of random effects was the sum of progeny plus 1/2 maternal grand progeny exceeding 5. However, to observe the effect of progeny number on the result of sire evaluations other lower limits of 20, 35 and 50 for this criteria were used and a summary of the results is shown in the appendix II. The model of analysis for this evaluation was assumed as follows:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{s} + \mathbf{Z}_2\mathbf{s} + \mathbf{e}$$

Where:

y is a vector of observations.

β is a vector of unknown fixed effects (herd-year-season).

s is a vector of random sire effects (breeding values of sires) and contains all bulls which are sufficiently represented as the sire and/or maternal grand-sire of the cows producing the records.

e is a vector of random residual effects.

$\mathbf{Z}_1$ is an incidence matrix relating elements of **s** as sires to **y**.

$\mathbf{Z}_2$ is an incidence matrix relating elements of **s** as maternal grand-sires to **y**.

X is an incidence matrix relating elements of β to **y**.

Some rows of $\mathbf{Z}_1$ and some rows of $\mathbf{Z}_2$ are all zero when pedigree information is not recorded or treated as unknown. Mostly there is one ½ in a row of $\mathbf{Z}_1$ and one ¼ in each row of $\mathbf{Z}_2$. Some columns of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are all zero, representing bulls which have no female progeny with records but are the sire or maternal grand-sire of other bulls with progeny records.

As matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are of the same order, these two matrices can be added together and then the model equation becomes:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{s} + \mathbf{e}$$

Where:

$$\mathbf{Z} = \mathbf{Z}_1 + \mathbf{Z}_2$$

In this case, every element of **Z** which relates a sire to **y** would be equal to ½ and every element of **Z** which relates a maternal grand-sire to **y** would be equal to ¼ and,

$$\text{Var}\begin{bmatrix} \mathbf{s} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

Where:

$\mathbf{G} = \quad \mathbf{A}\,\sigma^2_g$

$\mathbf{R} = \quad \text{diag}\{\sigma^2_e{}^*\}$, and,

$\mathbf{A}$ is numerator relationship matrix.

σ^2_g is additive genetic variance.

$\sigma^2_e{}^*$ is residual variance and has one of three values depending on whether sire and/or maternal grand-sire of the cow producing the record in known. Its derivation will be shown in the following section.

3.2.1.1- Calculating Residual Variance:

Depending on the existing information for the pedigree of each animal producing the record, different values for $\sigma^2_e{}^*$ as shown in the following were added to the appropriate positions depending on which ancestors were known.

First consider the following equation,

$$g_i = (1/2)\, g_s + (1/2)\, g_d + \phi$$

Where:

g_i is the total genetic effects present in the phenotype of each individual.

g_s is the genetic effects of sire of the animal.

g_d is the genetic effects of dam of the animal.

ϕ is Mendelian sampling effect resulting in meiosis with mean equal to zero and variance equal to 1/2 of genetic variance (σ^2_g) as proved in the following:

$$\begin{aligned}
Var(g_i) &= Var[(1/2)\, g_s + (1/2)\, g_d + \phi] \\
&= (1/4)\, Var(g_s) \\
&\quad + (1/4)\, Var(g_d) \\
&\quad + Var(\phi) \\
&\quad + (1/2)\,(1/2)\, 2\, Cov(g_s, g_s)
\end{aligned}$$

which is equal to 0 when sire and dam are not related.

$+ (1/2)\, 2\, Cov(g_s, \phi)$ which is equal to 0.

$+ (1/2)\, 2\, Cov(g_d, \phi)$ which is equal to 0.

Therefore, in populations with random mating and assuming sires and dams not related,

$$\sigma^2_g = (1/2)\sigma^2_{gs} + (1/2)\sigma^2_{gd} + \sigma^2_\phi$$

and assuming that both sire and dam genetic variances equal to σ^2_g, it can easily be seen that $\sigma^2_\phi = (1/2)\sigma^2_g$.

Now, we can define residual variances for different cases as follows:

Consider the real animal model,

$$y = \mu + g + e$$

Where:

y = observation on each animal (actual record).

μ = overall mean of the population and is common to all records.

g = total genetic effects present in the record.

e = total environmental effects present in the record, and,

$$\text{Var}(g) = \sigma^2_g \text{ and } \text{Var}(e) = \sigma^2_e$$

When none of the parents are known the residual variance is equal to:

$$\sigma^2_e{}^* = \sigma^2_g + \sigma^2_e$$

When only sire is known, we can write the model as the following:

$$y = \mu + (1/2)g_s + (1/2)g_d + \phi + e$$

Where:

$$\text{Var}[(1/2)g_s] = (1/4)\sigma^2_g$$
$$\text{Var}[(1/2)g_d] = (1/4)\sigma^2_g$$
$$\text{Var}(\phi) = (1/2)\sigma^2_g$$
$$\text{Var}(e) = \sigma^2_e$$

and all covariances are equal to zero, then, residual variance when only sire is known is equal to:

$$\sigma^2_e{}^* = \text{Var}[(1/2)g_d] + \text{Var}(\phi) + \text{Var}(e)$$

$$= \sigma^2_e + (3/4)\sigma^2_g$$

When both sire and maternal grand-sire are known,

$$g_d = (1/2)g_{mgs} + (1/2)g_{mgd} + \phi_d$$

where subscripts mgs and mgd denote maternal grand-sire and maternal grand-dam respectively, we can define the model as following:

$$y = \mu + (1/2)g_s + (1/4)g_{mgs} + (1/4)g_{mgd} + (1/2)\phi_d + \phi + e$$

therefore, residual variance in this case is equal to:

$$\sigma^2_e{}^* = \mathrm{Var}[(1/4)g_{mgd}] + \mathrm{Var}[(1/2)\phi_d] + \mathrm{Var}(\phi) + \mathrm{Var}(e)$$

$$= (1/16)\sigma^2_g + (1/4)(1/2)\sigma^2_g + (1/2)\sigma^2_g + \sigma^2_e$$

$$= (11/16)\sigma^2_g + \sigma^2_e$$

And finally, if only maternal grand-sire is known the residual variance is equal to:

$$\sigma^2_e{}^* = \mathrm{Var}[(1/2)g_s] + \mathrm{Var}[(1/4)g_{mgd}] + \mathrm{Var}[(1/2)\phi_d] + \mathrm{Var}(\phi) + \mathrm{Var}(e)$$

$$= (1/4)\sigma^2_g + (1/16)\sigma^2_g + (1/4)(1/2)\sigma^2_g + (1/2)\sigma^2_g + \sigma^2_e$$

$$= (15/16)\sigma^2_g + \sigma^2_e$$

3.2.1.2- Adding G^{-1} to the LHS Matrix:

A simple method for computing inverse of relationship matrix (Henderson, 1976 and Quaas, 1976) was used to construct elements of A^{-1}. Each element of A^{-1} was multiplied by the reciprocal of genetic variance to form corresponding element of G^{-1} and was added to appropriate positions. The simplified formula for positions of A^{-1} elements given in (Quaas *et. al.*, 1984) was applied. This formula for animal i, sire j and maternal grand-sire k is:

$$d_i = 1 - (1/4)a_{jj} - (1/16)a_{kk}$$

Where a_{jj} and a_{kk} are diagonal elements corresponding to sire i and maternal grand-sire k of animal i in the numerator relationship matrix and are equal to 1 when animals are not inbred. The following values were added to given positions in the LHS matrix:

	d_i^{-1}	to	a^{ii}
(-1/2)	d_i^{-1}	to	a^{ij} and a^{ji}
(-1/4)	d_i^{-1}	to	a^{ik} and a^{ki}
(1/4)	d_i^{-1}	to	a^{jj}
(1/8)	d_i^{-1}	to	a^{jk} and a^{kj}
(1/16)	d_i^{-1}	to	a^{kk}

Where a^{ij} refers to the element of A^{-1} in row i and column j.

A Fortran programme was written for absorbing HYS equations and adding G^{-1} to the LHS to form the following equations:

$$[Z'SZ + G^{-1}] [\hat{u}] = [Z'Sy]$$

Where:

$$S = R^{-1} - R^{-1}X (X'R^{-1}X)^{-1}X'R^{-1}$$

These equations were iteratively solved using Gauss-Seidel method. A flow diagram of the programme to explain the procedures of constructing and solving the equations is given in Appendix I.

3.2.2- Sire Evaluation with Inclusion of Genetic Groups.

In this analysis at first sires of each breed were evaluated using only their crossbred progeny and then all progeny of each sire, regardless of their breed, were taken into account. The model equation follows Quaas and Pollak (1981):

$$y = X\beta + ZQg + Zs + e$$

Where:

X, β, **Z**, **s** and **e** are as before in the model without inclusion of genetic groups. When sires of two breeds were jointly evaluated, breed of progeny was taken into account along with HYS to form contemporary groups.

ZQ is incidence matrix relating group of sires to daughters' records.

g is fixed group effect.

Construction of the linear equations and inverse relationship matrix was similar to when genetic groups were not included, with the exception that when adding $\mathbf{G}^{-1}$ to the absorbed equations, appropriate coefficients were added to genetic group equations. Because system of equations in this case are not full rank (Quaas and Pollak, 1981), one constraint was imposed to the group equations before iterative solution.

3.3- Correlation Between Sire Evaluations from different Methods.

The main objective of this study was to calculate correlation coefficients between results of sire evaluation from different methods. Ranking of sires of each breed in four methods of evaluation were compared to investigate how differently these methods rank the sires. To do this, a programme was written to separate common sires in each pair of evaluations. Sires were ranked based on their estimated breeding values in two analyses of each pair and Minitab statistical package was used to calculate correlation between ranks of the sires in two evaluation methods. For each trait ten different comparison was applied.

The pattern of comparisons was as follows:

type of comparison		
ALL	-	HH
ALL	-	HJ
ALL	-	JJ
ALL	-	JH
ALL	-	HJG
ALL	-	JHG
HH	-	HJ
JJ	-	JH
HH	-	HJG
JJ	-	JHG

ALL-HH represents comparison between evaluation of Holstein sires using all progeny with genetic groups included versus using only purebred progeny.

ALL-HJ represents comparison between evaluation of Holstein sires using all progeny with genetic groups included versus using only crossbred progeny without including genetic groups.

ALL-JJ represents comparison between evaluation of Jersey sires using all progeny with genetic groups included versus using only purebred progeny.

ALL-JH represents comparison between evaluation of Jersey sires using all progeny with genetic groups included versus using only crossbred progeny without

including genetic groups.

ALL-HJG represents comparison between evaluation of Holstein sires using all progeny with genetic groups included versus using only crossbred progeny with genetic groups included.

ALL-JHG represents comparison between evaluation of Jersey sires using all progeny with genetic groups included versus using only crossbred progeny with genetic groups included.

HH-HJ represents comparison between evaluation of Holsien sires using only purebred progeny versus using only crossbred progeny without including genetic groups.

JJ-JH represents comparison between evaluation of Jersey sires using only purebred progeny versus using only crossbred progeny without including genetic groups

HH-HJG represents comparison between evaluation of Holsien sires using only purebred progeny versus using only crossbred progeny with including genetic groups

JJ-JHG represents comparison between evaluation of Jersey sires using only purebred progeny versus using only crossbred progeny with including genetic groups

The expected correlation between EBVs of sires obtained from analysis of independent datasets was obtained using the following formula (Garrick, 1988):

$$r_{\hat{g}_1\hat{g}_2} = r_{g_1g_2} \cdot r_{g_1\hat{g}_1} \cdot r_{g_2\hat{g}_2}$$

Where:

$r_{\hat{g}_1 \hat{g}_2}$ = expected correlation between EBV of each sire in two independent analysis.

$r_{g_1 g_2}$ = correlation between true BVs of each sire in two methods .

$r_{g_1 \hat{g}_1}$ = correlation between true BV and its estimate in method one (square root of reliability of each sire in method one).

$r_{g_2 \hat{g}_2}$ = correlation between true BV and its estimate in method two (square root of reliability of each sire in method two).

The correlation between true BVs in independent data sets (r_{g1g2}) is assumed equal to unity if crossbred records are as useful as purebred records for evaluation of sires for purebred performance.

As the reliability of EBV of each sire in two methods is known and assuming the correlation between true breeding values of each sire in two methods to be equal to 1, it is easy to calculate the expected correlation between estimated breeding values of sires from two methods.

3.4- Calculating Effective Progeny Number, Variance of Prediction Error and Reliability.

Usually EBVs of sires are published with reliability of estimates, (r_{Π}^2), variance of prediction error (PEV) and effective number of progeny (n_e). Therefore these statistics were calculated for each sire.

Diagonal elements of the coefficient matrix after absorption of herd-year-season equations were adopted as effective number of progeny. In creating the equations, instead of writing the value of 1 for each incidence of each sire, the value of ½ was used to give the EBVs directly, and also the inverse of residual variance was multiplied to

each diagonal element. To remove the effect of these values on EPN, diagonal elements after absorption of HYSs were multiplied by 4 and the residual variance, respectively.

Variance of Prediction Error (PEV) was estimated as reciprocal of diagonal elements of coefficient matrix after absorption of HYS equations. To see how correct this estimation was, the exact values of diagonal elements of inverse of coefficient matrix were calculated using an iterative method. As calculation of each diagonal element needs several rounds of iteration and therefore is time consuming, this method was applied only for a few sires to compare the two methods. The logic of this method is illustrated as follows:

Suppose the objective is calculation of the 3rd diagonal element of the inverse of the matrix **A**. As the sum of products of all elements of row 3 of **A** by all elements of column 3 of its inverse matrix is equal to 1 (i.e. the 3rd diagonal element of identity matrix), the following equations in matrix form can be considered:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Solution for b_3 using iterative method is the 3rd diagonal element of the inverse of **A**.

Reliability of estimate of breeding value for each sire was approximated using the following formula:

$$r_{TI}^2(T) = (N_i - k)/N_i$$

Where:

$$k = \sigma_e^2/\sigma_s^2, \text{ and}$$
$$N_i = n_{e_i} + k a^{-1}_{ii}$$

n_{e_i} = effective number of progeny of sire i, found as the diagonal element for sire i in $\mathbf{Z'Z - Z'X(X'X)^{-}X'Z}$, and,

a^{-1}_{ii} = i_{th} diagonal element of the inverse of the numerator relationship matrix **A**.

This estimation of reliability does not account for off-diagonal elements of coefficient matrix after absorption and according to Wilmink and Dommerholt (1985) this may lead to bias in the estimates. True reliabilities were derived by calculating diagonal elements of inverse of LHS matrix using an iterative method.

The true reliability is equal to:

$$r_{TI}^2(T) = 1 - c_{ii}.k$$

Where c_{ii} is the i^{th} diagonal element of $[\mathbf{Z'Z - Z'X(X'X)^{-}X'Z} + k\mathbf{A}^{-1}]$ and k represents the ratio σ_e^2/σ_s^2.

In this study the coefficient matrix after absorption of HYS equations is equal to:

$$\mathbf{Z'R^{-1}Z - Z'R^{-1}X(X'R^{-1}X)^{-}X'R^{-1}Z + G^{-1}}$$

Assuming $\mathbf{R}^{-1} = \mathbf{I}(1/\sigma_e^2)$ and $\mathbf{G}^{-1} = \mathbf{I}(1/\sigma_s^2)$, therefore multiplying the above equations by σ_e^2/σ_e^2 gives:

$$(1/\sigma_e^2)[\mathbf{Z'Z} - \mathbf{Z'X(X'X)^{-}X'Z} + k\mathbf{A}^{-1}]$$

and its inverse is equal to:

$$\sigma_e^2[\mathbf{Z'Z} - \mathbf{Z'X(X'X)^{-}X'Z} + k\mathbf{A}^{-1}]^{-1}$$

Let d_{ii} represent diagonal elements of above matrix, then:

$$d_{ii} = \sigma_e^2 c_{ii}$$

Therefore,

$$r_{TI}^2 = 1 - (d_{ii}) \cdot (1/\sigma_e^2) \cdot k$$

$$= 1 - (d_{ii}) \cdot (1/\sigma_s^2)$$

which is the formula used in deriving true reliability of estimates of EBVs of sires in this study.

CHAPTER FOUR

CHAPTER FOUR

RESULTS

Means and standard deviations of different traits of each breed were calculated. As days in milking were different for different breeds, all record have been adjusted for this trait to calculate average production of each breed with the same length of lactation (250 days). Standard deviations for different breeds pooled within herd year season also are given.

Results of sire evaluations for four traits are presented separately in tables. For milk volume, five separate tables give information about: rank correlation of sire evaluations using different data sets, distribution of different statistics (Estimated Breeding Value, Effective Progeny Number, Prediction Error Variance and reliability of estimates), over-estimation of true reliability by estimation method, under-estimation of Prediction Error Variance by estimation method with different numbers of effective progeny, and correlation of true and estimated reliability (r_{TI}^2) as well as the correlation of true and estimated variance of prediction error (PEV) for different numbers of effective progeny (EPN). The expected correlation between sire evaluations using independent data sets are given where applicable. For milk fat, milk protein and days in milk, rank correlation of sire evaluations using different data sets as well as distribution of different statistics are given. Other information for these three traits are similar to those for milk volume.

Correlation of sire evaluations with different numbers of progeny also are given in Appendix II as extra information for those interested to see the effect of progeny number on these correlations.

4.1- General statistics for breeds:

Tables 4.1.1 through 4.1.3 represent average production (± standard deviation) of each breed before and after adjusting for lactation length as well as pooled standard deviation of each trait for different breeds respectively. The milk fat average of crossbreds is more than milk fat average for purebreds after adjusting for lactation length but the ratio of standard deviation to average in crossbreds is less than for purebreds. For milk volume, the variation in purebred Holstein is more than of those of other breeds and variation in purebred Jersey is less than those of other breeds. The two crossbreds demonstrated variation intermediate between the two purebreds. But the ratio of standard deviation to average production for all breeds are similar, indicating a relatively constant coefficient of variation.

Table 4.1.1: Mean and standard deviation of various traits for different breeds in the sire evaluation.

BREED[a]	N[b]	DAYS IN MILK	MILK VOLUME	MILK FAT	MILK PROTEIN	FAT (%)	PROTEIN (%)
HH	33906	248.74 ± 43.33	3143.61 ± 877.54	135.92 ± 37.67	104.76 ± 29.52	4.35 ± 0.49	3.34 ± 0.28
HJ	3449	238.11 ± 39.25	2793.40 ± 676.56	136.97 ± 34.54	100.17 ± 25.09	4.92 ± 0.52	3.59 ± 0.29
JJ	28623	247.92 ± 38.30	2334.82 ± 560.02	133.25 ± 34.27	94.14 ± 23.60	5.71 ± 0.56	4.03 ± 0.32
JH	2352	234.26 ± 40.07	2539.71 ± 631.93	132.02 ± 33.36	94.24 ± 23.65	5.23 ± 0.61	3.72 ± 0.34

a HH represents purebred Holstein cows.
HJ represents crossbred of Holstein and Jersey when sire is Holstein breed.
JJ represents purebred Jersey cows.
JH represents crossbred of Hostein and Jersey when sire is Jersey breed.

b Number of records used to calculate the statistics in the table.

Table 4.1.2: Mean and standard deviation of various traits for different breeds in the sire evaluation (adjusted to equal days in milk).

BREED	N	DAYS IN MILK	MILK VOLUME	MILK FAT	MILK PROTEIN	FAT (%)	PROTEIN (%)
HH	33906	250	3150.32 ± 647.59	136.14 ± 27.72	104.80 ± 21.22	4.35 ± 0.49	3.34 ± 0.28
HJ	3449	250	2938.38 ± 548.35	143.63 ± 26.74	105.06 ± 19.15	4.92 ± 0.52	3.59 ± 0.29
JJ	28623	250	2355.58 ± 446.42	133.99 ± 26.52	94.63 ± 17.61	5.71 ± 0.56	4.03 ± 0.32
JH	2352	250	2718.25 ± 538.31	140.84 ± 26.48	100.50 ± 18.34	5.23 ± 0.61	3.72 ± 0.34

Table 4.1.3: Standard deviation of various traits for different breeds (pooled within HYS).

BREED	DF	DAYS IN MILK	MILK VOLUME	MILK FAT	MILK PROTEIN
HH	23718	24.11	547.06	23.00	17.21
HJ	2017	20.57	436.24	21.98	15.56
JJ	22327	22.60	403.53	23.38	15.94
JH	1128	21.44	438.81	21.34	15.08

4.2- Milk Volume:

Results of sire evaluations for milk Volume are given in tables 4.2.1 to 4.2.5. Table 4.2.1 holds information about the distribution of Estimated Breeding Values (EBV) and Effective Numbers of Progeny (EPN) for sires as well as Prediction Error Variances (PEV) and Reliabilities (r_{TI}^2)of estimates. This information covers all sires having at least 5 progeny plus ½ grand progeny in the data set but not necessarily EPN = 5. More information about different statistics for sires having different EPN are also given in the tables.

The calculation of r_{TI}^2 requires inversion of the coefficient matrix which can be time consuming for big matrices (about 2000 by 2000). Approximate methods were used to estimate r_{TI}^2 in addition to calculation of nearly exact values by iteration. A comparison of these two r_{TI}^2 values was carried out for different EPN. True reliabilities were over-estimated and the distribution of over-estimation of r_{TI}^2 is given in table 4.2.2. As shown in the table, with increasing EPN the over-estimation of r_{TI}^2 decreases and the true r_{TI}^2 and its estimate get closer. However, the lowest limit of over-estimation was never less than 1%.

The same information as those for r_{TI}^2 are given for under-estimation of the PEV in table 4.2.3. Although not expected, under-estimation of PEV increases with increasing EPN, mainly because it is more likely sires having more progeny will be related and therefore the number of off-diagonal elements in coefficient matrix increases. This increase in the number of off-diagonal elements causes the reciprocal of diagonal elements in the coefficient matrix to be deviated from the diagonal elements of the

inverse coefficient matrix.

The correlation of r_{TI}^2 and its estimate as well as the correlation of PEV and its estimate for different number of effective progeny are given in table 4.2.4. This table shows that these statistics are highly correlated. This correlation slightly decreased for sires with EPN > 100 which can be due to smaller number of sires having more than 100 effective number of progeny.

Finally, the rank correlation of sire evaluations for different comparisons for milk volume are given in table 4.2.5. Where applicable, the expected correlation of the comparison are also given. The lowest correlation corresponds to the comparison of purebred records to crossbred records for Jersey sires (0.162) and the highest correlation corresponds to comparison of all records (purebred and crossbred records together) with genetic groups included to purebred records for Jersey sires (0.991). These correlations are for sires having more than 5 progeny plus ½ grand progeny in the data sets and the distribution of their progeny number is given in table 4.2.1.

Table 4.2.1: Distribution of different statistics in evaluation of sires (for milk volume).

	N[a]	MEAN	STDEV	SEMEAN	MINIMUM	MAXIMUM
EBV	2197	12.770	184.840	3.940	-997.000	1000.700
EPN	2197	22.340	103.010	2.200	0.000	1814.500
$\hat{\mathbf{PEV}}$	2197	123600.000	38651.000	825.000	1354.000	221541.000
$\hat{r}_{TI}^2$	2197	0.723	0.131	0.003	0.000	1.000
PEV	2197	128078.000	39656.000	846.000	1951.000	221541.000
r_{TI}^2	2197	0.422	0.179	0.004	0.000	0.990

a Number of sires used to calculate distributions.

EBV Estimated Breeding Value.

EPN Effective Progeny Number.

$\hat{\mathbf{PEV}}$ Estimated Prediction Error Variance.

$\hat{r}_{TI}^2$ Estimated Reliability for EBVs.

PEV Prediction Error Variance.

r_{TI}^2 Reliability of EBVs.

Table 4.2.2: Mean over-estimation(%) of r_{TI}^2 and other statistics with different number of EPN (for milk volume).

EPN	N	MEAN%	STDEV	SEMEAN	MINIMUM	MAXIMUM
n>0	2197	93.730	64.340	1.370	1.010	566.670
n>5	1239	72.880	38.730	1.100	1.010	176.000
n>10	457	39.450	24.780	1.160	1.010	92.860
n>15	278	24.460	17.610	1.060	1.010	67.310
n>20	216	17.934	13.615	0.926	1.010	45.161
n>25	177	13.342	10.148	0.763	1.010	37.313
n>30	160	11.216	8.105	0.641	1.010	30.986
n>35	149	9.986	6.927	0.567	1.010	27.027
n>40	140	9.092	6.117	0.517	1.010	23.377
n>45	134	8.557	5.685	0.491	1.010	20.253
n>50	128	8.023	5.234	0.463	1.010	20.000
n>100	81	4.629	2.618	0.291	1.010	10.112

Table 4.2.3: Mean under-estimation(%) of PEV and other statistics with different number EPN (for milk volume).

EPN	N	MEAN%	STDEV	SEMEAN	MINIMUM	MAXIMUM
n>0	2197	3.635	6.388	0.136	0.000	32.268
n>5	1239	4.008	5.905	0.168	0.067	32.268
n>10	457	5.622	6.116	0.286	0.235	32.268
n>15	278	7.086	6.278	0.377	0.411	32.268
n>20	216	7.258	6.091	0.414	0.586	32.268
n>25	177	7.625	6.134	0.461	0.711	32.268
n>30	160	7.814	6.202	0.490	0.711	32.268
n>35	149	7.909	6.277	0.514	0.711	32.268
n>40	140	8.185	6.310	0.533	1.131	32.268
n>45	134	8.384	6.335	0.547	1.256	32.268
n>50	128	8.440	6.425	0.568	1.256	32.268
n>100	81	10.320	7.130	0.792	2.431	32.268

Table 4.2.4:Correlation of true and estimated r_{TI}^2 and, true and estimated PEV with different EPN.

EPN	Correlation of $\hat{r}_{TI}^2$&r_{TI}^2	Correlation of $\hat{PEV}$&PEV
n>0	0.856	0.969
n>5	0.948	0.985
n>10	0.978	0.992
n>15	0.985	0.992
n>20	0.989	0.995
n>25	0.989	0.997
n>30	0.986	0.997
n>35	0.983	0.998
n>40	0.980	0.998
n>45	0.977	0.998
n>50	0.973	0.998
n>100	0.929	0.999

Table 4.2.5: Rank correlations of sire evaluations using different data sets (for milk volume).

Type of Comparison[1]	Number of Sires	Correlation	Expected[2] Correlation
ALL-HH	1092	0.972	*
ALL-HJ	86	0.658	*
ALL-JJ	1069	0.991	*
ALL-JH	82	0.277	*
ALL-HJG	86	0.667	*
ALL-JHG	82	0.277	*
HH-HJ	71	0.579	0.538
JJ-JH	41	0.162	0.546
HH-HJG	71	0.599	0.551
JJ-JHG	41	0.163	0.607

[1] Abbreviations in the table are as follows:

ALL-HH represents comparison between evaluation of Holstein sires using all progeny with genetic groups included versus using only purebred progeny.

ALL-HJ represents comparison between evaluation of Holstein sires using all progeny with genetic groups included versus using only crossbred progeny without including genetic groups.

ALL-JJ represents comparison between evaluation of Jersey sires using all progeny with genetic groups included versus using only purebred progeny.

ALL-JH represents comparison between evaluation of Jersey sires using all progeny with genetic groups included versus using only crossbred progeny without including genetic groups.

ALL-HJG represents comparison between evaluation of Holstein sires using all progeny with genetic groups included versus using only crossbred progeny with genetic groups included.

ALL-JHG represents comparison between evaluation of Jersey sires using all progeny with genetic groups included versus using only crossbred progeny with genetic groups included.

HH-HJ represents comparison between evaluation of Holstein sires using only purebred progeny versus using only crossbred progeny without including genetic groups.

JJ-JH represents comparison between evaluation of Jersey sires using only purebred progeny versus using only crossbred progeny without including genetic groups

HH-HJG represents comparison between evaluation of Holsien sires using only purebred progeny versus using only crossbred progeny with including genetic groups

JJ-JHG represents comparison between evaluation of Jersey sires using only purebred progeny versus using only crossbred progeny with including genetic groups

2 asterisks show that two data sets were not independent versus therefore the expected correlations were not calculated.

4.3- Milk Fat:

Results of sire evaluations for milk fat are given in tables 4.3.1 and 4.3.2. Table 4.3.2 holds information about distribution of EBVs and EPN for sires as well as PEVs and r_{TI}^2s of estimates. This information covers all sires having at least 5 progeny plus ½ grand progeny in the data set but not necessarily EPN = 5.

Rank correlations of sire evaluations for different comparisons for milk fat are given in table 4.3.1. Where applicable, the expected correlation of the comparison are also given. The lowest correlation corresponds to the comparison of purebred records to crossbred records with genetic groups included for Jersey sires (0.078) and the highest correlation corresponds to the comparison of all records (purebred and crossbred records together) with genetic groups included to purebred records for Jersey sires (0.993). These correlations are for sires having more than 5 progeny plus ½ grand progeny in the data sets and the distribution of their progeny number is given in table 4.3.2.

Table 4.3.1: Rank correlations of sire evaluations using different data sets (for milk fat).

Type of Comparison[1]	Number of Sires	Correlation	Expected[2] Correlation
ALL-HH	1092	0.985	*
ALL-HJ	86	0.570	*
ALL-JJ	1069	0.993	*
ALL-JH	82	0.098	*
ALL-HJG	86	0.588	*
ALL-JHG	82	0.094	*
HH-HJ	71	0.340	0.536
JJ-JH	41	0.110	0.546
HH-HJG	71	0.378	0.551
JJ-JHG	41	0.078	0.607

1, 2 see table 4.2.5

Table 4.3.2: Distribution of different statistics in evaluating sires (for milk fat).

EPN	N	MEAN	STDEV	SEMEAN	MINIMUM	MAXIMUM
EBV	2197	0.832	8.809	0.188	-35.600	44.500
EPN	2197	22.340	103.010	2.200	0.000	1814.490
$\hat{\text{PEV}}$	2197	278.220	87.000	1.860	3.000	498.700
$\hat{r}_{TI}^2$	2197	0.723	0.131	0.003	0.000	1.000
PEV	2197	288.300	89.270	1.900	4.400	498.700
r_{TI}^2	2197	0.422	0.179	0.004	0.000	0.990

4.4- Milk Protein:

Results of sire evaluations for Milk Protein are given in tables 4.4.1 and 4.4.2. Table 4.4.2 holds information about the distribution of EBVs and EPN for sires as well as PEVs and $r_{II}^2$of estimates. This information covers all sires having at least 5 progeny plus ½ grand progeny in the data set but not necessarily EPN = 5.

Rank correlations of sire evaluations for different comparisons for Milk Protein are given in table 4.4.1. Where applicable, the expected correlation of the comparisons are also given. The lowest correlation corresponds to comparison of purebred record to crossbred records for sires of the Jersey breed (0.080) and the highest correlation corresponds to comparison of all records (purebred and crossbred records together) with genetic groups included to purebred records for sires of Jersey breed (0.993). These correlations are for sires having more than 5 progeny plus ½ grand progeny in the data sets and the distribution of their progeny number is given in table 4.4.2.

Table 4.4.1: Rank correlations of sire evaluations using different data sets (for milk protein).

Type of Comparison[1]	Number of Sires	Correlation	Expected[2] Correlation
ALL-HH	1092	0.981	*
ALL-HJ	86	0.570	*
ALL-JJ	1069	0.993	*
ALL-JH	82	0.165	*
ALL-HJG	86	0.578	*
ALL-JHG	82	0.174	*
HH-HJ	71	0.371	0.523
JJ-JH	41	0.080	0.536
HH-HJG	71	0.394	0.539
JJ-JHG	41	0.088	0.598

[1, 2] see table 4.2.5

Table 4.4.2: Distribution of different statistics in evaluating sires (for milk protein).

EPN	N	MEAN	STDEV	SEMEAN	MINIMUM	MAXIMUM
EBV	2197	0.455	6.132	0.131	-32.900	34.900
EPN	2197	22.310	102.890	2.200	0.000	1812.150
$\hat{\text{PEV}}$	2197	155.810	48.040	1.020	1.800	273.800
$\hat{r}_{TI}^2$	2197	0.708	0.134	0.003	0.000	1.000
PEV	2197	161.420	49.210	1.050	2.600	273.800
r_{TI}^2	2197	0.410	0.179	0.004	0.000	0.990

4.5- Days in Milk:

Results of sire evaluations for Days in Milk are given in tables 4.5.1 and 4.5.2. Table 4.5.2 holds information about distribution of Estimated Breeding Values and Effective Numbers of Progeny for sires as well as Prediction Error Variances and Reliabilities of estimates. This information covers all sires having at least 5 progeny plus ½ grand progeny in the data set but not necessarily EPN = 5.

Rank correlations of sire evaluations for different comparisons for Days in Milk are given in table 4.5.1. Where applicable, the expected correlations of comparisons are also given. The lowest correlation corresponds to the comparison of purebred records to crossbred records with genetic groups included for sires of the Jersey breed (-0.018) and the highest correlation corresponds to the comparison of all records (purebred and crossbred records together) with genetic groups included to purebred records for sires of the Holstein breed (0.981). These correlations are for sires having more than 5 progeny plus ½ grand progeny in the data sets and the distribution of their progeny number is given in table 4.5.2.

Table 4.5.1: Rank correlations of sire evaluations using differentt data sets (for days in milk).

Type of Comparison[1]	Number of Sires	Correlation	Expected[2] Correlation
ALL-HH	1092	0.981	*
ALL-HJ	86	0.398	*
ALL-JJ	1069	0.980	*
ALL-JH	82	0.135	*
ALL-HJG	86	0.398	*
ALL-JHG	82	0.130	*
HH-HJ	71	0.235	0.501
JJ-JH	41	-0.013	0.516
HH-HJG	71	0.242	0.518
JJ-JHG	41	-0.018	0.583

[1, 2] see table 4.2.5

Table 4.5.2: Distribution of different statistics in evaluating sires (for days in milk).

	N	MEAN	STDEV	SEMEAN	MINIMUM	MAXIMUM
EBV	2197	-0.235	11.258	0.240	-56.000	55.100
EPN	2197	22.280	102.700	2.190	0.000	1808.540
$\hat{PEV}$	2197	607.610	183.250	3.910	7.700	1035.900
$\hat{r}_{TI}^2$	2197	0.682	0.140	0.003	0.000	1.000
PEV	2197	629.390	187.170	3.990	10.900	1035.900
r_{TI}^2	2197	0.392	0.181	0.004	0.000	0.990

CHAPTER FIVE

CHAPTER FIVE

DISCUSSION

Over-estimation of r_{TI}^2 by the approximate method used here was high when all sires were considered but for sires with greater effective progeny number over-estimation was decreased. This figure was around 8 % for sires having more than 50 effective progeny number. The comparison of true r_{TI}^2 and its estimate proves that estimating r_{TI}^2 by an approximate method for sires having low EPN would be greatly biased upward. Estimating r_{TI}^2 was done only as an additional piece of information. When the number of bulls used as sires of the next generation is high it would not be necessary to take these estimations into account for selection purposes. The reason is that the estimated breeding value has already taken account of reliability of the proof. In cases where a small number of bulls are used as sires of population, to avoid risk as a result of error in estimating breeding value, the reliability of estimates should be taken into account. However, whether reliabilities are considered or not the mean genetic gain would be the same but using sires with higher reliabilities would decrease the variation in genetic gain.

By increasing the EPN of sires, under-estimation of the prediction error variance was slightly increased which is mainly because of the increase in the percentage of related sires in the data set. These relationships come about because 1) bulls are related and 2) bulls are common sires in one or more contemporary group. When relationships are included the reciprocal of diagonal elements are not an accurate estimate of PEVs because the coefficient matrix deviates from sparseness (Weller et.al., 1985). When sires are not related the approximation method can be used as an

alternative to direct estimation but when relationships are high the best method would be direct calculation of PEV. However, the correlations of r_{TI}^2 and its estimate and PEV and its estimate were high which demonstrates the consistency of changes in them.

The correlation of ranks of Holstein sires evaluated for milk volume using only purebred progeny with ranks of the same sires evaluated using both purebred and crossbred progeny (ALL-HH) is high which supports using all progeny to estimate breeding value. The correlation for this type of comparison (ALL-JJ) for Jersey sires was also high. When using all progeny (both purebred and crossbred) to estimate breeding values in addition to HYS, the breed of animal should be taken into account to form contemporary groups otherwise breed effects would be confounded with breeding values.

The correlation of ranks of sires evaluated using only crossbred progeny with or without including genetic groups with ranks of the same sires evaluated using only purebred progeny was high for Holstein bulls (HH-HJ and HH-HJG) but for Jersey bulls (JJ-JH and JJ-JHG) were considerably less than expected. This result suggest that using only crossbred progeny to evaluate Jersey sires may give biased results due to unknown reasons. This problem is deserving of further investigation.

The correlation of ranks of Holstein sires evaluated using only crossbred progeny with or without including genetic groups with ranks of the same sires evaluated using both purebred and crossbred progeny (ALL-HJ and ALL-HJG) is high which again suggests that only crossbred progeny may be used to evaluate Holstein sires. These correlations (ALL-JH and ALL-JHG) are considerably lower for Jersey sires than that of Holstein sires and suggests that crossbred progeny do not give correct unbiased estimates of EBVs for Jersey sires.

When comparing ranks of sires evaluated using only crossbred progeny the correlations were similar whether genetic groups were included or not. This result suggests that for this data set including genetic groups in evaluating sires using only crossbred progeny was not necessary.

Similar results were obtained for evaluating sires for milk fat and milk protein. These results were also mostly similar for days in milk with the exception that correlation of ranks of Holstein sires evaluated with crossbred progeny with ranks of the same sires evaluated using purebred or both purebred and crossbred progeny was not as high as it was for other traits. This may be because of lower heritability of this trait (see appendix 4 for different estimates of h^2).

With increasing number of progeny per sire, the correlation of ranks evaluated with different data sets increased but the correlation of ranks of Jersey sires evaluated with crossbred progeny and their ranks evaluated with the other data sets remained low. Because of the increased reliability of estimates, the expected correlation between evaluations also increased.

The low correlations of ranks for Jersey sires evaluated with crossbred progeny with their ranks evaluated by other data sets has no clear explanation but several factors may cause this low correlation. Selective mating is one of the likely candidates. In New Zealand some farmers prefer to mate their cows producing low fat percentage with Jersey bulls and to mate those cows producing high fat percentage with Holstein bulls (P. Shannon, NZDB, pers. com.). The effect of corrective mating may be higher for Holstein cows mated to Jersey bulls than for Jersey cows mated to Holstein bulls. This is so because Jersey cows may be mated to Holstein bulls as a result of a grading-up process, rather than as a deliberate attempt at crossbreeding.

To see whether specific sires cause the low correlation or not, the EBVs estimated with different data sets were plotted using. Observing these plots confirms that there is no specific sire which reduces the correlation. These plots are given in appendix III.

Other possible explanations for reducing the correlation, could include different stocking rates for each breed, different maternal ability for each breed and specific combining ability of Jersey sires to Holstein cows.

The results of this analysis strongly supports the use of crossbred progeny records to evaluate Holstein sires (under New Zealand conditions). This makes increasing the number of progeny to evaluate dairy bulls possible without extra costs. Using this extra information increases reliabilities of breeding value estimates. The advantage of this method can be used in New Zealand where crossbreeding is part of dairy farming. As the direction of crossbreeding in New Zealand is to use Holstein sires to mate Jersey cows, the correlation of the ranks of Holstein sires evaluated using different data sets supports the use of this extra information at no extra costs. Using both purebred and crossbred progeny to evaluate Jersey bulls also can be practiced but further research is needed to investigate the reasons for the low correlation of ranks of Jersey bulls using crossbred progeny only with the ranks of the same sires evaluated with other progeny records. Before finding the appropriate models to evaluate Jersey bulls using their crossbred progeny, applying these records in Jersey sire evaluation is not recommended because the result would be biased. For further research, more information about the pedigree of animals and more precise percentage of breed composition of crossbred animals would be of assistance.

Plotting the EBVs of sires estimated using different data sets for different traits (Appendix III) shows that the low correlation of some comparisons is not because of

the presence of some specific sires. The existence of two lines are clearly observable in figures A.3.1 and A.3.3. It should be stated that the actual number of points for figures A.3.1 and A.3.3 were more than those presented in these figures. To prevent confusion of the plots, those points with equal integer part of two axes were omitted.

To find the reason for the presence of these two lines, coefficients of regression of EBVs of sires evaluated using all progeny on EBVs of the same sires evaluated using only purebred progeny were calculated. Tables 5.1 and 5.2 contain these coefficients for sires having different numbers of effective crossbred progeny as well as different proportions of effective crossbred progeny. As seen in the table 5.1 these coefficients are lower for sires having lower numbers of effective crossbred progeny or sires without crossbred progeny. The coefficient of regression is higher for sires having 20 or more effective crossbred progeny. To confirm that the number of crossbred progeny is the reason for the forming of two distinct lines, several sires were traced from the plot. Out of 54 sires which were traced 7 of them were located on the line with the slope of approximately equal to 1 and the remaining 47 sires were located on the line with the approximate slope of 0.9. Comparing the effective crossbred progeny number of these two groups of sires supports the hypothesis that differences in effective crossbred progeny number of sires resulted in them being located on these two lines. Sires located on the first line (slope = 1.0) had on average 24 crossbred progeny while those on the second line (slope = 0.9) had, on average, only 6 crossbred progeny. However, one could argue that there is a continuum between 6 and 24 crossbred progeny. Looking at table 5.1 reveals the lack of this continuum because coefficients of regression for sires having 0 to 20 crossbred progeny are nearly the same (0.90-0.92) whereas coefficients of regression for sires having more than 20 crossbred progeny are greater (about 1.00) and again nearly the same for different categories of crossbred progeny numbers. Although the reasons underlying the formation of the two lines are not known it seems that the number of crossbred

progeny is one of the likely reasons.

Table 5.2 shows that the proportion of crossbred progeny has no consistent effect on the regression coefficient. Therefore, it is the actual number of crossbred progeny rather than the proportion that is important in establishing the two distinct lines.

The different coefficients of regression suggest that sires evaluated using only purebred progeny would rank slightly different when evaluated using all purebred and crossbred progeny if they have only low numbers of crossbred progeny. Thus, if Holstein sires are to be evaluated using all their progeny, including all of their purebred and crossbred progeny would add to the precision of the proofs.

Disconnectedness of contemporary groups could also produce different plotted area through creating different bases for sires of each disconnected group. In this analysis, that was not the case, because, sires were from connected contemporary groups. This was proved by checking the connectedness of contemporary groups using a Fortran program (D.J. Garrick, pers. comm.). Checking the data file by this program showed that all contemporary groups which contained progeny of sires involved in the evaluation were connected.

The overall conclusion of this study is that crossbred progeny can be used in evaluating Holstein sires either separately or in combination with purebred progeny. For Jersey sires, crossbred progeny can only be used in combination with purebred progeny. Further investigation is needed to find the reasons which prevent the use of crossbred progeny in evaluating Jersey sires.

Table 5.1: Coefficients of regression of EBVs of Holstein sires evaluated using both pure- and crossbred progeny on the EBVs of the sires evaluated using only purebred progeny and with different effective crossbred progeny numbers (for milk volume).

Number of Crossbred progeny	Coefficient of regression
0	0.908
>0	0.926
0-5	0.904
5-10	0.906
10-15	0.923
15-20	0.928
20-25	1.010
25-30	0.997
30-35	0.992
35-40	1.130
40-45	*
45-50	*
>50	0.997

* = There were not enough number of sires in these categories to calculate the coefficient of regression.

Table 5.2: Coefficients of regression of EBVs of Holstein sires evaluated using both pure- and crossbred progeny on the EBVs of the sires evaluated using only purebred progeny and with different percentage of effective crossbred progeny (for milk volume).

Percent of progeny being crossbred	Coefficient of regression
0-10	0.920
10-20	0.892
20-30	0.936
30-40	0.952
40-50	0.852
50-60	0.924
60-70	0.887
70-80	0.897
80-90	0.947
90-100	0.983

REFERENCES

REFERENCES

Alba, J.DE and B.W. Kennedy, 1985. Milk production in the Latin-American Milking Crioll and its crosses with the Jersey. *Animal Production,* **41:** 143-150.

Ashmawy, A.A., 1987. Genetic and phenotypic parameters for production and stayability in British Friesian-Holstein cattle. *Egyptian Journal of Animal Production,* **25:** 117-123.

Berger, P.J., G.R. Luecke, and J.A. Hoekstra, 1989. Iterative algorithms for solving mixed model equations. *Journal of Dairy Science,* **72:** 514-522.

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. *Journal of Animal Science,* **58:** 1090-1096.

Brumby, P.J., 1959. The grazing behaviour of dairy cattle in relation to milk production, liveweight and pasture intake. *New Zealand Journal of Agricultural Research,* **2**: 797-807.

Bryant, A.M. and C.W. Holmes, 1985. Utilization of pasture on dairy farms. *Proceedings of conference on The Challenge: Efficient Dairy Production organized by the Australian and New Zealand Societies of Animal Production,* pp.48-63.

Coop, I., 1987. *Livestock in New Zealand.* Published by the New Zealand Society of Animal Production.

Dairy Statistics, 1988-89. Livestock Improvement, New Zealand Dairy Board

Dalton, C. and C. Rumble, 1985. *50 years of artificial insemination and herd improvement in New Zealand.* Auckland Livestock Improvement Association, Private Bag, Hamilton, New Zealand.

Dempfle, L., 1984. Principles of estimation of breeding values. In: Progeny testing methods in dairy cattle, *Bulletin of the International Dairy Federation.* **183**: 24-38.

Evans, D.A., 1969. Research on current sire proving procedures. *Proceedings of the New Zealand Society of Animal Production,* **29**: 77-86.

Everett, R.W. and L.P. Jones, 1985. Breeding values - Their calculation and potential use. *Proceedings of the conference on The Challenge: Efficient Dairy Production.* Australian and New Zealand Societies of Animal Production.

Falconer, D.S., 1981. *Introduction to Quantitative Genetics.* Second edition (reprint), Longman, London and New York.

Famula, T.R., E.J. Pollak and L.D. Van Vleck, 1983. Genetic groups in dairy sire evaluation under a selection model. *Journal of Dairy Science,* **66:** 927-934.

Freer, M., 1960. The utilization of irrigated pastures by dairy cows, II. The effect of stocking rate. *Journal of Agricultural Science,* **54**: 243.

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. thesis, Cornell University, New York, USA.

Garrick, D.J., 1991. Best linear unbiased prediction for across-flock/year breeding values. *The proceedings of New Zealand Society of Animal Production.* Volume 51, Otago University, Dunedin.

Gastinel, N., 1970. *Linear Numerical Analysis.* Herman, Paris.

Greenhalgh, E.M. and R.L. Quaas, 1986. Approximating prediction error variance for multiple trait sire evaluations. *Journal of Dairy Science,* **73:** 513-531.

Groeneveld, E. and M. Kovac, 1990. A generalized computing procedure for setting up and solving mixed linear models. *Journal of Dairy Science,* **73:** 513-531.

Hafez, E.S.E., 1964. Comparative analysis of grazing behaviour. *Proceedings of the 9th international congress,* Vol. **2**: 1039-1042.

Henderson, C.R., 1975a. Inverse of a matrix of relationship due to sires and maternal grandsires. *Journal of Dairy Science,* **58:** 1917-1921

Henderson, C.R., 1975b. Use of relationship among sires to increase accuracy of sire evaluation. *Journal of Dairy Science,* **58:** 1731-1738

Henderson, C.R., 1975c. Best linear unbiased estimation and prediction under a selection model. *Biometrics,* **31:** 423-447.

Henderson, C.R., O.V. Kempthorne, S.R. Searle and C.M. Von Krosigk, 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics,* **15:** 192-218.

Henderson, C.R., 1963. Selection index and expected genetic advance. In: *Statistical Genetics and Plant Breeding,* Edited by W.D. Hanson & H.F. Robinson. National Academy of Science-National Research Council, Washington, D.C.

Henderson, C.R., 1972. Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetics Symposium in honour of Dr. Jay L. Lush.*

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.

Hermas, S.A., C.W. Young, J.W. Rust, 1987. Genetic relationships and additive genetic variation of productive and reproductive traits in Guernsey dairy cattle. *Journal of Dairy Science,* **70:** 1252-1257.

Holmes, C.W. and G.F. Wilson, 1987. *Milk Production From Pasture.* Butterworths of New Zealand (Ltd) 205-207 Victoria St. CPO Bpx 472, Wellington New Zealand.

Holmes, C.W., 1984. *Dairy Production in New Zealand.* Dairy husbandry Department, Massey University, Palmerston North, New Zealand.

Holmes, C.W., 1990. Personal communication. Animal Science Department, Massey University, Palmerston north, New Zealand.

Hornbeck, R. W., 1975. *Numerical methods*. Quantum Publishers, Inc., New York, NY.

Jasiorowski, H., Z. Reklewski and M. Stolzman, 1983. Testing of different strains of Friesian cattle in Poland. I. Milk performance of F_1 paternal Friesian crosses under intensive feeding conditions. *Livestock Production Science,* **10**: 109-122.

Kennedy, B.W. and J.E. Moxley, 1975. Comparison of genetic group and relationship methods for mixed sire evaluation. *Journal of Dairy Science,* **58:** 1507-1514.

Kennedy, B.W., 1981. Bias and mean square error from ignoring genetic groups in mixed model sire evaluation. *Journal of Dairy Science,* **64:** 689-697.

Khan, U.U.N., 1986. The adaptability of crossbred dairy cattle to environmental conditions in Pakistan. *Dissertation Abstracts International, B (Sciences and Engineering),* **46:** 2191. University of Georgia, USA.

Lush, J.L., 1984. *Animal Breeding Plans*. Fifth facsimile printing, Iowa State University Press, Ames, Iowa.

Macmillan, K.L., 1982. Past and predicted trends in breeds, genetic improvement and breeding management in dairy cattle. *Proceedings of Ruakura Farmers' Conference (1982),* 34th conf., Hamilton, New Zealand, pp. 83-90.

MacMillan, K.L., 1985. Objectives of a breeding programme. *Proceedings of conference on The Challenge: Efficient Dairy Production organized by the Australian and New Zealand Societies of Animal Production.* pp.297-323.

MacMillan, K.L., D.M. Duganzich and A.M. Bryant, 1981. Production differences between Jersey and Friesian x Jersey cows in commercial dairy herds. *Proceedings of*

the New Zealand Society of Animal Production. **41**: 48-52.

Misztal, I. and D. Gianola, 1987. Indirect solution of mixed model equations. *Journal of Dairy Science,* **70:** 716-723.

Newell, P.B., 1973. *Sharemilking : A guide to owners and farm workers contemplating entering sharemilking agreements.* Ministry of Agriculture and Fisheries , Auckland, New Zealand.

NZDB, 1987-88. *Annual Report.* Livestock Improvement Division, New Zealand Dairy Board, Hamilton, N.Z.

NZDSSR, 1988-89. *New Zealand Dairy Sire Survey Register.* Livestock Improvement, Hamilton, N.Z.

NZJCBA, 1932. *The Jersey in New Zealand.* New Zealand Jersey Cattle Breed Association [Incorporated].

NZSER, 1988-89. *New Zealand Sire Evaluation Register- Traits other than production.* Livestock Improvement, Hamilton, N.Z.

Pirchner, F., 1984. History of progeny testing, *Bulletin of the International Dairy Federation.* **183**: 9-23.

Polastre, R., J.C. Milagress, P.S. Assis, C.A. Fre, 1987. Genetic and environmental effects on the performance of Holstein-Friesian x zebu cows. 4. Lactation length. *Revisat da Souedade Brasileira De Zootecnia,* **16:** 254-260.

Pollak, E.J. and R.L. Quaas, 1983. Definition of group effects in sire evaluation models. *Journal of Dairy Science,* **66:** 1503-1509.

Quaas, R.L. and E.J. Pollak, 1981. Modified equations for sire models with groups. *Journal of Dairy Science,* **64:** 1868-1872.

Quaas, R.L., 1976. Computing the diagonal elements and inverse of a large

numerator relationship matrix. *Biometrix*, **32:** 940-953.

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 genetic and breeding unit, Univesity of Armidale, Australia.

Romcevic, L., R. Lazarevic, S. Vasovic and N. Nikitovic, 1984. Variation of dairy performance in Black-Pied cattle. *Veterinaria Yugoslavia,* **33:** 485-492.

Rorato, P.R.N., R.B. Lobo, F.A.M. Duarte, M.A.R.DE Freitas, 1986. Estimates of phenotypic and genetic parameters for production traits in Holstein cows in Brazil. *Revista Brasileira de Genetica,* **9:** 261-269.

Sang, B.C, Y.Y. Cho, B.C. Chee, 1986. Estimates of heritability and genetic correlation for milk production traits in Holstein cattle. *Korean Journal of Animal Sciences,* **28:** 179-183.

Schaeffer, S.R. and B.W. Kennedy, 1986. Computing strategies for solving mixed model equations. *Journal of Dairy Science,* **69:** 575-579.

Searle, S.R., 1982. *Matrix Algebra Useful for Statistics.* John Wiley & Sons.

Searle, S.R., 1961. Part lactations. II. Genetic and phenotypic studies of monthly milkfat yield. *Journal of Dairy Science,* **44**: 282-95.

Sengar, M.S, N.S. Tomar, V.K. Arora, 1987. Genetic and phenotypic studies of some economic traits of Tharparkar cows. *Indian Veterinary Journal,* **64:** 146-149.

Services and Sires, 1990. Livestock Improvement, New Zealand Dairy Board, Hamilton, N.Z.

Simmonds, J.G., 1985. Factors influencing calving date of seasonal supply dairy herds in New Zealand. *Proceedings of conference on The Challenge: Efficient Dairy Production organized by the Australian and New Zealand Societies of Animal Production.* pp.245-252.

Stockdale, C.R. and K.R. King, 1980. The effect 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. Hus.*, **20**: 529.

Sung Ho, H., 1986. *A Review of Some Factors Affecting Milk Production From Pasture*. Dissertation , Dip.Agr.Sc., Massey University, New Zealand.

Swake, A. and E. Bruns, 1986. Considering relationships and genetic grouping in mixed model sire evaluation: A simulation study. *Livestock Production Science,* **14:** 123-138.

Swalve, H., L.D. Van Vleck, 1987. Estimation of genetic (co)variances for milk yield in first three lactations using an animal model and restricted maximum likelihood. *Journal of Dairy Science,* **70:** 842-849.

Szkotnicki, W.J., A.K.W. Tong, M.A. Sharaby, K.M. Krotch, L.P. Johnson and L.R. Schaeffer, 1978. Sire and cow evaluation in Brown Swiss, Canadienne, and Milking Shorthorn. *Journal of Dairy Science,* **61:** 497-505.

Ufford, G.R., C.R. Henderson and L.D. Van Vleck, 1979. An approximate procedure for determining prediction error variance of sire evaluations. *Journal of Dairy Science,* **62:** 621-626.

Valle, A., F.A.DE Moura Duarte, 1980. Production and genetic parameters in Holstein-Friesians in a humid tropical climate in Brazil. *Agronomia Tropical,* **28:** 177-193.

Van Vleck, L.D. and D.J. Dwyer, 1985a. Comparison of iterative procedures for solving eqautions for sire evaluation. *Journal of Dairy Science,* **68:** 1006-1014.

Van Vleck, L.D., 1987. Contemporary groups for genetic evaluations, *Journal of Dairy Science,* **70:** 2456-2464.

Van Vleck, L.D., E.J. Pollak and E.A. Bradford Oltenacu, 1987. *Genetics for the Animal Sciences.* W.H. Freeman and Company. New York.

Van Vleck, L.D. and D.J. Dwyer, 1985b. Successive overrelaxation, block iteration, and method of conjugate gradients for solving equations for multiple trait evaluation of sires. *Journal of Dairy Science,* **68:** 760-767.

Weller, J.I., H.D. Norman and G.R. Wiggans, 1985. Estimation of variance of prediction error for best linear unbiased prediction models with relationships included. *Journal of Dairy Science,* **69:** 930-938.

Wickham, B.W., 1989. The New Zealand Dairy Herd: Genetic progress and future expectations. *Dairy farming annual,* Massey University, Palmerston North, New Zealand.

Wickham, B.W., M.A. Belsley , R.G. Jackson and W. Rumball, 1978. Evidence of genetic improvement of New Zealand dairy cattle. *New Zealand Journal of Experimental Agriculture,* **6**: 101-113.

Wilmink, J.B.M. and J. Dommerholt, 1985. Approximaet reliability of best linear unbiased prediction in models with and without relationships. *Journal of Dairy Science,* **68:** 946-952.

APPENDICES

APPENDIX I

APPENDIX I

DESCRIPTION OF THE FORTRAN PROGRAM:

72,480 lines of data each containing 118 characters and representing information for one animal were written in the original data file.

To sort data based on Herd-Year-Season, data file was read into a vector and 'qsort' subroutine was called from Unix operating system. To determine the type of sorting a function was written.

Type of evaluation was selected interactively (e.g. Holstein sires with their purebred progeny etc.)

Lower limit for the progeny plus ½ grand-progeny number was asked interactively.

Subroutine for changing the actual identification number of bulls to sequential numbers was called.

Number of Paternal Great grand sire , Paternal Grand sire, Maternal Grand sire and sires were counted.

Length of season for HYS was asked.

Parameters like phenotypic variances and heritabilities were given.

R^{-1} elements for different cases were calculated.

Reading data from sorted data file started.

Subroutine to change actual calving date to sequential numbers was called.

Appropriate coefficients for each HYS were written into system of equations at the end of reading data for each HYS and the subroutine to absorb HYS equations was called.

Last HYS was absorbed at the end of reading data file.

Effective Progeny Numbers were written into an array.

Appropriate values for G^{-1} for cases with and without grouping were calculated and were added to LHS matrix.

Subroutine to solve equations using Gauss-Seidel iterative method was called.

Decision was made to either estimate or calculate Variance of Prediction Error and Reliabilities of estimates. If calculation was needed, subroutine to calculate these values was called.

Results were written into files and/or screen.

List of Functions and Subroutines:

1- Function to determine type of sorting.

This function was written to determine order of sorting records (Ascending/descending) and also to show the parts of records upon which sorting should be done. Sorting was done by subroutine "qsort" of Unix operating system.

2- Function to estimate r_{TI}^2.

This function was written to estimate r_{TI}^2 for each sire based on the formula given in the text.

3- Subroutine to change actual IDs to sequential numbers.

As sires come from different herds and their ID number has wide range which is not suitable in setting equations and also arranging them based on their age was necessary in writing inverse relationship matrix, a subroutine was written to change sires' actual ID numbers to sequential and chronological numbers. Relations between the actual ID numbers and new coded numbers was maintained to use after evaluation.

4- Subroutine to count progeny and grand progeny numbers.

As mentioned in the text, a lower limit for the number of progeny plus ½ grand progeny for each sire was imposed as a condition for sires to be included in the evaluation. This subroutine counts the progeny plus ½ grand progeny of each sire.

5- Function to find position of each element of half-stored LHS matrix in array form.

Because large number of sires are being evaluated, the coefficient matrix is large. As this matrix is symmetric and sparse, using advantage of sparseness and symmetry would reduce the amount of memory needed. Therefore the diagonal elements plus half of the off-diagonal elements of coefficient matrix are stored in a vector and this subroutine finds the exact position of each element (row and column number).

6- Subroutine to change date of calving to sequential numbers.

Date of calving is used for generating herd-year-season and it is necessary to change it to sequential numbers to make measurement between dates easier. This subroutine changes the calving date to numbers 1 to 365 for each year.

7- Subroutine to absorb HYS equations.

Before solving equations for random effects, the fixed effects are absorbed. When data file is being read to form equations, after reading records of each HYS, that HYS is absorbed. This procedure saves considerable amounts of time and disk space in comparison to when at first all fixed and random effect equations are formed and then fixed effect equations are absorbed. Special features like compacting records of each HYS to reduce its size and speeding the calculation up was considered.

8- Subroutine to solve equations using Gauss-Seidel iterative method and advantages of sparseness.

In this subroutine, at first the half stored matrix was moved to smaller vector to keep only nonzero elements and two other vectors were used to keep the numbers of rows and columns for these elements. Then, Gauss-Seidel iterative method was used to solve the equations. To check the time of finishing the solving procedure, a convergence criterion was adopted.

9- Subroutine to calculate PEV and true r_{TI}^2.

To calculate true r_{TI}^2 and Prediction Error Variance, inverting the coefficient matrix is needed. Because this matrix is too large to be inverted directly, the iterative method was applied to calculate the diagonal elements of inverse matrix. The procedure is explained in the text.

APPENDIX II

Different Comparison of Correlations of Sire Evaluations

Using Differenet Sets of Data for Different Traits

Table A.2.1: Comparison of correlations of sire evaluations using different data sets for Milk Volume when different lower limits for progeny plus ½ grand progeny numbers were used:

	n>5[a]			n>20			n>35			n>50		
	N	1	2	N	1	2	N	1	2	N	1	2
ALL-HH	1092	0.972	*	218	0.965	*	136	0.981	*	103	0.977	*
ALL-HJ	86	0.658	*	41	0.658	*	32	0.683	*	23	0.733	*
ALL-JJ	1069	0.991	*	170	0.985	*	90	0.960	*	70	0.969	*
ALL-JH	82	0.277	*	30	0.277	*	20	0.683	*	17	0.679	*
ALL-HJG	86	0.667	*	41	0.667	*	32	0.690	*	23	0.733	*
ALL-JHG	82	0.277	*	30	0.277	*	20	0.653	*	17	0.679	*
HH-HJ	71	0.579	0.538	27	0.555	0.777	20	0.838	0.848	16	0.824	0.881
JJ-JH	41	0.162	0.546	20	0.463	0.754	14	0.512	0.824	14	0.464	0.824
HH-HJG	71	0.599	0.551	27	0.556	0.782	20	0.857	0.851	16	0.841	0.883
JJ-JHG	41	0.163	0.607	20	0.465	0.776	14	0.507	0.841	14	0.464	0.841

[a] Number of progeny plus ½ grand progeny.
N Number of sires used in comparison.
1 Correlations of EBVs of sires calculated using two different data sets.
2 Expected correlations of EBVs of sires.
Note: Expected correlations are only given for comparisons using independent data sets.

Table A.2.2: Comparison of correlations of sire evaluations using different data sets for Milk Fat when different lower limits for progeny plus ½ grand progeny numbers were used:

	n>5			n>20			n>35			n>50		
	N	1	2	N	1	2	N	1	2	N	1	2
ALL-HH	1092	0.985	*	218	0.973	*	136	0.991	*	103	0.989	*
ALL-HJ	86	0.570	*	41	0.626	*	32	0.512	*	23	0.466	*
ALL-JJ	1069	0.993	*	170	0.931	*	90	0.962	*	70	0.969	*
ALL-JH	82	0.098	*	30	-0.059	*	20	0.043	*	17	0.262	*
ALL-HJG	86	0.588	*	41	0.625	*	32	0.474	*	23	0.454	*
ALL-JHG	82	0.094	*	30	-0.029	*	20	0.094	*	17	0.262	*
HH-HJ	71	0.340	0.536	27	0.522	0.777	20	0.573	0.848	16	0.485	0.881
JJ-JH	41	0.110	0.546	20	0.244	0.754	14	0.354	0.824	14	0.370	0.824
HH-HJG	71	0.378	0.551	27	0.522	0.782	20	0.578	0.851	16	0.499	0.883
JJ-JHG	41	0.078	0.607	20	0.222	0.776	14	0.354	0.841	14	0.370	0.841

Table A.2.3: Comparison of correlations of sire evaluations using different data sets for Milk Protein when different lower limits for progeny plus ½ grand progeny numbers were used:

	n>5			n>20			n>35			n>50		
	N	1	2	N	1	2	N	1	2	N	1	2
ALL-HH	1092	0.981	*	218	0.969	*	136	0.989	*	103	0.986	*
ALL-HJ	86	0.570	*	41	0.747	*	32	0.674	*	23	0.364	*
ALL-JJ	1069	0.993	*	170	0.937	*	90	0.960	*	70	0.959	*
ALL-JH	82	0.165	*	30	0.278	*	20	0.330	*	17	0.369	*
ALL-HJG	86	0.578	*	41	0.765	*	32	0.672	*	23	0.360	*
ALL-JHG	82	0.174	*	30	0.299	*	20	0.384	*	17	0.363	*
HH-HJ	71	0.371	0.523	27	0.625	0.766	20	0.771	0.840	16	0.790	0.873
JJ-JH	41	0.080	0.536	20	0.352	0.744	14	0.356	0.816	14	0.351	0.815
HH-HJG	71	0.394	0.539	27	0.633	0.770	20	0.772	0.842	16	0.790	0.876
JJ-JHG	41	0.088	0.598	20	0.354	0.766	14	0.359	0.833	14	0.363	0.833

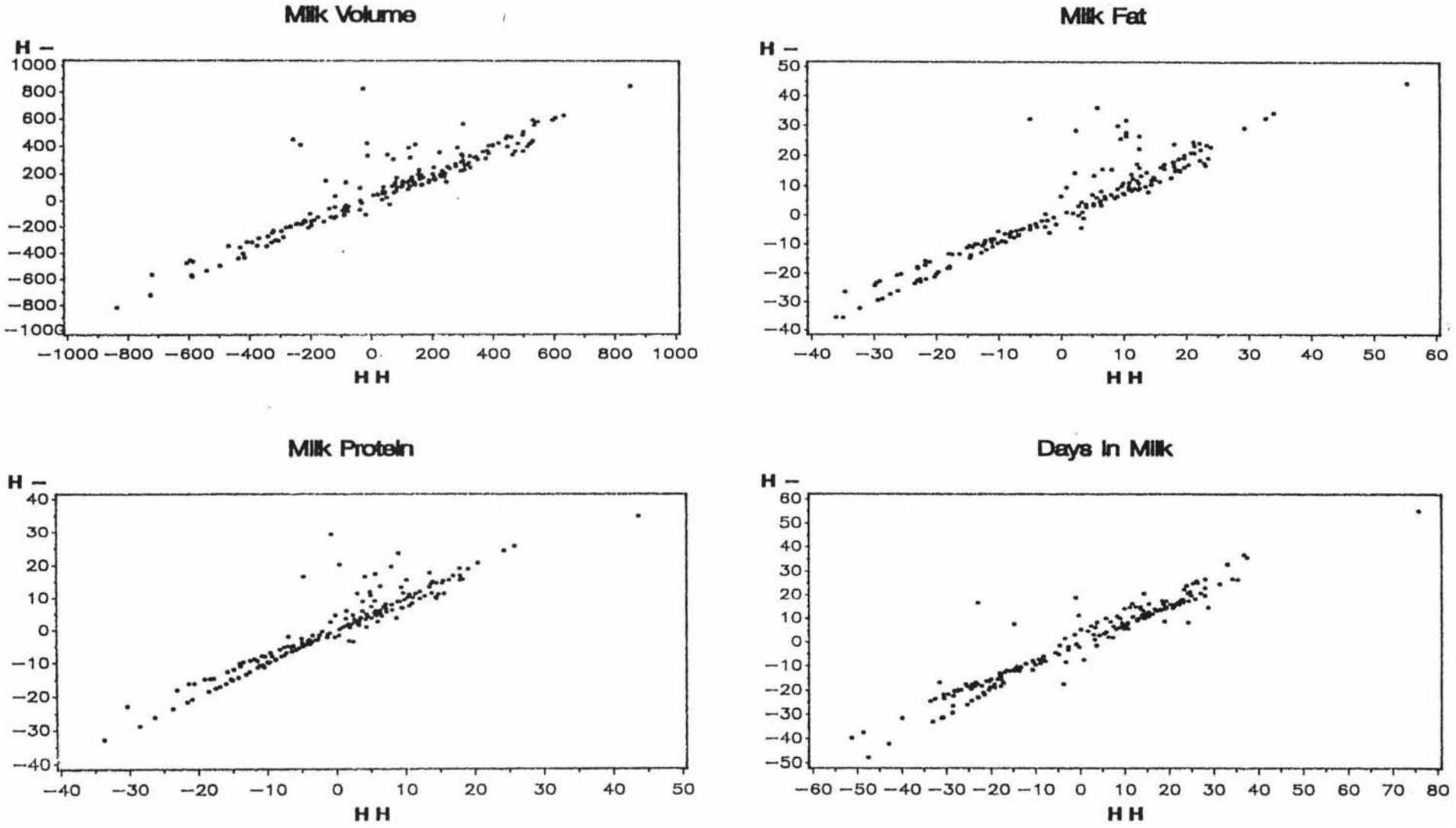
Table A.2.4: Comparison of correlations of sire evaluations using different data sets for Days in Milk when different lower limits for progeny plus ½ grand progeny numbers were used:

	n>5			n>20			n>35			n>50		
	N	1	2	N	1	2	N	1	2	N	1	2
ALL-HH	1092	0.981	*	218	0.982	*	136	0.987	*	103	0.986	*
ALL-HJ	86	0.398	*	41	0.488	*	32	0.423	*	23	0.364	*
ALL-JJ	1069	0.980	*	170	0.971	*	90	0.948	*	70	0.959	*
ALL-JH	82	0.136	*	30	0.024	*	20	0.224	*	17	0.369	*
ALL-HJG	86	0.398	*	41	0.457	*	32	0.401	*	23	0.360	*
ALL-JHG	82	0.130	*	30	0.017	*	20	0.209	*	17	0.363	*
HH-HJ	71	0.235	0.501	27	0.205	0.750	20	0.282	0.827	16	0.263	0.862
JJ-JH	41	-0.013	0.516	20	0.042	0.726	14	0.235	0.801	14	0.235	0.801
HH-HJG	71	0.242	0.518	27	0.208	0.755	20	0.283	0.829	16	0.263	0.865
JJ-JHG	41	-0.018	0.583	20	0.020	0.751	14	0.235	0.819	14	0.224	0.819

APPENDIX III

Figure A.3.1: Comparing Two Data Sets In Ranking Holstein Sires For Different Traits

(Number of Sires In Each Comparison = 1092)

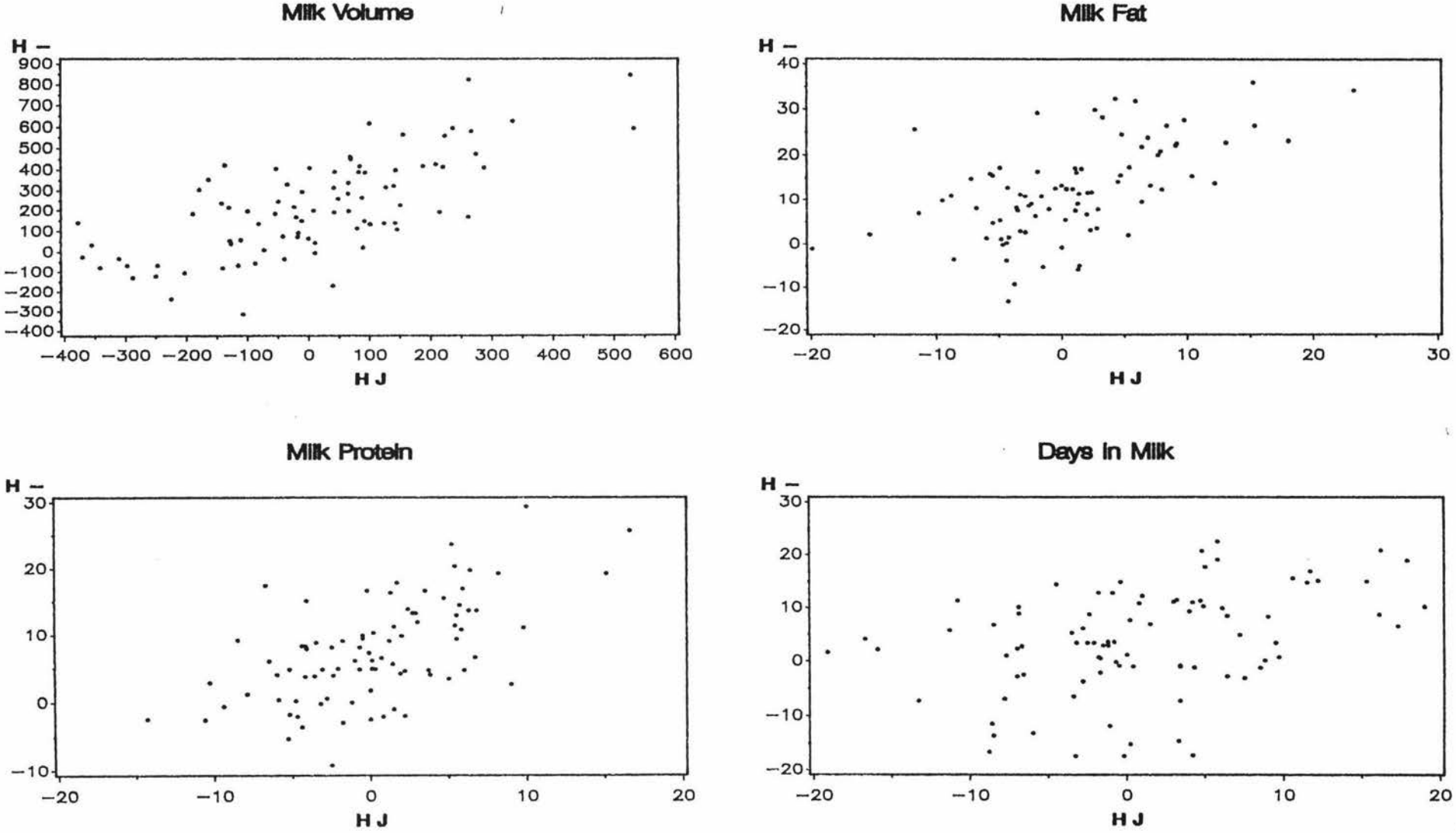


H H = EBVs Using Only Purebred Progeny

H — = EBVs Using All Progeny With Including Genetic Groups

Figure A.3.2: Comparing Two Data Sets in Ranking Holstein Sires For Different Traits

(Number of Sires in Each Comparison = 86)

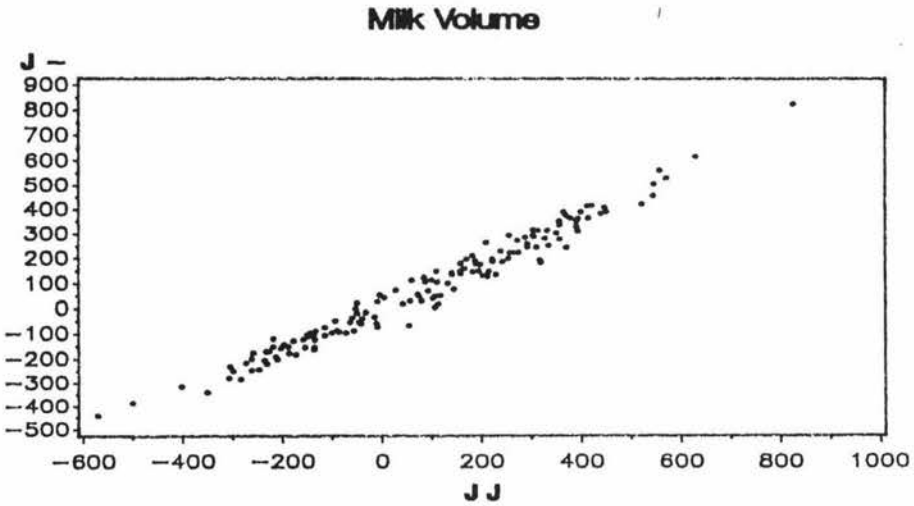


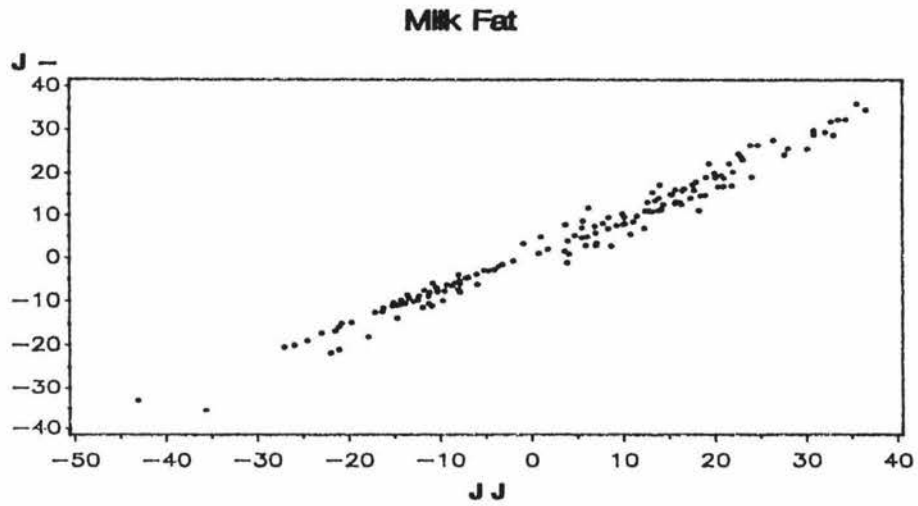
H J = EBVs Using Only Crossbred Progeny Without Including Genetic Groups

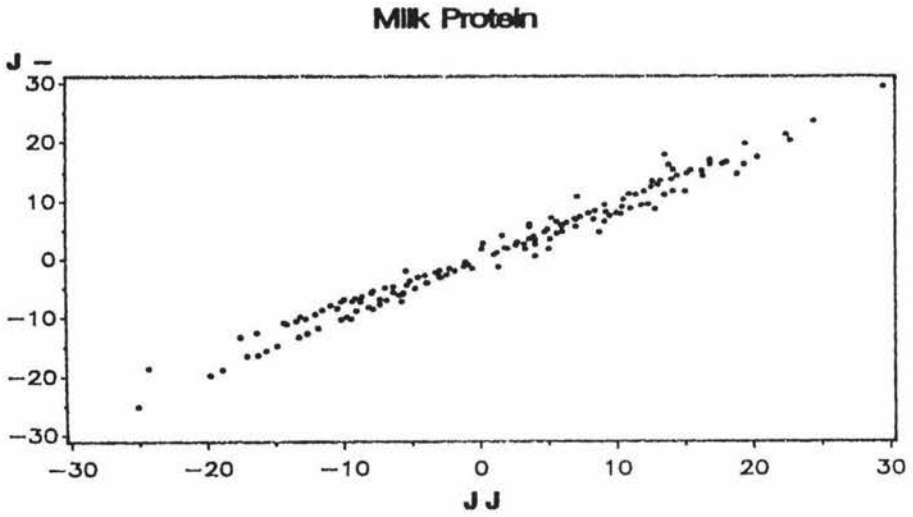
H – = EBVs Using All Progeny With Including Genetic Groups

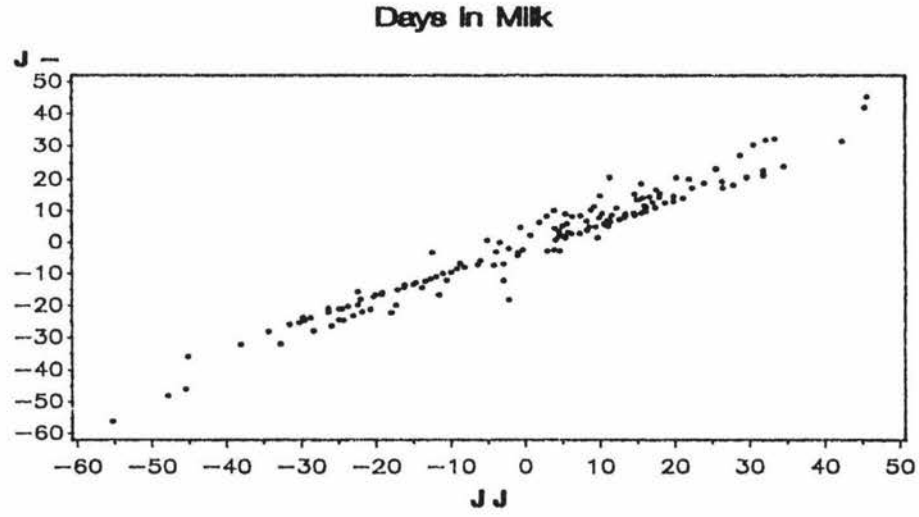
Figure A.3.3: Comparing Two Data Sets In Ranking Jersey Sires For Different Traits

(Number of Sires In Each Comparison = 1069)







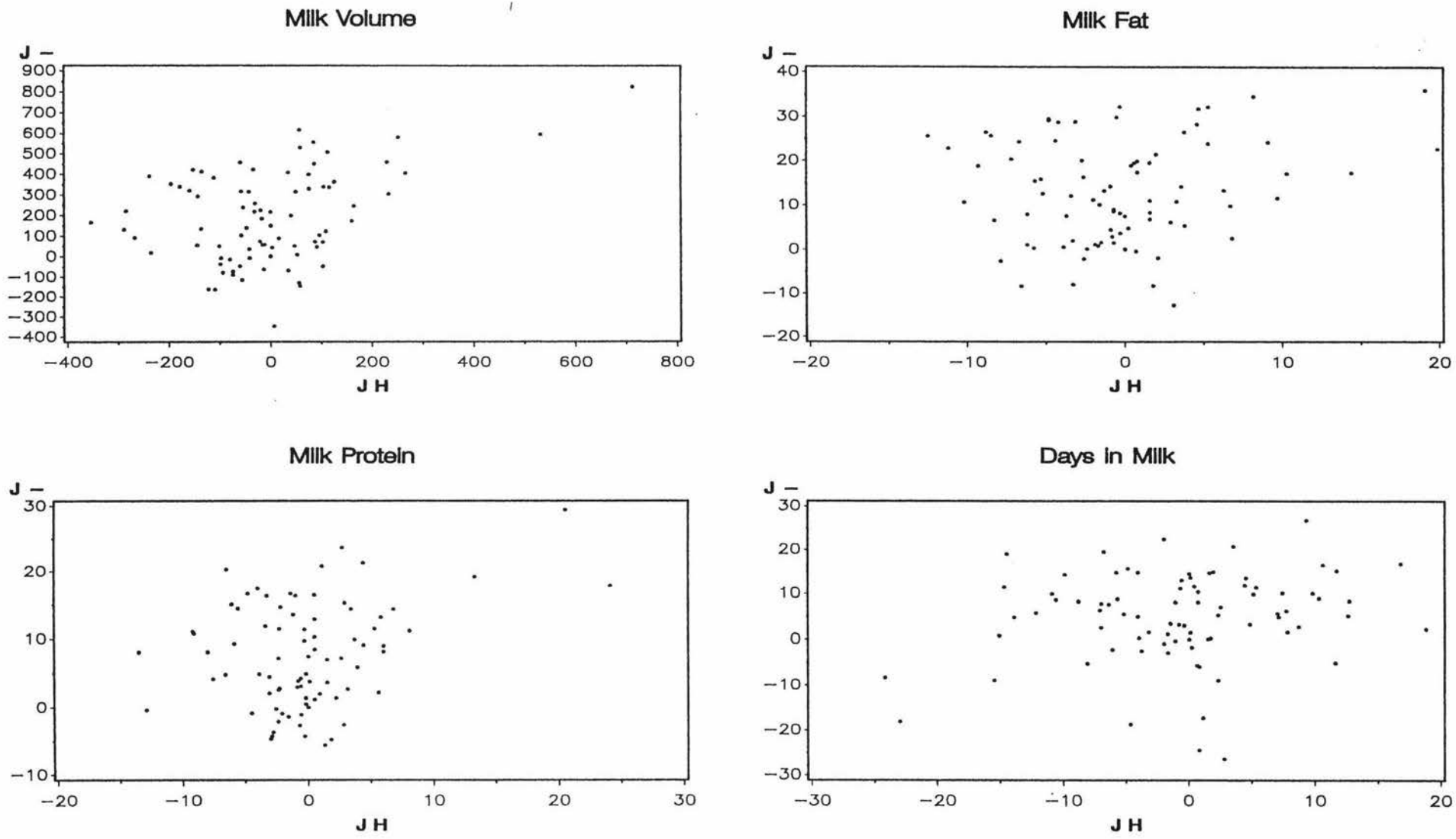


J J = EBVs Using Only Purebred Progeny

J – = EBVs Using All Progeny With Including Genetic Groups

Figure A.3.4: Comparing Two Data Sets in Ranking Jersey Sires For Different Traits

(Number of Sires in Each Comparison = 82)

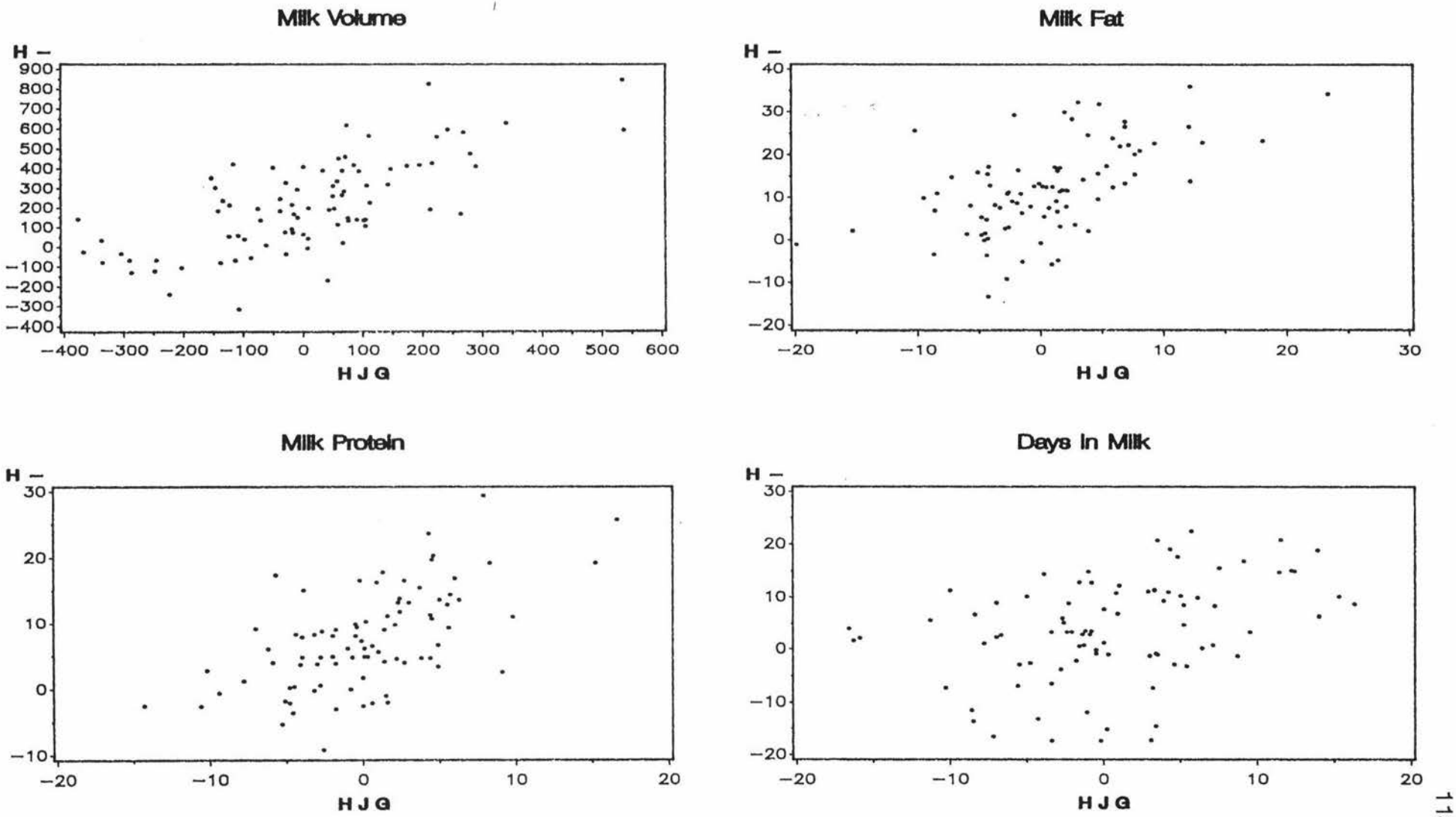


J H = EBVs Using Only Crossbred Progeny Without Including Genetic Groups

J – = EBVs Using All Progeny With Including Genetic Groups

Figure A.3.5: Comparing Two Data Sets in Ranking Holstein Sires For Different Traits

(Number of Sires in Each Comparison = 86)

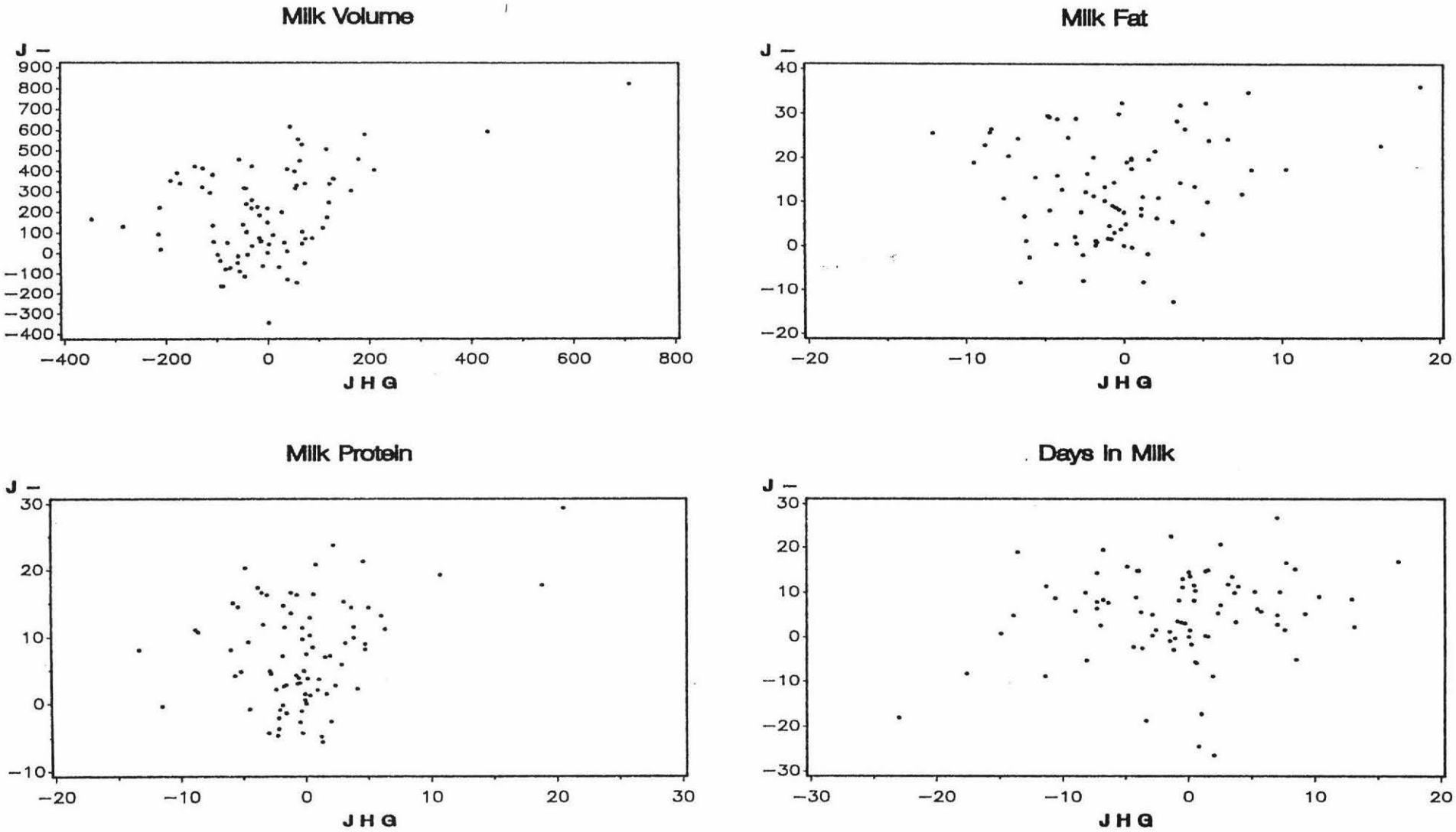


H J G = EBVs Using Only Crossbred Progeny With Including Genetic Groups

H – = EBVs Using All Progeny With Including Genetic Groups

Figure A.3.6: Comparing Two Data Sets in Ranking Jersey Sires For Different Traits

(Number of Sires in Each Comparison = 82)

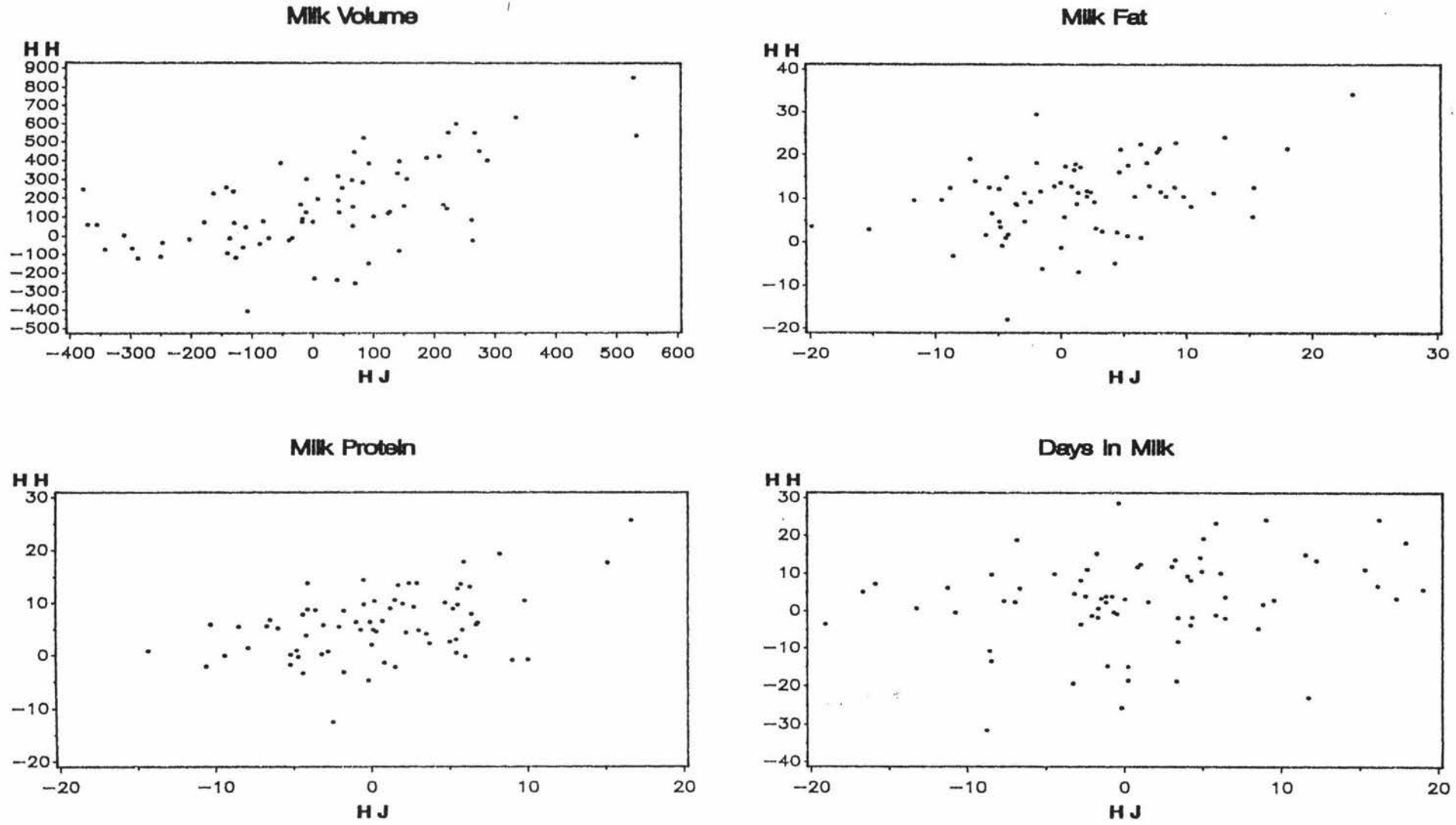


J H G = EBVs Using Only Crossbred Progeny With Including Genetic Groups

J — = EBVs Using All Progeny With Including Genetic Groups

Figure A.3.7: Comparing Two Data Sets In Ranking Holstein Sires For Different Traits

(Number of Sires In Each Comparison = 71)

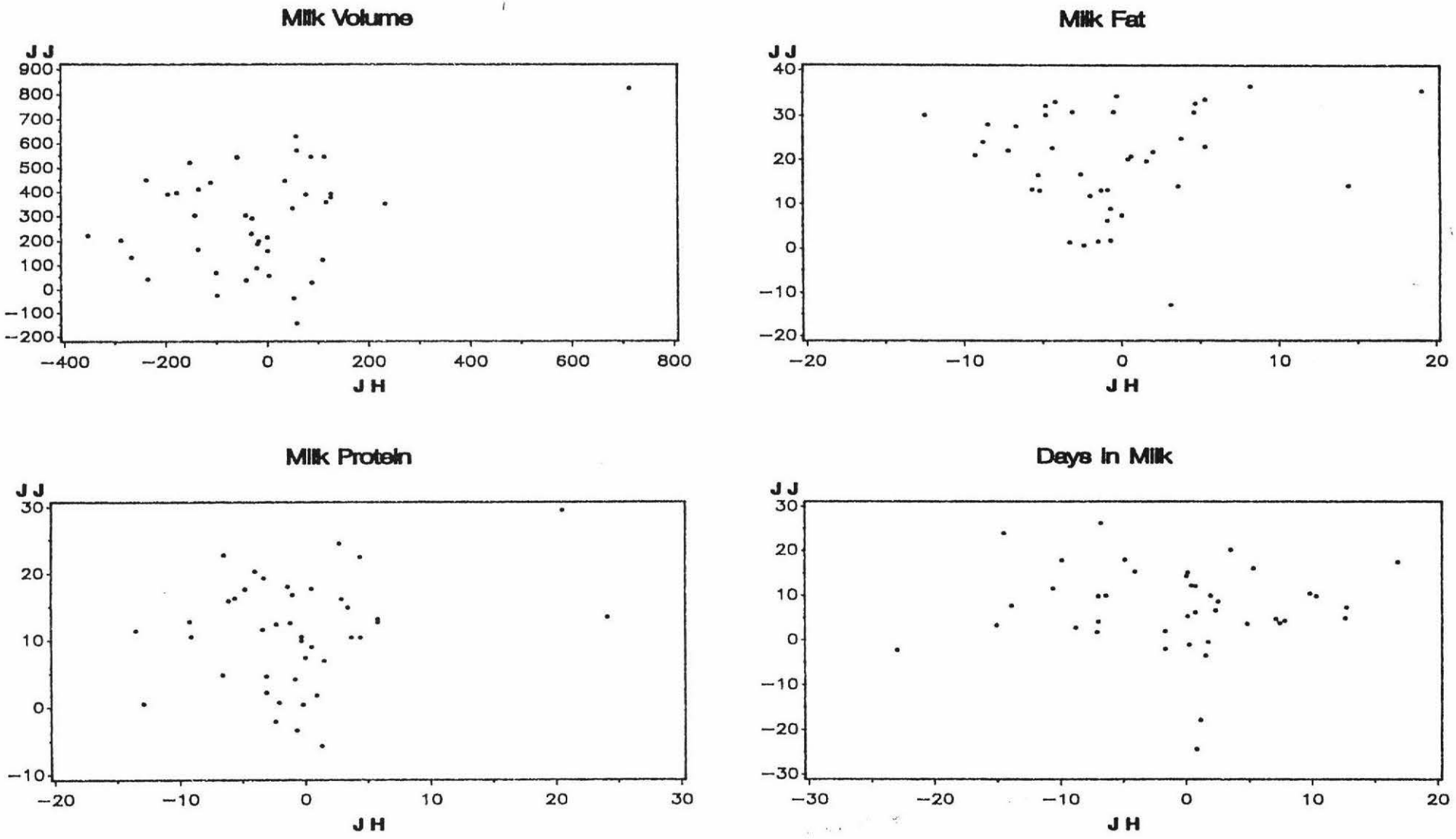


H H = EBVs Using Only Purebred Progeny

H J = EBVs Using Only Crossbred Progeny Without Including Genetic Groups

Figure A.3.8: Comparing Two Data Sets In Ranking Jersey Sires For Different Traits

(Number of Sires In Each Comparison = 41)

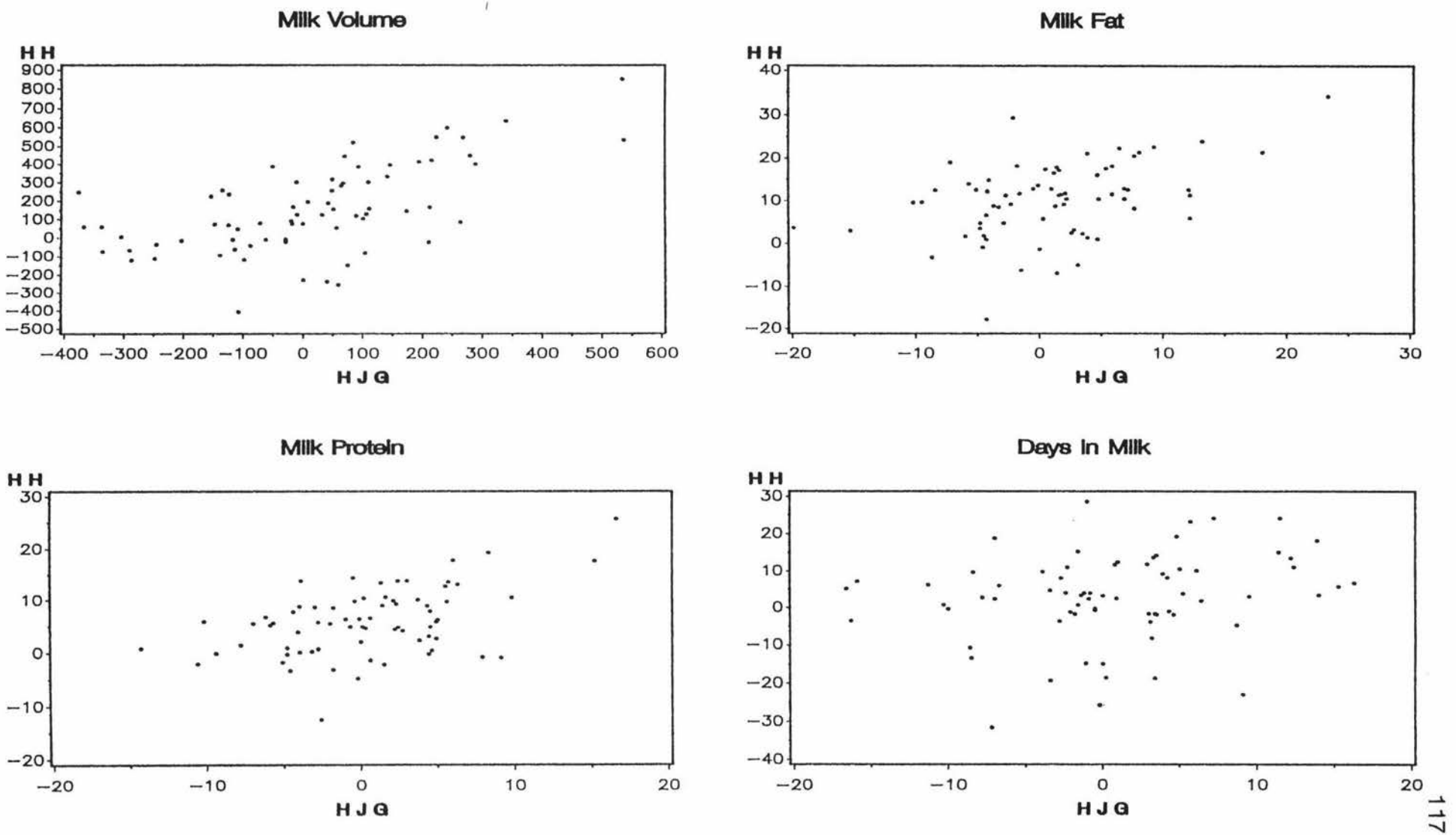


J J = EBVs Using Only Purebred Progeny

J H = EBVs Using Only Crossbred Progeny Without Including Genetic Groups

Figure A.3.9: Comparing Two Data Sets in Ranking Holstein Sires For Different Traits

(Number of Sires in Each Comparison = 71)

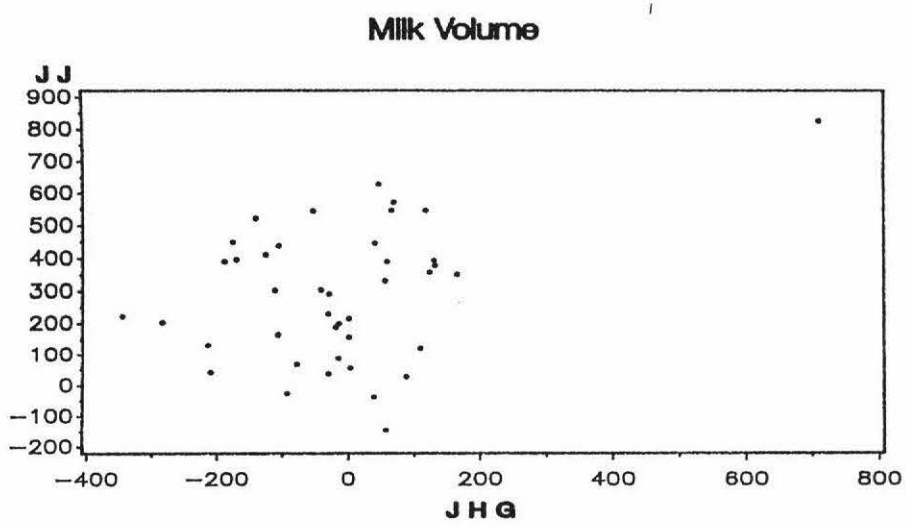


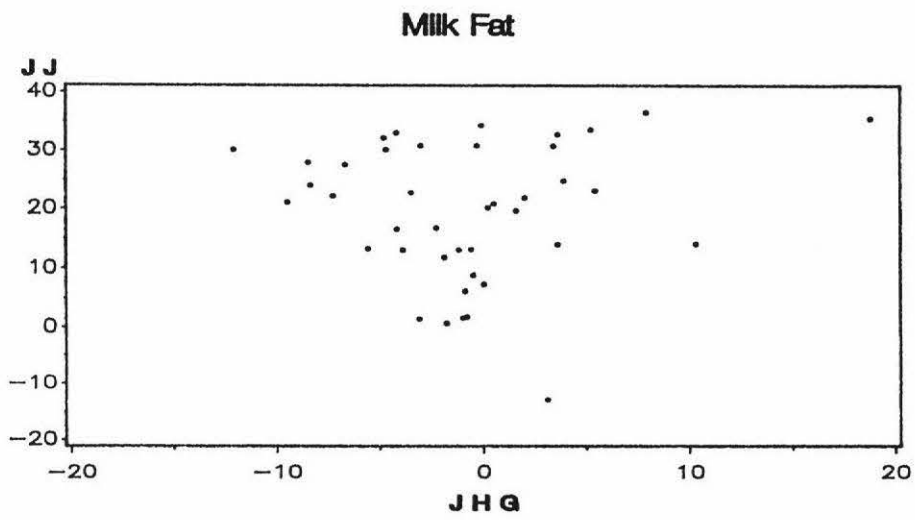
H H = EBVs Using Only Purebred Progeny

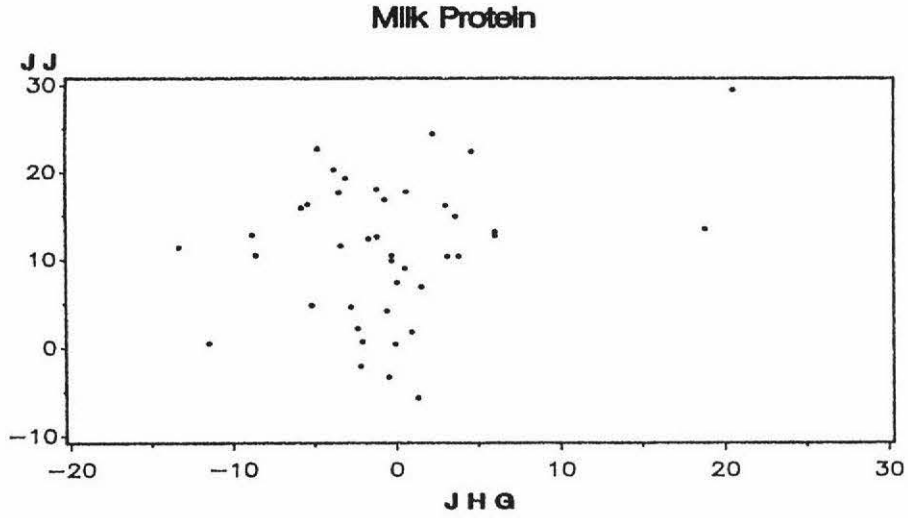
H J G = EBVs Using Only Crossbred Progeny With Including Genetic Groups

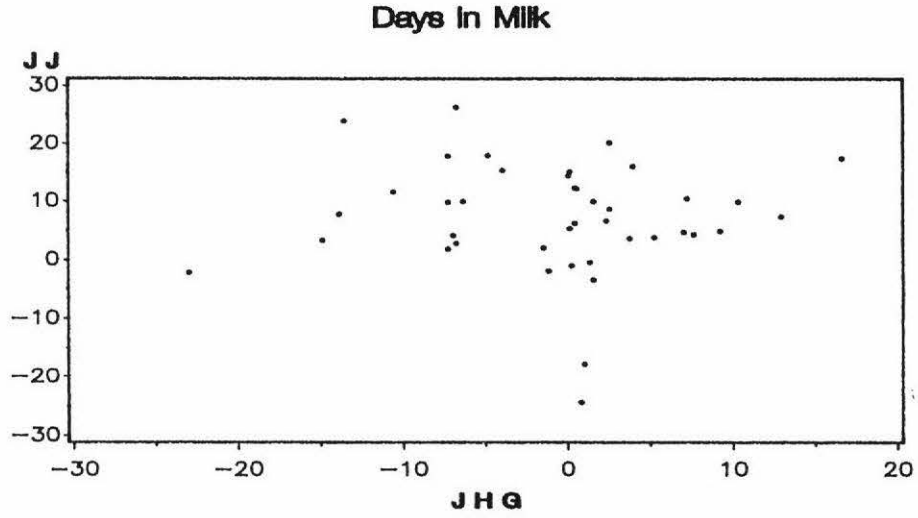
Figure A.3.10: Comparing Two Data Sets in Ranking Jersey Sires For Different Traits

(Number of Sires in Each Comparison = 41)









J J = EBVs Using Only Purebred Progeny

J H G = EBVs Using Only Crossbred Progeny With Including Genetic Groups

APPENDIX IV

Table A.4.1: Different estimates of heritabilities of dairy performance traits.

DIM[1]	MILK	FAT	PROTEIN	FAT%	BREED	AUTHOR(S)
0.13 ± 0.15	0.15 ± 0.16	0.11 ± 0.15	-	0.46 ± 0.31	Black-Pied	Romcevic et.al., 1984.
0.20 to 0.76	0.15 to 0.94	-	-	-	Crossbred of Holstein-Frisian and Jersey with Sahival and Red Sindhi	Khan, 1986.
0.17 ± 0.24	0.23 ± 0.26	-	-	-	Tharparkar	Sengar et.al.,1987.
0.13 ± 0.10	0.24 ± 0.12	0.26 ± 0.12	-	0.77 ± 0.15	Guernsey	Hermas et.al., 1987.
0.41	0.44	-	-	0.28	Dairy Criollos, Jersey and their reciprocal cross and back cross [2]	Alba and Kennedy, 1985.
0.08	0.28	-	-	0.35	same as above[3]	
0.10 ± 0.04	0.23 ± 0.06	0.24 ± 0.07	-	0.38 ± 0.09	Holstein	Sang et.al., 1985.
-	0.21 ± 0.06	0.23 ± 0.06	-	0.25 ± 0.07	Holstein	Rorato et.al., 1986
0.16	0.21	-	-	0.23	Holstein-Friesian	Valle & Moura Duarte, 1980.
0.12 ± 0.05	-	-	-	-	H-Friesian x zebu	Polastre et.al., 1987.
-	0.33	-	-	-	Holstein-Friesian	Swalve and Van Vleck, 1987.
-	0.24 ± 0.03	0.23 ± 0.03	0.22 ± 0.03	-	Holstein-Friesian	Ashmawy, 1987.

[1] Days in Milk.
[2] all lactation.
[3] lactations without presence of the calf.