This is to certify that the

dissertation entitled

ANALYSIS OF THE ERROR VARIANCE-COVARIANCE MATRIX
OF FIRST, SECOND AND THIRD LACTATIONS
IN MEXICAN HOLSTEINS

presented by

Manuel Villarreal

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Animal Science

Major professor

Date 10/28/85

0-12771

ANALYSIS OF THE ERROR VARIANCE-COVARIANCE MATRIX
OF FIRST, SECOND AND THIRD LACTATIONS
IN MEXICAN HOLSTEINS

By

Manuel Villarreal

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requeriments
for the degree of

DOCTOR OF PHILOSOPHY

Department of Animal Science

1985

tr

la

re

co

la

au

es

fo

re

ho

ex

or

co

ef

in

co

ABSTRACT

ANALYSIS OF THE ERROR VARIANCE-COVARIANCE MATRIX OF FIRST, SECOND AND THIRD LACTATIONS IN MEXICAN HOLSTEINS

by

Manuel Villarreal

Errors from each record of the same cow have been treated as independent of each other when analyzing later lactations. Need to incorporate more flexibility to this relation by allowing errors from each lactation to be correlated with each other and to vary from lactation to lactation in accordance with a first order non-stationary autoregressive process has been suggested. Use of recursive estimation techniques would be possible if residual errors follow this pattern. More accurate estimates of a cow's real ability could be obtained from these relationships.

This study tests the assumptions that there is homogeneous error variance-covariance among lactations; and examines the hypothesis that errors might follow a first order non-stationary autoregressive process.

Milk yield records of 3999 Mexican Holsteins with three consecutive lactations were used.

The model applied to each lactation included random effects of sire, fixed effects of herd-year-season, and days in lactation fitted as a covariable. Residuals were computed from solutions to the mixed model equations. The

var

res

the

err

hig

pro

ord

tha

var

was

rej

cor

cow

car

abi

acc

mod

ord

variance-covariance matrix S was produced from a vector of residuals for each lactation from each cow.

The S matrix is not homogeneous ($p < 0.01$) according to the Box test for homogeneous variance. In the S, matrix error variance increases with time and error covariances are higher in adjacent lactations. Using a maximum likelihood procedure to test the hypothesis that errors follow a first order autoregressive model was rejected ($p < .01$). A model that incorporated a parameter representing the possible variance increment caused by the maturing effect of the cow was also rejected ($p < .01$). Although these models were rejected they fitted S better than a model with no correlation of errors. Justifications would be that if a cow produces an unusually good or bad yield initially her caretaker would have a high or low expectation of her ability managing her accordingly, also cumulative effects of accidents or diseases could be carried over lactations. To model the error structure of repeated lactations, a first order non-stationary process should be used.

DEDICATION

I would like to dedicate this thesis to the following people: to my parents Maria Estela C. de Villarreal and Fernando Villarreal who instilled in me the importance of education; to my brothers and sisters for their example; to aunt Emma and uncle Fernando Riou remembering their support and friendship; to my American family Carlitos, Leticia and Gumecindo Salas for helping me to be me again, and finally, to the many friends who make life so worthwhile.

R. And

major

sugge

also

Drs. J

suppo

membe

Walte

exper

his he

suppor

F

Scien

gradua

for Dr

apprec

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

lact

lact

the

made

effec

those

recor

seaso

to th

repre

condi

recor

real

envir

"S
o
t
H
a
t
e
o

The av

I INTRODUCTION

The amount of milk that a cow produces varies from one lactation to another. The variation in milk yield from lactation to lactation may be attributed to variations in the environment prevailing at the time the observation is made.

A goal of animal breeders has been to estimate effects of environment and correct the cow's production for those effects. For dairy cattle, it is common to correct records for age, times milked per day, days in milk, and season of freshening. This process corrects or standardize to the average of the whole population.

Each adjusted or corrected record can be considered to represent the real ability of the cow under those "standard conditions" plus or minus some random error. The corrected record of the same cow in the next lactation has the same real ability of the animal plus or minus another error for environmental conditions. As pointed out by Lush (1945):

> "So far as temporary environmental conditions are concerned, these errors remaining in the corrected records will be independent of each other. Hence if all the records of the animal are averaged together, some of these will have positive errors, and others will have negative errors which will tend to cancel the positive ones."

The average lifetime production of a cow can be expressed as:

1

Ave

w

The

anim

The

more

vali

assum

anima

lacta

ʹ

!

ʹ

ʹ

s

c

$$\text{Average Lifetime Production} = \begin{array}{c}\text{Real ability}\\\text{of the animal}\end{array} + \text{SUM (errors)}/n$$

where: SUM is the summation of all n errors.

The average of n observations is the real ability of the animal and the average of the errors that did not cancel. The amount of error in this average becomes smaller with more observations if errors were really random.

Animal breeders have considered this assumption to be valid and random nature of the errors is a standard assumption in most of the research in animal breeding.

A linear model for estimating breeding value of dairy animals and predicting future milk production from repeated lactations follows:

$$Y_{ijkn} = hys_i + s_j + c_{ijk} + e_{ijkn}$$

where:

Y_{ijkn} is the milk production for the nth lactation of the kth cow in the ith herd-year-season sired by the jth sire,

hys_i is the fixed effect of the ith herd-year-season in which this lactation was initiated,

s_j is the random effect of the jth sire,

c_{ijk} is the random effect of kth cow, daughter of the jth sire in the ith herd,

From

c_{ij}

with

perm

temp

be ra

resid

Ve a

a_{ijkn}

real

lifet

for a

Yield

T

over t

that

that w

lactat

lactat

incorp

e_{ijkm} is the residual or random error.

From genetic theory, the cow effect in a given lactation, c_{ijkn}, can be expressed as:

$$c_{ijkn} = a_{ijkn} + d_{ijkn}$$

with a_{ijkn} being viewed as the genetic and unseparable permanent environmental contribution and d_{ijkn} as a temporary environmental contribution.

The s_j sire, the a_{ijkn} and, d_{ijkn} effects are taken to be random having zero means and variances Vs, Va, Vd. The residual effects e_{ijkn} have zero mean and common variance Ve and are uncorrelated with each other and with s_j and a_{ijkn}.

Inherent to this model are the assumptions that a cow's real producing ability remains constant over her productive lifetime, that the variance of the milk yields is constant for all lactations and the covariances for all pairs of milk yields are constant.

The assumptions that the cow's effect remains the same over time has been challenged. Harville (1979) pointed out that a cow might incur an ailment, such as acute mastitis that would affect adversely her producing ability in the lactation of initial occurence and in all sequent lactations. He suggested that more flexibility could be incorporated into the model by letting the temporary

env

lac

proc

Manso

estima

design

Their

assump

not ap

They p

of lact

tions c

environmental part of the cow effect vary from lactation to lactation in accordance with a first order autoregressive process, i.e., by replacing d_{ijkn} by d_{ijknt}:

$$d_{ijknt} = \emptyset\, d_{ijkn,t-1} + v_{ijknt}$$

where:

$\emptyset$ = autogregressive coefficient

$d_{ijkn,t-1}$ = temporary environmental effect in previous observations

v_{ijknt} = uncorrelated random variable with zero mean and variance Vv.

Mansour, Nordheim, and Rutledge (1985) presented a model to estimate variance components in a repeated measurements design, assuming non-stationary autoregressive errors. Their rationale was that in many sets of data the assumptions of constant variance and constant covariance do not appear warranted.

They presented as an example, the variance-covariance matrix of lactational yield of milk fat for the first three lactations of 11,613 cows reported by Butcher and Freeman (1969):

$$\begin{vmatrix} 6398 & 3322 & 2623 \\ 3323 & 7003 & 3514 \\ 2623 & 3514 & 7467 \end{vmatrix}.$$

Note

increa

the de

relax

A

y

w

y

M

a

t

e

In this

factors

factors

Fo

follows

e_{ij}

Note the increasing value along the principal diagonal, the increasing value 3323, 3514 on the first minor diagonal, and the decreasing value of the covariances in the first row.

Mansour, Nordheim, and Rutledge (1985) proposed to relax the assumption of independence of the error terms.

An example is a two way mixed linear model:

$$y_{ij} = M + a_i + t_j + e_{ij}$$

where:

y_{ij} is the milk yield recorded for the ith subject at the jth time,

M   is the overall fixed mean,

a_i is the random effect of the ith animal,

t_j is the fixed effect of the jth time and

e_{ij} is a random error term,

       e_{ij} and a_i are independently and identically distributed (i.i.d.) with $N(0, Ve)$ and $N(0, Va)$.

In this model Ve represents the temporary environmental factors and Va the genetics and permanent environmental factors.

For the same conditions on a_i but now assuming that e_{ij} follows a first order autoregressive model:

$$e_{ij} = \emptyset \, e_{i(j-1)} + v_{ij}$$

where:

$e_{i(j-1)}$ is error term in time $j-1$,

v_{ij} is uncorrelated random variable i.i.d. $N(0, Vv)$,

$\emptyset$ is autoregressive coefficient.

The symmetric variance-covariance structure of three milk yields associated with the model is:

$$\left|\begin{array}{ccc} Va+Ve & Va+\emptyset Ve & Va+\emptyset^2 Ve \\ & Va+(1+\emptyset^2)Ve & Va+\emptyset(1+\emptyset^2)Ve \\ Symmetric & & Va+(1+\emptyset^2+\emptyset^4)Ve \end{array}\right|.$$

With simulated data, the authors calculated maximum likelihood estimates of $\hat{V}a$, $\hat{V}e$, and $\hat{\emptyset}$. They found substancial improvement of the autoregressive model in the Butcher and Freeman (1969) data over the model that did not use autoregressive statistics. These findings could have a biological justification. The authors' arguments for the positive association among errors are: First, if a cow initially produces an unusually good or bad yield, it would be natural to think that her caretaker would have a high or low expectation of her ability and would provide

sequent management, accordingly. Although a portion of the initial yield is due to genetic and permanent environmental effects, when Ve is large relative to Va, the major impact of the differential management would be reflected better by autoregression. A pathogenic infection carried over from one lactation to the next might cause a low production in both lactations.

. Rothchild and Henderson (1979b) also relaxed the assumption of error covariance being set to zero. They explained that in a sire evaluation model with no cow component, to estimate variances and covariances for first and second lactations an error covariance must be included because the same cow has both records, and covariances between errors must exist.

If errors are correlated and if this correlation has a first order autoregressive pattern, there would be the potential to use recursive estimation techniques in the sire evaluation for later lactations case. These techniques are such that it is unnecessary to store old data to process new data. Also, more accurate estimates of the real ability of the cow could be obtained from these relationships.

Objectives of this study are:

--Estimate the error variance-covariance structure of a series of three lactations.

--Test the assumptions commonly made by animal breeders that there is homogeneous error variance and constant covariance between each pair of lactations.

--Examine the hypothesis presented by Harville (1979) and Mansour, Nordheim, and Rutledge (1985) that errors might follow a first order non-stationary autoregressive process.

Because of the computational simplicity of the following tasks, they too were performed:

--Estimate sire variance components for each lactation.

--Estimate heritability of milk yield for each lactation.

--Estimate sire solutions for each lactation.

--Rank the sires for each lactation.

--Calculate the rank correlation among sires.

II LITERATURE REVIEW

II.1 Studies on Later Lactations

II.1.1 Overview: First vs. Later Lactations in Sire Evaluation

Milk yield is and will continue to be the most important trait selected for in dairy cattle. Currently, dairy cattle improvement depends on the emphasis placed upon increasing milk yield.

Cows are ranked by Dairy Herd Improvement Associations according to criteria such as actual production, production ability, or cow index. Bulls are ranked on milk yield of their daughters. These methods have excelled for selecting those cows and bulls to maximize milk output of the herd.

Sire evaluation methods have evolved rapidly. Up to 25 years ago most methods of evaluating bulls used dam-daughter comparisons. The spread of artificial insemination, which gave the possiblity of having daughters of a bull in many herds, facilitated the introduction of methods comparing daughters with their herdmates. Onset of the computer made it possible to implement recent advances of variance component estimation on which most sire evaluation methods depend. One important development has been the introduction of the Best Linear Unbiased Prediction (BLUP) by Henderson (1975). This methodology could compensate for the effect of

selection and relationships between sires.

Because milk yield of the dairy cow is measured through her productive life, several records of a daughter could be used to evaluate a dairy bull. The first, second or any or all the lactations of daughters would be adequate to estimate the genetic value. However, developers of each sire evaluation system have chosen to use one lactation over the others. In the progeny test of dairy bulls, selection decisions are based principally on production by daughters in first lactation. The rationale behind the use of first lactation was summarized by Cassell and McDaniel (1983), who said:

> "Yield of first lactation offers advantages to
> dairy breeding researchers. Results
> are available earlier, and measurements exist
> for more cows than from later records. Extra-
> neous sources of variation such as injury,
> days dry following lactation, and preferential
> treatment are less likely to influence yield
> of first lactation."

Mao (1982) pointed out other arguments for the use of first lactation records.

> "-that the first lactation is an adequate
>   indicator of a cow's lifetime performance;
> -that the computation is much simpler without
>   having had to include a cow effect in the
>   model;
> -that the supposition of different genes
>   and genic actions for each lactation can be
>   avoided;
> -that the assumption of no sire by age of cow
>   interaction can be avoided;
> -that the potential bias due to selection can

>be also avoided, since not all cows are
>allowed by farmers to continue for more lacta-
>tions, and those who did make later lactations
>records are superior for one reason or the
>other to those cows that were culled; and
>-that the inclusion of all lactations
>records would only increase the accuracy of
>evaluation on those bulls who already had
>plenty of daughters, but the accuracy of their
>evaluation is of primary interest."

There are advantages for using later lactations for sire evaluation. Shaeffer (1982) summarized some of these:

>"-increase the number of records;
>-increase the number of records per
>fixed effect e.g. per like herd-year-seasons;
>-reduces the standard error of prediction for sire"

Let us document these arguments. An implicit condition for using first lactations is that the same genes influence production in first and later lactations. A genetic correlation less than unity would mean that some genetic control of later lactations is independent of genetic merit for yield of first lactation. The range of estimates of genetic correlation between first and second lactations is from .75 to .92 (Cassell and McDaniel, 1983).

The inclusion of later lactations in sire evaluations systems has some advantages over first lactation evaluation systems. One of the advantages of the use of all lactations is the increase in accuracy in evaluation of sires for milk yield. Ufford et al. (1979) in a study comparing first versus all lactations by Best Linear Unbiased Prediction (BLUP) found that the use of all lactation records decreased

the variance of prediction error of sire solutions so that 15 daughters per sire with all lactations gave accuracy of sire values, equivalent to 25 daughters with only first lactations. He concluded that the genetic progress per year from selection of bulls to sire daughters would be expected to be 10 to 15% greater for all lactation records than for only first lactation records.

A major criticism of evaluation systems employing first lactation records as estimates of performance is that the working life of dairy cattle may be reduced by placement of too much emphasis on early lactations. If a cow lasted in the herd for several lactations, it means that on several occasions the milk producer has assessed her and found her satisfactory. The length of life of a cow in a herd, or longevity, does not only depend upon milk yield but on other characteristics including reproductive efficiency and health.

Robertson and Barker (1966) said:

>"We can then look on longevity as the ulti-
>mate character in the evaluation of different
>ways of selecting dairy sires, not for its own
>sake but as a numerical indicator of the
>judgements of milk producers."

Lifetime performance of the cow is a determining factor of her profitability in a dairy operation, and later records may contain useful information relative to lifetime profitability. A positive correlation between yield of first

la

su

li

pu

ti

by

ge

se

wi

ha

se

in

an

un

en

in

a

mi

pro

pro

hig

ver

stu

dau

daug

lactation and traits of lifetime performance such as
survival, length of herdlife, total performance, total
lifetime yield and lifetime profitability are found in all
published work concerning these matters, the remaining ques-
tions to be answered are: how accurate are they predicted
by yield of first lactation and will optimum progress for
genetic gain in economically important traits be reached by
selecting only on first yield? (Cassell and McDaniel, 1983).

There are just as many literature references concerned
with disadvantages of later records in sire evaluation. Per-
haps a valid criticism of later lactation is that cows with
second or later lactations have been selected. This
introduces selection bias into the sire evaluation. Wickham
and Henderson (1977) suggested that large biases seem
unlikely because a small percentage of cows is culled at the
end of the first lactation. Some other authors have
investigated whether survivors of first lactations are not
a random sample of all first lactation cows with respect to
milk yield. Van Vleck and Henderson (1963) approached this
problem by studying daughters of sires in six classes of
production (based on their proofs). They found that a
higher proportion of daughters of lower production bulls
were culled at the end of the first lactation. In their
study, cows not having second records were 11 to 18% of all
daughters of high sires versus 27 to 30% of all the
daughters of low production sires. Keown et al. (1976)

g
s
d
r
g
h
a
H
f
f
he
fo

He

studying effects of selection bias on sire evaluation in 1969 records, found that those cows culled following second records exceeded those culled after first by 675 kg in first lactation yield. A similar positive relationship between first lactation deviation and sequent performance and survival was found by Hinks (1966). He reported that whereas the starters had a positive deviation from herd average of +51, the survivors to second lactation had a average deviation of +363. These results indicate that survivors of first lactations are not a random sample of all first lactation cows with respect to milk yield.

Another difficulty for the use of later records is genetic differences in rate of maturity. Hargrove (1974) studying the rate of maturity of dairy females, examined differences between first and later lactations from 19,000 records. Deviations were not subject to influences of genetic trend and selection on first yield. He reported heritabilities of differences in rate of maturity of .1, approximately one-fourth the heritability for milk yield. Hickman and Henderson (1955) studied the milk yield increase from first to second lactation records and age at first freshening for data from 4,000 cows. They also found the heritability for rate of maturity was one-third to one-fourth the heritability of milk yield.

Hillers and Freeman (1965) took another approach to measure differences between sires in rate of maturity. They

used the regression within actual production on age at first calving as a measurement of rate of maturity. Data were from 76 California herds with freshenings between 23 and 35 months of age. Sire differences in rate of maturity were significant. A general conclusion of this study indicates that differences among sires in rate of maturity are real, but heritabilities are low enough to discourage direct selection. Freeman (1973) in a review article on age adjustment of production records concluded that sire differences in rate of maturity imply that age factors to adjust records to some base by a common curve may obscure genetic differences between sires. This will cause unequal ranking for sires on first and later lactation. Cassell and McDaniel (1983) concluded to the discussion on this matter:

> "Thus we face the problem of knowing that age adjustment factors may obscure genetic diffe-rences yet to make no adjustment for age in sire evaluation places such evaluation at the mercy of progeny age distribution. Effective removal of only those effects of age common to all cows across all herds and progeny groups is a challenge for future work."

There are sire evaluation systems that use all lacta-tions (USDA-Modified Contemporary Comparison), only first lactations (Canadian Record of Performance, Northeastern Sire Evaluation Program) and a combination of two separate evaluations, for first and first and second combined (Israel sire evaluation system).

Although it may be safer to avoid potential problems of

including later lactation records in sire evaluation, there are advantages such as cow evaluations generated as a by-product of all records of a cow in a sire evaluation system. The use of all records in sire evaluation may be even necessary for progeny testing programs of small populations or populations with fewer dairy records, as for less developed countries like Mexico (McDowell, 1983).

II

of s

lact

vali

usin

daugh

and c

betwe

sire

the s

selec

They

proce

Holst

were

The av

in re

evalu

Perfor

(1978)

15 ∎

II.1.2 Studies Comparing Sire Values Using First and Later Lactations

Considerable effort has been directed to study results of sire evaluation methods using first, second, or later lactations. One purpose of these studies has been to validate the assumption of equality of the ranking of bulls using different lactations.

Tomaszewski et al. (1975) from data from two samples of daughters of 133 Holstein bulls for only first lactations and only second lactations, reported correlations of .64 between sire values for milk yields. They also found that a sire value for first lactation can predict with 85% accuracy the sire value for second lactation.

Wickham and Henderson (1977) studied the effect of selection in evaluating sires by second lactation data. They developed a BLUP procedure to estimate biases. This procedure was used to obtain estimates of biases for 1109 Holstein sires. Evaluations for first and second lactations were estimated with a set of equations for a mixed model. The authors found that the magnitude of the bias was small in relation to the rank correlation of .8 between sire evaluations by first and second lactation.

First, second and third lactation Ontario Records of Performance were analyzed separately by Nicholson et al. (1978) to evaluate 246 Holstein sires. These authors also

ana

witl

metl

yie.

rec

lac

incr

The

lact

Howe

Nich

lact

and

diff

tion

leas

foun

sugge

becau

lact

sires

cow/y

lacta

Casse

analyzed second and third records, correcting for selection
with the method suggested by Lush and Shrode (1950). This
method consisted of subtracting the average of the first
yield to the yield of those that made a second lactation
record times the regression of second lactation on first
lactation yield. This adjustment for selection tends to
increase the variation known to be reduced by selection.
The correlation among proofs was higher for second and third
lactations adjusted for selection. This range was .71 to .75.
However, there were differences in the ranking of the sires.
Nicholson et al. (1978) did not see any need to use later
lactation for sire evaluation.

Bar-Anar (1975) studied the relationship between first
and second lactations using the Israeli method (cumulative
difference) and data representing 106 sires with two lacta-
tions from each daughters. For each sire there were at
least 60 daughters in each lactation group. The author
found the correlation between tests to be of .76. He
suggested that for the sire test program in Israel and
because no extra cost is incurred by evaluating the second
lactation a gain of 7% in accuracy in the estimation of
sires will give an estimated yield increase of 7.5 kg milk
cow/year in daughters of proven bulls.

Modified Contemporary Comparisons for first and second
lactations in the same and different herds were analyzed by
Cassell et al. (1983a). Data were from 200 Holstein sires

with 500 or more daughters in each data set. They also used an adjustment to account for effects of selection on later lactations for evaluations by the same method used by Nicholson (1978) and originally proposed by Lush and Shrode (1950). Correlations between first lactation evaluations in one independent data set and second lactation evaluations in the other were .87 and .84. The correlation between first and adjusted second evaluation was .87 for both independent data sets. First lactation evaluations accounted for only about 75% of the variation in second record evaluation convinced the authors of a need for some form of later record sire evaluation at least for selection of sires of sons.

Lofgren et al. (1983) studied effects of culling on sire evaluation using the same data of 677,800 daughters of 200 widely used Holstein sires, those used by Cassell et al. (1983a). From these data, 10, 20, and 30% of second records were eliminated based upon least yield of milk in first lactation. Evaluations by BLUP of sires were obtained separately for both records and for culled groups. They used a model that included effects of cow when more than one lactation was used. Second lactations were adjusted for selection following the method of Lush and Shrode (1950). All correlations between evaluations of first and both were .95 and were unaffected by increased culling. Because these data are the same as those used by Cassell et al. (1983a),

authors were able to compare two methods of sire evalua-
tions. Evaluations by BLUP and MCC procedures were affected
similarly by culling. In both cases, culling reduced
standard deviations of evaluations by second records and re-
duced correlations between evaluations by first and second
records. The two sire evaluation procedures also ranked
bulls nearly identically.

Cassell et al. (1983b) in another study compared the
impact of culling or selection on sire evaluation by three
mixed model procedures: 1) single trait evaluation of first
and second lactations, 2) evaluation of both lactations
together including a random component for cow effect, and 3)
a multiple trait procedure where first and second lactation
evaluations were calculated simultaneously. Data were first
and second lactation records from the dairy herd at North
Carolina State University where each of 130 females produced
second records regardless of first lactation yield. The
relationship matrix of the 45 sires represented in the data
was included in all models. Culling was simulated at
intensities of 10, 20, and 30% on deviations of first
lactation from population means. The main conclusion of
this study favor the use of the multiple trait approach
because of its reduction of standard errors of prediction.
The use of both lactations, according to the authors,
appears to be a reasonable alternative for sire evaluation
until multiple trait procedures become computationally

feasible.

One can summarize the results of these studies comparing the sire evaluations using first and later lactation. The correlation among them is high but not perfect. Rankings of sires change from one evaluation to the other. Sire values for first lactation account for only from 75 to 85% of the sire evaluations for the second indicating a need for the evaluation for later lactations at least for sires of sons. These studies also indicate that selection bias cannot be completely responsible for differences in sire evaluations.

II.1

tere

first

condu

param

(1960

milk

lacta

third

that b

unity,

lactat

Ba

animals

nested

.35, .24,

lactati

ones av

average

But

tion rec

pairs of

first thi

were obta

II.1.3. Genetic Parameters of Milk Yield for First and Later Lactations.

Emphasis on selection on first lactation yield has centered attention on the same genes influencing production in first and later lactations. A number of authors have conducted studies with the purpose of estimating genetic parameters for first, second, and later lactations. Freeman (1960) found by daughter-dam regression heritabilities for milk of .36, .24, and .26 for first, second, and third lactations. Genetic correlations among first, second, and third lactations were between .93 and .98. He suggested that because these correlations are consistently less than unity, different sets of genes influence milk in different lactations.

Barker and Robertson (1966) using records of 10,965 animals, 80 bulls, and a model with all random effects nested within fixed effects reported heritabilities of .35, .24, and .23 for milk yield of first, second, and third lactations. Genetic correlations of first yield with later ones averaged .80 but between second and third yields averaged .91.

Butcher and Freeman (1968) using 12,500 Holstein lactation records estimated the relationship between various pairs of lactations of the same cow and heritabilities of first through fourth lactations. Heritability estimates were obtained by intrasire regression of daughter on dam.

Heritabilities were .37, .25, .35, and .40 for first, second, third, and fourth lactations. Repeatabilities estimated by intracow correlation were from .54 to .50 for the first through fifth lactation.

The same authors in another study (Butcher and Freeman, 1969) used the same data to estimate genetic parameters and relationships between pairs of lactation by five procedures.

1) Intraclass correlations for all cows that have the first record of the pair, regardless of whether the second record was present.

2) Intraclass correlations using only those cows with both records of the pair under consideration.

3) The regression of the second record of the pair on the first record of the pair.

4) A Maximum Likelihood procedure described by Curnow (1961).

5) A procedure to estimate parameters free of the effects of selection on the independent variable developed by the author.

The maximum likelihood procedure (Curnow, 1961) works as if the only factor that determines if a cow has a second record is the magnitude of her first one, with no culling, variances of all records are equal, and all first records are available, whether the cow has a second record.

The method derived by Butcher and Freeman (1969) constructs estimates of variances and covariances free of effects of selection by regression techniques, and then these relationships are computed by simple correlations.

we

inc

The

par

to

pro

acc

howe

effe

proc

vari

gene

lact

Hende

and 3

the M

and Qu

covar

trait

ance b

includ

for fir

include

therefo

Ranges for heritabilities for first to third lactations were .17 to .39, .11 to .35 and .11 to .40. These results indicate differences in methods of computing heritabilities. The authors suggested that the method used to estimate these parameters should be determined by how well the data conform to conditions necessary for unbiased estimates. The procedure derived to remove effects of selection seemed to accomplish the desired results. The authors suggested, however, that to establish firmly that the procedure removes effects of selection would have to be done by simulation procedures.

Perhaps the most comprehensive attempt to investigate variance and covariance components, heritabilities, and genetic and phenotypic correlations between first and second lactations milk records was the study of Rothschild and Henderson (1979a). They used data of 423,314 first records and 339,182 selected second records. The authors extended the Maximum Likelihood (ML) algorithm presented by Henderson and Quaas (1977) to estimate sire and error variance and covariances for the two trait model when records on one trait were missing. The authors included the error covariance because in a sire evaluation model, no cow component is included. Therefore, to estimate variances and covariances for first and second lactations an error covariance must be included because the same cow has both records, and therefore covariances between errors must exist. From

result

mates

select

practi

requir

hours

ties f

correl

authors

first

T

correl

to cul

daught

Maximu

trait

editin

third

herd-y

effect

in thi

results with simulated data the authors found that ML esti-
mates of variances and covariances were unaffected by 50%
selection within fixed effects. The authors discussed the
practicality of ML estimation, and they mentioned that time
required for one round of iteration was between 28 and 30
hours for an IBM 370-138 with 512 K of memory. Heritabili-
ties for first and second records were .41 and .35. Genetic
correlation between first and second records was .92. The
authors estimate matrix of error variance-covariance for the
first two lactations is as follows:

$$
\begin{vmatrix}
900,561 & 507,430 \\
507,430 & 1,032,391
\end{vmatrix} .
$$

Tong et al. (1979) estimated heritabilities and genetic
correlations for the three lactations from records subject
to culling. Lactation records on 13,544 cows representing
daughters of 90 sires. The authors used a Restricted
Maximum Likelihood (REML) procedure derived from a multiple
trait model discribed by Schaeffer et al. (1978). After
editing, the data yielded 5,036 second lactations and 1,500
third lactations. Their model contained fixed effects of
herd-year-season and age of cow at calving and random
effects of sire and error. Error covariances were ignored
in this model. After ten rounds of iterations

heritabilities of milk yield for first, second, and third lactations were .26, .19 and .17. The genetic correlations were between first and second .89, first and third .85, and second and third .89. This study shows that genetic correlations between adjacent lactations are higher than between nonadjacent lactations.

Meyer (1983) developed a REML procedure for estimation of genetic parameters for later lactations of dairy cattle. She derived an algorithm for a two way mixed model allowing for missing observations while taking account of residual covariances.

The author presented an example of data consisting of 2,247 first lactations, 1701 second lactations, and 1,186 third lactations of British Friesians. With small herd-year-season (HYS) subclasses, the data structure was improved by adding records for daughters of proven sires calving in the same herd-year-season. As variation between proven sires is expected to be reduced by selection of proved sires, these sires were treated as fixed effects (i.e., their daughters contributed information to the components within but not between sires). The model was also extended further by adding four regression coefficients per lactation to correct for the systematic effects of lactation length (linear), month of calving within season (linear), and calving age (linear and quadratic). Heritabilities of milk yields for first, second, and third

lactat

correl

betwee

The au

merely

variat

mate o

lactat

Th

al. (19

Likelil

account

used by

Because

not app

covaria

varianc

Th

(1979)

compone

of free

bias fr

lactations were .33, .34, and .28. Estimates of genetic correlations were 1. between first and second lactation .98 between second and third, and.83 between first and third. The author mentioned that the data set considered was chosen merely to illustrate the procedure; therefore, sampling variation was too large to allow inferences. Authors' estimate of error variance covariance matrix for the three lactations is as follows:

$$\begin{vmatrix} 557.31 & 371.67 & 405.91 \\ 371.67 & 928.51 & 565.45 \\ 405.91 & 565.45 & 1,008.67 \end{vmatrix}.$$

The papers of Rothschild and Henderson (1979), Tong et al. (1979), and Meyer (1983) are attempts to apply Maximum Likelihood and Restricted Maximum Likelihood procedures to account for bias from cow selection. The REML algorithm used by Tong et al. was for residual covariances of zero. Because they are repeated measures on the same cow, this is not appropriate. The authors argued however, that the error covariances are likely to be small relative to error variances.

The algorithm derived by Rothschild and Henderson (1979) is a ML procedure, which means that variance components estimates were biased upwards by loss of degrees of freedom from estimating herd-year-season effects. This bias from ML is greatest for error components because the

denominator is the number of observations. For REML the
degrees of freedom are the number of observations minus the
number of fixed effects that is used to estimate variance
components.

Estimation of genetic parameters for later lactations
has been an important subject of study for animal breeders.
A purpose has been to try to demonstrate that different
lactations are genetically the same trait. Failing to
demonstrate this would justify the use of first only records
or multiple trait procedures for sire evaluation for milk
yield.

Estimation procedures often are subject to bias from
selection because cows that remain in the herd are likely to
have been selected for higher yields. Later lactations have
been used as   typical example for analysis of data under
selection (Henderson, 1975).

The genetic correlation between lactations is greater
than.80 and genetic correlations between adjacent
lactations are higher than between nonadjacent lactations.
Several authors have speculated that the significant
difference of the genetic correlation between lactations may
have something to do with a interaction of sire by age or
sire by parity. Sire variance across all three lactations
also differed (Tong et al.,1979). This may indicate that
the sire effect is different for each lactation. The
difference of genetic correlations could be explained by

differ

be use

is ger

H

for la

possib

Mo

mental

and man

records

differe

differences in sire effect in rate of maturity. This could
be used to support that milk yield in different lactations
is genetically the same trait.

Heritabilities for milk yield are consistently lower
for later lactations. Butcher and Freeman (1968), listed
possible explanations for these differences as they said:

> "a) If all genes have equal effects, first
> lactations are controlled by more pairs of
> genes than second lactation or, if the same
> number or genes control both lactations, they
> have larger effects on first lactation.
>
> b) The presence of a genetic maternal effect
> that gradually decreases in importance could
> cause the estimate of heritability of first
> lactation to be larger.
> c) The presence of constant genetic effects
> on second lactations would lower estimates of
> heritability of second lactations."

Most authors seem to favor the opinion that environ-
mental factors in the lifetime of the cow, such as diseases
and management, will increase variation of later
records.This could obscure genetic variation resulting in
different estimates of genetic parameters.

II.2.

measu

may

are u

once c

treatm

invest

treatm

a sequ

includ

designs

In

methods

rationa

pointed

a block

(2) to t

study tr

Time

which ari

II.2 Analysis of Sequential or Repeated Observations

II.2.1 Repeated Measurements Experiments

It is practice in a wide range of experiments to repeat measurements over time on the same experimental units which may be a plot, animal, or plant. Experimental designs which are used in these situations are:

(1) Those in which experimental units are treated once only during the experiment or receive the same treatment repeatedly. These include those experiments that investigate growth curves and profile studies like block and treatment designs.

(2) Those in which each experimental unit receives a sequence of treatments over successive times. These include rotational trials, change over and superimposed designs.

In animal experiments, analysis by repeated measurement methods has some advantages. Gill (1978) reviewed the rational for performing experiments in this manner. He pointed out three major reasons, (1) to use each subject as a block to conserve resources and reduce experimental error, (2) to test for residual effects of treatment, and (3) to study trends of individual responses to treatments.

Time is a factor in such experiments and the issue which arises relates to possible correlation between

successive observations. Gill and Hafs (1971) referring to
this problem said:

> "the structure of underlying causes of
> variation often is over-simplified in
> analysis of such data. Simplifications
> usually fall into one or two categories;
> 1) Analyzing factorial experi ments as if
> completely randomized, when the structure
> is really a split plot and 2) Ignoring the
> correlations of errors induced by repeated
> measurements."

Recognition that measurements within animals are likely
to vary less than those among animals leads to the study of
the random effect of individual subjects. This
characterizes the split plot design where the error term no
longer represents variation among homogeneous subjects
treated alike but measures variation within subjects treated
alike.
Constancy of error variance and independence of errors,
the underlying assumptions for univariate analysis of
variance, often are violated in the analysis of repeated
measurements.

In the analysis of repeated observations, error terms
from different times are correlated. When this occurs, it
is said that the error terms are autocorrelated or serially
correlated. This happens when the errors associated with
observations in a given time carry over into future times.

When error terms are functionally related to the mean,
for example, periods with larger means also have larger

varian

will h

over c

occurs

T

ment c

are tr

monitor

Le

$$
\begin{vmatrix}
y_1 \\
y_2 \\
\cdot \\
\cdot \\
\cdot \\
y_n
\end{vmatrix}
$$

whe

least sq

more expl

variances, and constant error variance or homoscedasticity will be unreasonable. When the variance is not constant over observations, it is said that heteroscedasticity occurs.

To illustrate these conditions in the repeated measurement case, let us examine a simple model in which n animals are treated alike and each animal is measured r times to monitor changes of variable y.

Let us write a model describing this relatonships.

$$
\begin{vmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{vmatrix} = \begin{vmatrix} 1 & x_{1,t} & x_{1,t-1} & x_{1,t-2} & \cdots & x_{1,r} \\ 1 & x_{2,t} & x_{2,t-1} & x_{2,t-2} & \cdots & x_{2,r} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{n,t} & x_{n,t-1} & x_{n,t-2} & \cdots & x_{n,r} \end{vmatrix} \begin{vmatrix} b_0 \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_r \end{vmatrix} + \begin{vmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{vmatrix}
$$

where:

   y = column vector of observations for the dependent variable y,

   X = matrix giving r repeated measurements of variable x, the first column of 1's representing the intercept term,

   b = column vector of unknown parameters $b_1, \ldots, b_r$,

   e = column vector of n error terms associated with each observation.

The previously mentioned underlying assumption of least squares methods on the error term can be represented more explicitly by deriving the error variance by rules of

expecta

We

Pe

E =

expectations.

We can write:

$$E(e\ e') = E \begin{vmatrix} e_1 \\ e_2 \\ . \\ . \\ e_n \end{vmatrix} (e_1,\ e_2\ .\ .\ .\ e_n).$$

Performing the multiplication we obtain:

$$E = \begin{vmatrix} e_1\ e_1 & e_1\ e_2\ .\ .\ .\ e_1\ e_n \\ e_2\ e_1 & e_2\ e_2\ .\ .\ .\ e_2\ e_n \\ . & .\quad\quad .\quad . \\ . & .\quad\quad .\quad . \\ e_n\ e_1 & e_n\ e_2 \quad\quad e_n\ e_n \end{vmatrix}.$$

Ap

the pre

E

Wit

otherwis

E(e

bec

E(e

wh

Thi

Applying the expectation operator E to each element of
the preceding matrix we obtain:

$$
E(e\ e') = \begin{vmatrix}
E(e_1e_1) & E(e_1e_2) & \cdots & E(e_1e_n) \\
E(e_2e_1) & E(e_2e_2) & \cdots & E(e_2e_n) \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
E(e_ne_1) & E(e_ne_2) & \cdots & E(e_ne_n)
\end{vmatrix}.
$$

With homoscedasticity and no correlation serial or
otherwise, the matrix reduces to:

$$
E(e\ e') \begin{vmatrix}
Ve & 0 & \cdots & & 0 \\
0 & Ve & \cdots & & 0 \\
\cdot & \cdot & \cdots & & \cdot \\
\cdot & \cdot & \cdots & & \cdot \\
\cdot & \cdot & \cdots & & \cdot \\
0 & 0 & \cdots & & Ve
\end{vmatrix}
$$

because

$E(e_te_t) = Ve$ and $E(e_t\ e_r) = cov\ e_t\ e_r = 0$

 where $r \neq t$.

This is a homogeneous variance-covariance matrix. In

other w

replica

observa

terms a

E(

Ve

To

situatio

from dist

correlati

E(e

other words, each error term, observed across all possible
replications, has the same variance regardless of the
observations and there is no correlation among the error
terms associated with different observations.

E(e e') also can be written by an identity matrix:

$$
Ve \begin{vmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & & \ldots & 1 \end{vmatrix} = VeI.
$$

To visualize the heteroscedastic (heterogeneous)
situation where the error term for each observation is drawn
from distributions with different variances but without
correlation among error terms, we write E(e e') as:

$$
E(e\ e') = \begin{vmatrix} Ve_1 & 0 & 0 & \ldots & 0 \\ 0 & Ve_2 & & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & & \ldots & Ve_n \end{vmatrix}.
$$

T

matrix

the er

B

e_t as a

e_t

Thi

beginnin

first ob

becomes

$(1 + \beta$

The

be obtai

$E(e^2$

To illustrate a heterogeneous variance-covariance matrix let the error term in period t, e_t be a function of the error term in period t-1 plus a random component v_t

$$e_t = \emptyset\ e_{t-1} + v_t\ .$$

By extending this equation backwards one can write each e_t as a function of only previous v_t

$$e_t = \emptyset^2\ v_{t-2} + \emptyset\ v_{t-1} + v_t$$

$$= \emptyset^3\ v_{t-3} + \emptyset^2\ v_{t-2} + \emptyset\ v_{t-1} + v_t$$

$$= \text{SUM}(i= 1,t)\ \ \emptyset^i\ v_{t-i}\ .$$

This is if that the series extends prior to the beginning of our sample of observations, so that even our first observation corresponds to a high value for t. As t becomes large with $\emptyset<1$, the series becomes:

$$(1 + \emptyset^2 + \emptyset^3 + \ .\ .\ .)\ v_t = 1/(1-\emptyset^2)\ v_t$$

The variance of the error term in autocorrelation can be obtain from the expectation, this can be written:

$$E(e^2{}_t) = E\ (e^2{}_t) + \emptyset^2\ E(e^2{}_{t-1}) + \emptyset^4\ E(e^2{}_{t-2}) + \ .\ .\ .$$

$$= Ve\ (1 + \emptyset^2 + \emptyset^4 + \ .\ .\ .) = Ve/(1-\emptyset)$$

E

Th

expecte

Conse

correla

$$VCV =$$

The

ciated;

than a p

associat

process

before t

present c

of relat

What

current c

observati

repeated

Gill

$$E(e_t \, e'_{t-s}) = Ve\emptyset.$$

This expressions for $E(e_t^2)$ and $E(e \, e')$ imply that the expected correlation between e_t and e_{t-1} is $\emptyset$.

Consequently, $E(e_t \, e_{t-1})$ for first order serial correlation the variance covariance matrix could be written:

$$VCV = Ve \begin{vmatrix} 1 & \emptyset & \emptyset^2 & \cdots & \emptyset^{n-1} \\ \emptyset & 1 & \emptyset & \cdots & \emptyset^{n-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \emptyset^{n-1} & \emptyset^{n-2} & \emptyset^{n-3} & \cdots & 1 \end{vmatrix}.$$

There are several ways in which the errors can be associated; periods of this association also could extend more than a previous observation. The number of periods of this association is known as the order; i.e., a second order process assumes that observations made 2 and 1 periods before the present observation are correlated with the present observation. Later in this review some other types of relationships will be presented.

Whatever the extent or number of lags influencing the current observation, there is a correlation of repeated observations. This is a major problem in analyzing repeated measurements.

Gill and Hafs(1971) indicated that experience calls for

examin

decidi

univar

varian

typica

the im

illust

to spl

mals.

having

standa

procedu

Th

measure

basic a

indepen

variati

By this

were di

advant

"
t
l
c
v
H
m
t
i
j

Ho

examination of the variance covariance matrix before deciding the analytical approach because the validity of the univariate split-plot analysis depends on the homogenous variance-covariance matrix. The authors in presenting the typical split-plot case to animal scientists, pointed out the importance of random assigment of animals. They also illustrated the severe errors of interpretation if one fails to split residual errors into portions among and within animals. It seems to be a concensus that for balanced data having a homogeneous variance-covariance matrix the standard univariate analysis of variance is the best procedure for analysis of repeated measurements.

There are other procedures used to analyze repeated measurements. Multivariate procedures do not require the basic assumptions of homogeneity of variance and independence of errors because they deal with simultaneous variation of two or more variables (Sokal and Rohlf,1969). By this method repeated observations are treated as if they were different variables. Gill(1981) listed some of the advantages of multivariate procedures:

> "It offers opportunity to examine not only
> the original data but multivariate sets of
> linear functions (usually contrast) of
> observational units without requiring equal
> variances and covariances among these units."
> He also added: "The principal advantage of
> multivariate analysis is that it permits valid
> tests of the repeated factor(e.g. time) and its
> interactions with treatments without
> justifications."

However, Morrison (1967) demonstrated that unless the

seri

abso

than

recon

data

been

that

repea

there

are f

measu

combi

each

measu

corre

Ano

corre

time

al. (

avera

Box a

series

The a

of the

serial correlation between two periods exceeds .5 in absolute value, univariate procedures are more sensitive than the multivariate T^2 test (Hotelling's T). It has been recommended (Gill,1978) that lowly but uniformly correlated data should be analyzed by univariate methods.

Another drawback of the multivariate procedures has been pointed out by Gupta and Perlman (1974). They found that the power of the test declined as the number of repeated measurements in one animal (unit) increased; therefore, one has to be careful with inferences when there are fewer animals than sampling points.

There are other approaches to the analysis of repeated measurements, Gill (1979) presented a procedure that combines significance levels of correlated univariate tests, each performed on data from a different periods of measurement. This method is based in intratreatment correlations between data from all pairs of periods.

Another approach to deal with the problem of serial correlation is to adjust for serial correlation or to use time methods such as autoregression analysis. Chistian et al. (1978) opted to remove serial correlation by a moving average time series process of first order as described by Box and Jenkins (1970). The data generated by the time series analysis later were analyzed by conventional methods. The authors fail to provide theoretical support to the basis of their analysis. Gill (1981) mentioning this method

comment

value b

Ga.

study t

efficie

An

product

curves

paramet

those

was ass

correl

The au

were s

value

from o

data t

autoco

autore

betwee

transf

simila

square

analys

varian

ters."

commented that users of this procedure were skeptical of its value because of its low power.

Gallant and Goebel (1976) demostrated in a monte carlo study that the autoregressive estimators are the most efficient linear unbiased estimators in finite samples.

Anderson (1981) in a study of the _Tribolium_ egg production curves as a model for dairy cattle lactation curves used this approach. He compared estimates of parameters obtained with the autoregressive process with those obtained with standard regression. In this study, it was assumed that only three previous egg counts were correlated with the current count, i. e., only three lags. The autoregression coefficients of each of the three lags were solved by Yule-Walker equations. To determine the value of the autoregression coefficients, autocorrelation from ordinary least squares was estimated. The original data then were adjusted by the coefficients to remove autocorrelation. The author postulated that autoregression, which is based on the autocorrelation between ordinary least squares residuals may be considered a transformation of the data. Estimates of parameters were similar for each method (autoregressive and ordinary least squares). Therefore, the author concluded, "Autoregression analysis improved the accuracy of the estimates of error variance and not the accuracy of estimates the model parameters."

induce

cattle

to cor

measur

hetero

for a

heteroc

correct

split-p

way the

Kesner et al. (1981) in a study to show that estradiol induces preovulatory Luteinaizing Hormone (LH) surge in cattle, used a similar approach to that of Anderson (1981) to correct for autocorrelations caused by the repeated measurement nature of the experiment. Facing a heterogeneous variance-covariance matrix the authors opted for a logarithmic transformation to correct for heterogeneous variance and an autoregressive process to correct for serial correlation. Further, they performed a split-plot analysis with transformed data improving in this way the accuracy of the estimate of error.

st
observe
princip
researc
linear
standar
multiva
opinion
linear
models.
particu
the sam
time se
fields,
analysi
univari
correla
flexibl
For exa
for est
individ

It
change
be deri
measure

II.2.2 Analysis of Time Series Data

Studies in which the responses of each individal are
observed on two or more occasions represent one of the
principle research strategies in animal and medical
research. Much of the literature on specifying and fitting
linear models for serial measurements use methods based on
standard regression, analysis of variance (ANOVA) and
multivariate linear models (Ware, 1985). Harville's (1977)
opinion on this matter is that it is a mistake to think of
linear models only in terms of ordinary regression and ANOVA
models. To do so is to miss many potential applications. In
particular, not all the observations may have been taken at
the same time so that the observations are regarded best as
time series. Such time series data are common in many
fields, such as economics. Ware (1985) indicated that
analysis of serial measurements should be viewed as a
univariate regression analysis of responses with
correlated errors. Such a formulation suggests new and more
flexible approaches to modeling and estimation of parameters
For example, Wilson et al. (1981) presented the methodology
for estimation of biological parameters when the within
individual observations errors are autocorrelated.

It is timely now to discuss approaches to the study of
change or behavior of a variable throughout time. This can
be derived intuitively from the study of repeated
measurements. As an example, consider a series of

measu

be re

the h

may h

varia

to fa

expla

varial

varial

it is

variah

the ba

the va

T

called

observ

hormone

systema

the var

for the

its beh

its pas

future

The

in cases

determin

measurements of the variable y throughout time. This could be represented as a time function y(t), which may represent the historical performance of some biological variable; it may have moved up or down partly in response to other variables. However, much of its movement may have been due to factors that one simply may have not been able to explain. It may be difficult to relate to other biological variables, or data are not available for those explanatory variables that are believed to affect y(t). In this case, it is no longer possible to predict future movements in a variable by relating it to a set of other variables; instead the basis of prediction depends solely on past behavior of the variable and that variable alone.

The study of a single sequence of observations is called time series analysis. This is the study of repeated observations of a single variable, e.g., the study of a hormone during the estrous cycle. If some kind of systematic behavior like a trend or a cyclical pattern is in the variable to study, one can attempt to construct a model for the time series which does not offer an explanation for its behavior in terms of other variables but does replicate its past behavior in a way that might help to forecast its future behavior.

The choice of a time series model usually will result in cases where little information is known of the determinants of the variable of primary concern, and a

suf:

a tl

thro

move

out

the

the

simul

it as

diffe

that

is as

rando

is po

assoc

$y_1...y$

varia

distr

to al

distr

espec

is ne

sufficiently large amount of data is available to construct a time series of reasonable length (number of measurements throughout time).

A time series model accounts for patterns in past movements of a particular variable. Durbin (1960) pointed out the purposes for studyng time series was to 1) describe the data, 2) explain the behavior of the data, 3) forecast the behavior, 4) control the system, and 5) study simultaneous variation of several series.

A simple view of the time series model is to think of it as a sophisticated method of extrapolation; however, difference arises because time series analysis presumes that the series has been generated by a random process; it is assumed that each y_1, y_2, ... y_t in the series is drawn randomly from a probability distribution. In this way, it is possible to infer something about the probabilities associated with alternative future values of the series.

Theoretically, it is assumed that the observed series y_1...y_t is drawn from a set of jointly distributed random variables, i.e., that there exists some probability distribution function, $p(y_1...y_t)$ that assigns probabilities to all possible combinations of y_t.

Because complete specification of the probability distribution for a time series is almost always difficult, especially if the series has more than six data points, it is necessary to constuct a simplified model of the time

ser.

be a

acco

The

but

Nels

actu

know

the s

is ca.

const

consta

fixed

equati

past d

A

ging in

we have

Product

series which explains its randomness. For example, it may
be assumed that the $y_1,...,y_t$ are correlated with each other
according to a simple first order autoregressive process.
The actual distribution of y's may be much more complicated,
but this simple model may be a reasonable approximation.
Nelson (1973) talking about the problem of finding the
actual joint distribution concluded:

> "In practice, of course, we must first
> attempt to infer from the data what mechanism
> it is that generated the data. Thus, the past
> history of the time series is called upon to
> do double duty: First it must inform us about
> the particular mechanism which describes its
> evolution through time and second, it allows
> us to put that mechanism to use in forecasting
> the future"

Another characteristic of time series that one should
know is whether the underlying random process that generated
the series is invariant with respect to time. Such a series
is called stationary; this requires the series to have a
constant mean and to fluctuate about that mean with a
constant variance. This is because if the random process is
fixed in time, it is possible to model the process with an
equation with fixed coefficients that can be estimated from
past data.

A non-stationary series will be one that has been chan-
ging in average over time, presenting a trend; as an example
we have some economical variables like the Gross National
Product which has been growing steadily; for this reason

alone

in 193

P

few of

statio

encount

one or

or will

Th

differe

order o

stationa

w_t

where:

w_t

If a seri

series wo

differenc

$w_t =$

Time

alone its random properties in 1980 are different from those in 1933.

Pindyck and Rubinfeld (1981) mentioned that probably few of the time series that one meets in practice are stationary. However, many of the non-stationary time series encountered have the property that if they are differenced one or more times, the resulting series will be stationary or will be called homogeneous.

The number of times that the original series must be differenced before a stationary series results is called the order of homogeneity. A first order homogeneous non-stationary time series (w_t) will be:

$$w_t = y_t - y_{t-1} = Dy_t$$

where:

w_t = observations t of the w stationary series resulting from the subtracting of the observations $y_t - y_{t-1}$ of the nonstationary series.

If a series happened to be second order homogeneous, the series would be stationary by subtracting the first differences. It can be shown in this way

$$w_t = D^2y_t = Dy_t - Dy_{t-1}.$$

Time series data often are analyzed on the basis of

linea

proce

proce

(Box

analy

formul

Y

where:

Y

m

S.

e

As we

or the

and cou

Y_t

The seas

by seaso

a comple

data in

of Census

of the Ce

linear models in which errors are generated by stochastic processes like moving average processes, autoregressive processes, or mixed autoregressive moving average processes (Box and Jenkins,1970)

Let discuss some of the basic models in time series analysis. Each observation can be represented with this formula:

$$Y_t = m_t + S_t + e_t$$

where:

Y_t = observation in time "t"

m_t = trend or trend cycle effect

S_t = periodic or seasonal effect

e_t = random error.

As we discussed earlier, by differencing we can eliminate m_t or the trend or trend cycle. The formula will be reduced and could be written:

$$Y_t = S_t + e_t.$$

The seasonal variation (S_t) in a time series can be removed by seasonal adjustment techniques that basically consist of a complex averaging of past estimates. National economic data in the United States usually is adjusted by the Bureau of Census of the U.S. Department of Commerce (X-11 Variant of the Census Method II Seasonal Adjustment Program, 1967)

by thi

behav:

cycles

other

produc

decrea

curren

M

averag

autorec

Ir

a weigł

autoreg

past v

A

observ

error (

of an

Y_t

where:

θ_1

e_t

M

An auto

by this method of adjustment. Seasonality is a cyclical behavior that occurs on a regular calendar basis, i.e., cycles with a periodicity that is annual, monthly or any other unit of calendar time. For example, Peruvian anchovy production shows seasonal once every 7 years in response to decreased supply brought about by cyclical changes in ocean currents.

Models used to study time series are: 1) moving average, 2) autogregressive, and 3) mixed model, the autoregressive-moving average model.

In a moving average model, y_t is derived completely by a weighted sum of current and lagged random errors. In the autoregressive model, y_t depends on a weighted sum of its past values and a random error term.

A moving average model of order q where each observation y_t is generated by a weighted average of random error going back q periods (lags) can be written in the form of an equation and denoted by MA(q):

$$y_t = M + e_t - \theta\ e_{t-1} - \theta_2\ e_{t-2} - \cdots \theta_q\ e_{t-q}$$

where:

$\theta_1 \cdot \cdot \cdot \theta_q$ = fixed parametes

$e_{t-1} \cdot \cdot \cdot e_{t-q}$ = error terms times t to q

M = mean of the series.

An autoregressive model of order p where each observation y

is ge

going

the cu

denot

Y

where:

ø

M

e.

An exar
AR(1) c
writter

Yt

Th
referrec
of the

Ide
major pr
indicate

Man
average c
logical

is generated by a weighted average of t past observations
going back p periods, together with a random disturbance in
the current period can be written with this equation and
denoted as AR (p):

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdot \cdot \cdot + \emptyset_p Y_{t-p} + M + e_t$$

where:

$\emptyset_1 \cdot \cdot \cdot \emptyset_p$ = weights of 1 to q past observations

M = constant associated with the mean of the TS

e_t = error term for time t.

An example of a simple autoregressive model would be the
AR(1) or first order autoregressive process that can be
written in the equation form:

$$Y_t = \emptyset Y_{t-1} + e_t.$$

The number of past periods used in these models is
referred to as the order, the number of lags, or the memory
of the series.

Identifying the order of the underlying process is a
major problem. The knowledge of the autocorrelations will
indicate the appropriate order.

Many time series cannot be modelled as purely moving
average or as purely autoregressive. For these cases a
logical extension of the models AR and MA is used.

A

(p,q)

ARMA

$Y_t = f$

The co

I

differ

a mode

autore

denote

indica

T

in ani

them i

series

Resear

series

lactat

Anothe

satisf

animal

A mixed autoregressive moving average process of order (p,q) can be written with this equation and denoted as an ARMA (p,q) model:

$$Y_t = \emptyset \; Y_{t-1} + \cdots + \emptyset \; Y_{t-p} + M + e_t - \theta_1 e_{t-1} \cdots - \theta_q \; e_{t-q}.$$

The components of this model already have been defined.

If a series is non-stationary, it can be homogenized by differencing the series to produce the stationary series w_t; a model using this series is said to be integrated. An autoregressive model with a differenced series will be denoted as ARI (p, d) were the number of differences used is indicated by the letter d.

Time series models presented have not been used widely in animal breeding. Perhaps one of the reasons for not using them is that traditionally the traits of interests are short series, i.g., total lactation yield, annual wool production. Researchers have not thought of these observations as time series data. But other important traits like growth and lactation curves have been considered time series data. Another deterent for the use of these models is because satisfactory methods for analysis of serial measurements in animal breeding are not applied easily.

III.1.

T

Mexico

progra

Provo,

United

Univers

Th

associa

breeds.

States

either

Th

from 19

34,472 s

remainir

The

1966 Nat

Herd Imp

Th

Mexican

ied. T

III MATERIALS AND METHODS

III.1 <u>Data</u>

III.1.1 <u>Source</u> <u>of</u> <u>Data</u>

The "Asociacion de Criadores Holstein-Friesian de Mexico" sponsors a DHIA (Dairy Herd Improvement Association) program. Records are processed by the regional Center at Provo, Utah, and completed milk records are forwarded to the United States Department of Agriculture (USDA) and Cornell University. Fat test are not recorded in Mexico.

The herds consist mainly of cattle registered with the association plus grade Holsteins and crossbreds of other breeds. A number of cows were imported from the United States and Canada. Semen used to breed these cows is from either the United States, Canada, or Mexico.

The data bank contained 152,331 records for calvings from 1969 to 1980. These data are 45,655 first lactations, 34,472 second lactations, 26,345 third lactations, and the remaining are fourth and later lactations.

The record format is similar to that adopted at the 1966 National Computing Workshop and presented in the Dairy Herd Improvement Letter ARS 44-22 (1970).

This population may not be a true random sample of the Mexican dairy cattle population because animals are identified. They should constitute a superior population because

the DH

in ter

III.1.

R

For the

subset

first t

consecu

herds d

the pre

followi

Restric

arbitra

group w

confoun

the DHI herds in Mexico are under better management and are in temperate areas of the country.

III.1.2 **Data Screening and Editing Procedures**

Records were deleted for several reasons including:

Breed other than Holstein,

Incorrect Identification,

Unidentified Sire, and

Abnormal termination of lactation.

For the purpose of this study it was necessary to form a subset of data with the records of cows that completed the first three lactations. There were 5259 cows with three consecutive lactations. Also no cow in this subset changed herds during the three first lactations. In addition to the previous requirements, more records were deleted for the following reasons:

Sires having fewer than six records,

Lactations following in a single herd-year-season group, and

Records of daughters of sires disconnected.

Restrictions on the number of daughters per sire to six was arbitrary. Lactations falling in a single herd-year-season group were deleted because all other effects would be confounded with herd-year-season and would have not contri-

buted

number

A

the fi

the th

climat

throug

Mexico

type o

temper

T

each l

357, a

lactat

T

steps

matrix

I

approa

severa

lactat

could

buted to estimation. For this reason there are different number of cows in each lactation.

After screening and editing there were 3989 cows for the first lactation, 3985 cows for the second, and 3990 for the third lactation. They were distributed in 9 years, 2 climatic seasons, and 80 herds. Seasons are dry (October through March) and rainy (April through September). In Mexico seasonal management changes are determined by the type of feeding and feedstuffs available, not by temperature. This is reason for only two seasons.

There were different numbers of herd-year-seasons for each lactation also. Number of of herd-year-seasons are 361, 357, and 371 for the models for first, second, and third lactations, respectively.

III.2 Strategy to Calculate the Error Variance-Covariance Matrix

To understand methods of this study an overview of the steps needed to construct the error variance-covariance matrix of three lactations is necessary.

In analyzing dairy records there have been three approaches taken by researchers. When they studied all or several lactation of cows, they used the so-called multiple lactation model. This model incorporated the cow effect and could be described by:

where

When
lacta

where

In the
and th

A
multip
mixed
that c
multiv
colums
Y_i to
vector

$$y_i = \text{Fixed and Random Effects} + \text{Cow} + e_i$$

where:

$$e_i = \text{random error}.$$

When only one lactation was studied at a time, a single lactation model is used, and it could be described by:

$$y_i = \text{Fixed and Random Effects} + u_i$$

where:

$$u_i = \text{cow}_i + e_i.$$

In the latter case the residual u_i contains the cow effect and the error effect e_i.

Another approach to study multiple lactations is multiple trait analysis. This procedure is an extension of mixed procedures designed to analyze simultaneously traits that can be explained by similar model equations. Unlike multivariate analysis where y_i are lined up side by side as colums of a matrix Y, the multiple trait operation requires y_i to stack up on top of the next to become a long column vector, y. Therefore, a basic multiple trait model is:

$$
\begin{vmatrix}
y \\
y \\
\cdot \\
\cdot \\
\cdot \\
y
\end{vmatrix}
$$

where:

Each l

observ

trait

model.

statis

treate

covari

(1983)

residu

T

varian

model

Estimat

$$
\begin{vmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_t \end{vmatrix} = \begin{vmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & X_t \end{vmatrix} \begin{vmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_t \end{vmatrix} + \begin{vmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0 & Z_t \end{vmatrix} \begin{vmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_t \end{vmatrix} + \begin{vmatrix} u_1 \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_t \end{vmatrix}
$$

where: y_1 is vector of trait 1 (lactation 1) equation being X fixed effects and Z random effects for traits 1 to t,

b_1 is vector of solutions for fixed effects,

s_1 is vector of solutions for random sire effects,

u_1 is vector residuals lactation 1.

Each lactation is a different trait instead of repeated observations of the same cow. There are several multiple trait algorithms that have been used in the later lactation model. Walter and Mao, (1985) compared assumptions and statistical properties of each of them. Tong et al. (1979), treated each lactation as a different trait with error covariances zero. Rothschild and Henderson (1979) and Meyer (1983) developed algorithms that do not require zero residual covariances.

The approach used in this study to calculate the variance-covariance matrix was to fit a single lactation model to each of three lactations to obtain solutions. Estimates y_{ij} were obtained by each observation y_{ij} adjusted

for s

by sub

E

form a

residu

as sug

I

The E

duced

matri

varia

summa

for solutions. The individual residuals then were computed by subtracting $\hat{y}_{ij}$ from observation y_{ij}.

Estimated residuals $\hat{u}_{ij}$ of ith lactation in the jth cow form a vector of residuals. With the three vectors of residuals, one for each lactation, the matrix E was formed as suggested by Morrison (1967);

$$
E = \begin{vmatrix}
\hat{u}_{11} & \hat{u}_{21} & \hat{u}_{31} \\
\hat{u}_{12} & \hat{u}_{22} & \hat{u}_{32} \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
\hat{u}_{1n} & \hat{u}_{2n} & \hat{u}_{3n}
\end{vmatrix}.
$$

The E matrix then is premultiplied by E' (i.e., E'E is produced) to obtain the sum of squares and products SSCP matrix, which divided by the degrees of freedom becomes variance-covariance matrix VCOV named S.

Steps to estimate the variance covariance matrix S is summarized:

-Obtain solutions from fitting a single lactation model to each of the three lactations.

-Estimate $\hat{y}_{ij}$ from solutions.

-Subtract $y_{ij} - \hat{y}_{ij}$ to obtain residuals $\hat{u}_{ij}$.

-Form matrix E with the vectors of residuals $\hat{u}_{in}$.

III.3.

A

applie

y

where:

y

Mu

s_t

Hy

-Produce the SSCP matrix by multiplying E' with E.

-Divide SSCP by degrees of freedom to obtain S.

III.3 Model

III.3.1 Equations and Characteristics of Model

A model describing each of the variables of interest was applied to each lactation.

$$Y_{tijk} = Mu_t + s_{ti} + HYS_{ij} + f_{t1} DIM_{tijk} + e_{tijk}$$

where:

Y_{tijk} is the unadjusted milk yield for the tth lactation (t = 1,2,3) for the ith sire in the jth herd-year-season from a population of registered Holsteins in Mexico DHI, having their sires identified and lactating from 1969 to 1980.

Mu_t is the tth mean of the named fixed effects.

s_{ti} is the random effect of the ith sire in the tth lactation; i = i, ..., 193;

HYS_{tj} is the fixed effect of the jth herd-year-season in the tth lactation in which a cow freshened. The j has a different number of factor depending of the lactation

being fitted $j = 361$, $j = 357$, and $j = 371$
for $t = 1,2,3$.

f_{t1} is the regression coefficient of y_{tijk} on DIM

DIM_{tijk} is the effect of days in milk in the $tijk^{th}$ subclass, and

e_{tijk} is the error effect associated with y_{tijk}.

Factors HYS_{tj} and DIM_{tijk} are fixed effects and y_{tijk}, s_{ti}, and e_{tijk} are random.

$E(y) = Xb$, and the variance of y is $V = Z G Z' + R$. Because the objective of this study is to examine the variance-covariance structure of the residuals, no specific assumptions about errors are set other than multivariate normal distribution of the three dimensional vector of errors corresponding to the three lactations; Cov (s, e) = 0 is no correlation between error and sire; (Vs) = G is that there is no covariance between s's, in other words, no additive genetic relationship and independent sampling between s's; and that each sire effect is drawn from the same population with mean zero and variance Vs.

Sire effect, s, is normally distributed;

no correlation between the ranking of s_{ti} and the number of observations for s_{ti};

interactions of sire with herd-year-season effects are

trivi

s

wherea

facto

I

Y

where:

Y

X

b

Z

s

e

Ch

matrix

$E($

$E($

trivial and negligible; and

sire and residual effects are of primary interest whereas herd-year-season and days in milk are nuisance factors.

Presenting the model in matrix form one obtains:

$y = Xb + Zs + e$

where:

y is the observation vector on either the first, second, or third unadjusted lactation

X is an n x p incidence matrix where n = number of cows and p is the number of parameters to estimate, number of herd-year-seasons (HYS) and one for column Mu. It contains 1's and 0's corresponding to the presence or absence of observations in the HYS; and for observation a 1in the column for Mu and the observation for the covariate.

b is a vector of length nHYS+1 containing the unknown constants of the fixed effects
$b' = [\ u,\ HYS_1 \ldots HYS_n,\ b_1]$

Z is a n x 193 incidence matrix containing 1's and 0's corresponding to the presence or absence of observations within each sire.

s is an 193 by 1 vector of unobservable random effects for s, $s_1 \ldots s_{193}$

e is a n by 1 vector of unobservable random residuals corresponding to y.

Characteristics of the model also can be expressed in matrix form:

$E(y) = Xb$

$E(s) = 0$

where

For s

vari

I

I

$COV(s,e) = 0$

$V_y = V = ZGZ' + R$

where:

$G = V_s = E(s's)$

$\qquad = E[s - E(s)][s - E(s)]'$

$$G = E \begin{vmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ s_{193} \end{vmatrix} \begin{vmatrix} [s_1 & s_2 & \cdots & s_{193}] \cdot \end{vmatrix}$$

For s all covariances are zero and both have homogeneous variance i.e., I Ve.

$$G = \begin{vmatrix} Ve & 0 & \cdot & \cdot & 0 \\ 0 & Ve & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & Ve \end{vmatrix} = I_{193}Ve$$

R is then

$R = V_e = E(ee') = E[e - E(e)][e - E(e)]'$

$$
E = \begin{vmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{vmatrix} \qquad [e_1 \ e_2 \ \cdot \ \cdot \ \cdot \ e_{4076}] \ .
$$

For each independent model it is assumed that R has homogeneous variance, i.e., IVe.

$$
R = \begin{vmatrix} Ve & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & Ve & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & Ve \end{vmatrix} = I_n Ve
$$

In this way the variance-covariance matrix for all random factors of each model can be written

$$
Var \begin{vmatrix} y \\ s \\ e \end{vmatrix} = \begin{vmatrix} V_n & ZVs & I_n Ve \\ & I_{193}Vs & 0 \\ Sym & & I_n Ve \end{vmatrix} .
$$

III.3.2 Numerical Example

To describe methods used for setting up and solving mixed model equations, it is often beneficial to do so in an example.

A hypothetical example of ten cows in three herd-year-seasons (HYS) with two covariates is illustrated. Data are in Tables 3.1 and 3.2.

Table 3.1 - Data for example problem.

Cow	Sire Identification	HYS	Milk Yield kg/100	Cov 1	Cov 2
1	1	1	600	396	305
2	1	2	700	401	380
3	1	3	800	304	280
4	1	1	580	400	370
5	2	2	420	365	305
6	2	2	560	426	360
7	2	3	910	550	410
8	3	1	1010	370	320
9	3	1	340	400	340
10	3	3	580	410	380

Table 3.2 - Sire by Herd-year-season Distribution.

HYS	Sire 1	2	3	SUM_i
1	2	0	2	4
2	1	2	0	3
3	1	1	1	3
SUM_i	4	3	3	10

The model can be written in matrix notation:

$$y = Xb + Zs + e$$

where:

y   is the observation vector of milk yield

X   is an n x p incidence matrix, where n = 10 number of cows, p is number of parameters to estimate, and one for column Mu. It contains 1's and 0's corresponding to the presence or absence of observations in the HYS_i, and for each observation a 1 in the column for Mu and observations for each covariate.

b   is a vector of length 6 containing the unknown constants of the fixed effects

Z   is a 10 x 3 incidence matrix containing 1's and 0's corresponding to the presence or absence of observations within each sire.

s   is a 3 x 1 vector of unobservable random effects for s,

e   is a 10x 1 vector of unobservable random residuals corresponding to y.

$$[\ y\ ] = [\qquad\qquad X \qquad\qquad]\ [\ b\ ]$$

y			X				b
600	1	1	0	0	396	305	Mu
700	1	0	1	0	401	380	HSY_1
800	1	0	0	1	304	280	HYS_2
580	1	1	0	0	400	370	HYS_3
420	1	0	1	0	355	305	COV_1
560	1	0	1	0	426	360	COV_2
910	1	0	0	1	550	410	
1010	1	1	0	0	370	320	
340	1	0	0	1	400	340	
580	1	0	0	1	410	380	

$$[\quad\quad Z \quad\quad] \quad [\ s\ ] \quad + \quad [\ e\ ]\ .$$

$S1$	S_2	S_3
1	0	0
1	0	0
1	0	0
0	1	0
0	1	0
0	1	0
0	0	1
0	0	1
0	0	1
0	0	1

$$\begin{vmatrix} S_1 \\ S_2 \\ S_3 \end{vmatrix} \qquad \begin{vmatrix} e_{111} \\ e_{122} \\ e_{133} \\ e_{114} \\ e_{225} \\ e_{226} \\ e_{237} \\ e_{318} \\ e_{329} \\ e_{3210} \end{vmatrix}$$

The normal equation are then:

$$\begin{vmatrix} X'X & X'Z \\ Z'X & Z'Z \end{vmatrix} \begin{vmatrix} b \\ s \end{vmatrix} = \begin{vmatrix} X'y \\ Z'y \end{vmatrix}$$

and become for the example:

$$
\begin{bmatrix} & & \mathbf{X'X} & & & & & \mathbf{Z'Z} & & \end{bmatrix} \; [\mathbf{b,\ s}]'
$$

	X'X					Z'Z			[b, s]'
10	4	3	3	4022	3454	4	3	3	Mu
4	4	0	0	1566	1335	2	0	1	HYS$_1$
3	0	3	0	1192	1045	0	2	1	HYS$_2$
3	0	0	3	1264	1070	2	0	1	HYS$_3$
4022	1566	1192	1264	1652234	1406665	1501	1341	1180	COV$_1$
3450	1335	1045	1070	1406665	1205850	1335	1075	1040	COV$_2$
4	2	0	2	1501	1335	4	0	0	S$_1$
3	1	2	0	1341	1075	0	3	0	S$_2$
3	1	1	1	1180	1040	0	0	3	S$_3$

[**X'y, Z'y**]'.

$$
= \begin{bmatrix}
2{,}530 \\
1{,}680 \\
2{,}290 \\
2{,}633{,}360 \\
2{,}249{,}600 \\
\hline
2{,}680 \\
1{,}890 \\
1{,}930
\end{bmatrix}
$$

The variance of y = Vy = ZGZ' + R

$$
\text{Vy} =
\begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
Ve & 0 & 0 \\
0 & Ve & 0 \\
0 & 0 & Ve
\end{bmatrix}
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1
\end{bmatrix}
$$

[Z] [G] [Z']

10x3   3x3   3x10

$$
+
\begin{bmatrix}
Ve &   &   &   &   &   &   &   &   &   \\
   & Ve &   &   &   &   &   &   &   &   \\
   &   & Ve &   &   &   &   &   &   &   \\
   &   &   & Ve &   &   &   &   &   &   \\
   &   &   &   & Ve &   & 0 &   &   &   \\
 0 &   &   &   &   & Ve &   &   &   &   \\
   &   &   &   &   &   & Ve &   &   &   \\
   &   &   &   &   &   &   & Ve &   &   \\
   &   &   &   &   &   &   &   & Ve &   \\
   &   &   &   &   &   &   &   &   & Ve
\end{bmatrix}
.
$$

[R]

$$= \begin{bmatrix} & & & & & Vy & & & & & \end{bmatrix}$$

$$\begin{bmatrix}
Vs+Ve & Vs & Vs & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & Vs+Ve & 0 & Vs & Vs & Vs & 0 & 0 & 0 & 0 \\
 & & Vs+Ve & 0 & 0 & 0 & Vs & Vs & Vs & Vs \\
 & & & Vs+Ve & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & & Vs+Ve & 0 & 0 & 0 & 0 & 0 \\
\text{Symmetric} & & & & & Vs+Ve & 0 & 0 & 0 & 0 \\
 & & & & & & Vs+Ve & 0 & 0 & 0 \\
 & & & & & & & Vs+Ve & 0 & 0 \\
 & & & & & & & & Vs+Ve & 0 \\
 & & & & & & & & & Vs+Ve
\end{bmatrix} .$$

Henderson (1975) derived mixed model equations (MME) by a maximum likelihood approach assuming normality. He proved that fixed constants from MME are equivalent to those from normal equations of generalized least squares and that predictors from MME are best linear unbiased predictor (BLUP) with or without normality assumption. Prediction refers to the likely outcome that one would expect of a future occurrence from a class, which is randomly out of a population of classes defined by a factor. The mixed model equation MME of best linear unbiased predictors are (Mao, 1983):

$$
\begin{vmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z +G^{-1} \end{vmatrix} \begin{vmatrix} b \\ s \end{vmatrix} = \begin{vmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{vmatrix}
$$

Then, $X'R^{-1}X$ develops from:

[X'   ]

1	1	1	1	1	1	1	1	1	1
0	0	1	0	0	0	1	0	0	1
0	0	0	0	1	1	0	0	1	0
1	1	0	1	0	0	0	1	0	0
410	400	370	550	426	355	400	304	401	396
380	340	320	410	360	305	370	280	380	305

6x10

[R^{-1}]

1/Ve	0	0	0	0	0	0	0	0	0
	1/Ve	0	0	0	0	0	0	0	0
		1/Ve	0	0	0	0	0	0	0
			1/Ve	0	0	0	0	0	0
				1/Ve	0	0	0	0	0
					1/Ve	0	0	0	0
						1/Ve	0	0	0
	Sym						1/Ve	0	0
								1/Ve	0

$$
\begin{array}{c}
[\qquad\qquad X \qquad\qquad] \\
\left|
\begin{array}{cccccc}
1 & 1 & 0 & 0 & 396 & 305 \\
1 & 0 & 1 & 0 & 401 & 380 \\
1 & 0 & 0 & 1 & 304 & 280 \\
1 & 1 & 0 & 0 & 400 & 370 \\
1 & 0 & 1 & 0 & 355 & 305 \\
1 & 0 & 0 & 1 & 550 & 410 \\
1 & 1 & 0 & 0 & 370 & 320 \\
1 & 0 & 0 & 1 & 400 & 340 \\
1 & 0 & 0 & 1 & 410 & 380
\end{array}
\right| .
\end{array}
$$

Similiarly:

$$
\begin{array}{c}
[\qquad X'R^{-1}Z \qquad] \\
\left|
\begin{array}{ccc}
4 & 3 & 3 \\
2 & 0 & 1 \\
0 & 2 & 1 \\
2 & 0 & 1 \\
1501 & 1341 & 1180 \\
1335 & 1075 & 1040
\end{array}
\right| \qquad 1/Ve.
\end{array}
$$

Because G^{-1} is:

$$
\left|
\begin{array}{ccc}
1/Vs & 0 & 0 \\
0 & 1/Vs & 0 \\
0 & 0 & 1/Vs
\end{array}
\right| .
$$

$$
\begin{bmatrix} & Z'R^{-1}\,Z \;+\; G^{-1} & \end{bmatrix}
$$

$$
\begin{vmatrix}
4 + 1/Vs & 0 & 0 \\
0 & 3 + 1/Vs & 0 \\
0 & 0 & 3 + 1/Vs
\end{vmatrix} \quad 1/Ve
$$

$$
\begin{bmatrix} & & & X'R^{-1}\,X & & \end{bmatrix}
$$

$$
\begin{vmatrix}
10 & 4 & 3 & 3 & 4022 & 3454 \\
4 & 4 & 0 & 0 & 1566 & 1335 \\
3 & 0 & 3 & 0 & 1192 & 1045 \\
3 & 0 & 0 & 3 & 1264 & 1070 \\
4022 & 1566 & 1192 & 1264 & 1652234 & 1406665 \\
3450 & 1335 & 1045 & 1070 & 1406665 & 1205850
\end{vmatrix} \; 1/Ve
$$

In the same way:

$$
\begin{bmatrix} & & Z'R^{-1}X & & \end{bmatrix}
$$

$$
\begin{vmatrix}
4 & 2 & 0 & 2 & 1501 & 1335 \\
3 & 1 & 2 & 0 & 1341 & 1075 \\
3 & 1 & 1 & 1 & 1180 & 1040
\end{vmatrix} \; 1/Ve
$$

The right hand side is:

$$
\begin{vmatrix}
2{,}530 \\
1{,}680 \\
2{,}290 \\
2{,}633{,}360 \\
2{,}249{,}600 \\
2{,}680 \\
1{,}890 \\
1{,}930
\end{vmatrix}.
$$

The mixed model equations are multiplied by R, and G^{-1} is multipled by Ve/Vs, which is the variance ratio for error and sire components. In doing this R^{-1} is cancelled from both sides leaving a ratio of Ve/Vs = L in the diagonal of the Z'Z portion.

The final MME is this:

10	4	3	3	4022	3454	4	3	3
4	4	0	0	1566	1335	2	0	1
3	0	3	0	1192	1045	0	2	1
3	0	0	3	1264	1070	2	0	1
4022	1566	1192	1264	1652234	1406665	1501	1341	1180
3450	1335	1045	1070	1406665	1205850	1335	1075	1040
4	2	0	2	1501	1335	4+Ve/Vs	0	0
3	1	2	0	1341	1075	0	3+Ve/Vs	0
3	1	1	1	1180	1040	0	0	3+Ve/Vs

HYS_1		2,530
HYS_2		1,680
HYS_3	=	2,290
COV_1		2,633,360
COV_2		2,249,600
s_1		2,680
s_2		1,890
s_3		1,930

.

To solve for these equations a L ratio of 15, derived from estimates of heritability for milk yield was used. The

heritability of milk yield in dairy cattle has been estimated to be around .25. The paternal half-sisters heritability is:

$$h^2 = .25 = 4Vs / (Ve + Vs), \text{ set } Vs = 1,$$

then: $.25 = 4(1) / (1 + Ve),$

and $(4 /.25)-1 = Ve = 15,$

$$L = Ve/Vs = 15/1 = 15.$$

then:

$$L = 15.$$

Solutions are:

$$
\begin{vmatrix}
HYS_1 \\
HYS_2 \\
HYS_3 \\
COV\ 1 \\
COV\ 2 \\
s1 \\
s2 \\
s3
\end{vmatrix}
=
\begin{vmatrix}
559 \\
492 \\
684 \\
.64 \\
-.54 \\
7.37 \\
-3.46 \\
-3.91
\end{vmatrix}.
$$

III.3.3 <u>Programming</u> <u>Strategy</u>

The actual labor of setting up the mixed model equations requires understanding matrix manipulations. In programming mixed model equations one cannot read the records and create the incidence (0's and 1's) matrix X and premultiply it to create X'X. This would require much computer memory, making this alternative inefficient. Therefore, it was required to set the mixed model equations, directly. To reduce further the number of equations within manageable limits and still obtain the exact same solutions as if all the equations were solved, a mathematical procedure called absorption was used (Mohammad et al. 1985). For example, in this case one may have thousands of HYS's, and solutions for HYS's in this study are not of great importance, then one may absorb HYS into sire and covariate group.

As described by Searle (1971), it is possible to reduce the mixed model equation by partitioning the X'X matrix into four submatrices and the X'y vector into two subvectors. This procedure is called absorption by partition or block absorption because it involves reduction of equations as a block after the normal equation is partitioned. To illustrate this process we will absorb the HYS submatrix into the Sire submatrix.

For this example, only the Sire, and HYS portion of the normal equation will be used, the covariate will ignored for simplicity porpuses. The row and column for Mu will be ignored, because the sum of the effects equals the Mu row and column.

The normal equation will be of this form:

$$
\begin{vmatrix} Z'Z & Z'X \\ X'Z & X'X \end{vmatrix} = \begin{vmatrix} Z'y \\ X'y \end{vmatrix}
$$

$$
\begin{vmatrix}
4 & 0 & 0 & 2 & 1 & 1 \\
0 & 3 & 0 & 0 & 2 & 1 \\
0 & 0 & 3 & 2 & 0 & 1 \\
2 & 0 & 2 & 4 & 0 & 0 \\
1 & 2 & 0 & 0 & 3 & 0 \\
1 & 1 & 1 & 0 & 0 & 3
\end{vmatrix} = \begin{vmatrix}
2680 \\
1890 \\
1930 \\
2530 \\
1680 \\
2290
\end{vmatrix} .
$$

To absorb HYSs submatrix into sire submatrix we will use the following formula

$$
Z'Z_{X'X\ absorbed} = Z'Z - Z'X\ (X'X)^{-1}\ X'Z\ .\ .\ .\ .(1)
$$

$$
Z'Z_A = \begin{bmatrix} Z'Z \end{bmatrix} - \begin{bmatrix} Z'X \end{bmatrix} \begin{bmatrix} (X'X)^{-1} \end{bmatrix} \begin{bmatrix} X'Z \end{bmatrix}
$$

$$
Z'Z_A = \begin{vmatrix}
4 & 0 & 0 \\
0 & 3 & 0 \\
0 & 0 & 3
\end{vmatrix} - \begin{vmatrix}
2 & 1 & 1 \\
0 & 2 & 1 \\
2 & 0 & 1
\end{vmatrix} \begin{vmatrix}
1/4 & 0 & 0 \\
0 & 1/3 & 0 \\
0 & 0 & 1/3
\end{vmatrix} \begin{vmatrix}
2 & 0 & 2 \\
1 & 2 & 0 \\
1 & 1 & 1
\end{vmatrix} .
$$

$$\begin{vmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{vmatrix} - \begin{vmatrix} .5 & .33 & .33 \\ 0 & .66 & .33 \\ .5 & 0 & .33 \end{vmatrix} \begin{vmatrix} 2 & 0 & 2 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{vmatrix}$$

$$\begin{vmatrix} 1.66 & 1 & 1.33 \\ 1 & 1.66 & .33 \\ 1.33 & .33 & 1.33 \end{vmatrix}$$

$$Z'Z_A = \begin{vmatrix} 4 & 0 & 3 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{vmatrix} - \begin{vmatrix} 1.66 & 1 & 1.33 \\ 1 & 1.66 & .33 \\ 1.33 & .33 & 1.33 \end{vmatrix} = \begin{vmatrix} 2.33 & -1 & 1.33 \\ -1 & 1.33 & -.33 \\ 1.33 & -.33 & 1.66 \end{vmatrix}.$$

To absorb the HYS subvector y'X into the sire subvector y'Z
we use the following formula

$$Z'y_{X'yabsorbed} = Z'y - Z'X (X'X)^{-1} X'y \quad \ldots \quad (2)$$

$$y'Z_A = \begin{array}{ccc} [Z'y] & [\quad Z'X (X'X)^{-1} \quad] & [X'y] \end{array}$$

$$y'Z_A = \begin{vmatrix} 2680 & .5 & .33 & .33 & 2530 \\ 1890 & - & 0 & .66 & .33 & 1680 \\ 1930 & .5 & 0 & .33 & 2290 \end{vmatrix}$$

$$\begin{vmatrix} 2680 & 2588.33 \\ 1890 & - & 1883.33 \\ 1930 & 2028.73 \end{vmatrix} = \begin{vmatrix} 91.67 \\ 6.67 \\ -98.33 \end{vmatrix}.$$

The new normal equations with the nuisance effect absorbed
will be:

$$\begin{vmatrix} 2.33 & -1 & -1.33 \\ -1 & 1.33 & -.33 \\ -1.33 & -.33 & 1.66 \end{vmatrix} \begin{vmatrix} S_1 \\ S_2 \\ S_3 \end{vmatrix} = \begin{vmatrix} 91.67 \\ 6.67 \\ -98.33 \end{vmatrix}.$$

This procedure is, nevertheless, inadequate for large data
sets and computers because it will require set up and
storage of the entire matrix.

An alternative method is to use a row and column
technique. This absorbs one class at a time while setting
up equations pertaining to the effects of interest. This
technique requires that data be sorted by classes of the
factor being absorbed. Additional sorting by classes of the
factors being adjusted reduces the programming task, but it
is not required.

The mixed model equation will be constructed from all
records of one HYS group at a time. After all of its records
are read, it will be absorbed immediately into the rest of
the equation. In this way, it never will be necessary to
store more than one HYS at a time in computer memory nor
will execution time be spent to obtain solutions for HYS.

This process will be illustrated with the same data
sample. The algorithm is as follows. First, data are
sorted by HYS; then sires are sorted within HYS. The data

sorted in this manner are presented in Table 3.3.

Table 3.3- Example data sorted by HYS and sire sorted
within HYS.

HYS	Sire	Cow	Milk Yield kg/100	Cov1	Cov2
1	1	1	600	396	305
1	1	4	580	400	370
1	3	8	1010	370	320
1	3	9	340	400	340
2	1	2	700	401	380
2	2	5	420	365	305
2	2	6	560	426	360
3	1	3	800	304	280
3	2	7	910	550	410
3	3	10	580	410	380

Because we are going to absorb HYS into the matrix con-
taining sires and the covariates, let us call this matrix $Z'Z$
and the elements of this matrix A(i row, j column). The
right hand side vector will be RHS and its elements r_i.

For our example, a memory array of n(number of sires) +
2(number of covariate) squared for $Z'Z$ and a vector space of
n+2 for RHS.

We will read the first HYS into $Z'Z$, and element A_{ij}
represents the number of daughters of the sire A_{ij}, the r_i
elements of RHS are the totals of milk yield of the
daughters of sire i. For the covariates part, the A_{ij}
elements are totals of the covariate for sire as well as the
sums of squares and crossproducts of the covariates in the

RHS, the r_i are crossproducts of the covariates with milk yield.

As the first HYS is read, it will accumulate in a vector the total number of observation for each sire as well as totals for the covariates in a vector ($Vect_{HYS}$) of order n+2 x 1. We also will store the total number of observation in each HYS in the scalar n_{HYS} and the total milk yield in scalar y_{HYS}.

$$
\left[\begin{array}{cccc}
 & Z'Z_A & & \\
A_{11} & A_{12} & \cdot & A_{15} \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
A_{15} & \cdot & \cdot & A_{55}
\end{array} \right]
\qquad
\left[\begin{array}{c}
RHS \\
r_1 \\
\cdot \\
\cdot \\
\cdot \\
r_5
\end{array} \right]
$$

After reading all the records contained in each HYS we will have $Vect_{HYS}$ and the scalars n_{HYS} and y_{HYS}.

$Vect_{HYS}$

$$\begin{vmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ \cdot \\ V_5 \end{vmatrix}$$

$[n_{HYS}]$ $[Y_{HYS}]$

where:

n_{HYS} = number of records in HYS_i

Y_{HYS} = Z'y's in HYS_i.

The new elements for matrix Z'Z and vector RHS will be generated by:

$$A_{ij} = A_{ij} - Vect_i \times Vect_j/n_{HYS} \quad \cdots \quad (3)$$

$$r_i = r_i \quad - Vect_i \times Y_{HYS}/n_{HYS} \quad \cdots \quad (4).$$

After absorption of one HYS, the vector Vect and scalars n_{Hys} and Y_{HYS} are zeroed out for the next HYS. With this procedure only one pass of the data is required to complete the absorption and setup the normal equations. This procedure will be illustrated with the data used previously. Remember that herd-year-seasons will be absorbed into sires and covariates reducing our matrix by three to a 5 x 5 which equals the number of sires plus the number of covariates.

The first $Z'Z_A$ matrix is formed after four records falling in the first HYS class are read.

```
[                    Z'Z_A                        ]
[ S1      S2      S3      Cov1      Cov2 ] [ Vect ] [ RHS ]

  2        0       0       796       675       2       1180

  0        0       0         0         0       0          0

  0        0       2         0         0       2       1350

796        0     770    613716    523180    1566     979300

675        0     660    523180    447925    1335     836400
```

$$y_{HYS} = [2,530] \qquad\qquad n_{HYS} = [4].$$

To form the absorb matrix $Z'Z_A$ and vector RHS we use the formulas (3) and (4), i.e.,

A11 = 2 - (2x2)/4 = 1   r1 = 1180-(2x2530)/4 = -85
A13 = 0 - (2x2)/4 = -1
A31 = 0 - (2x2)/4 = 1   r2 = 0 - 0 = 0
A33 = 2 - (2x2)/4 = -1
A14 = 796 - (2x1566)/4 = 13   r3 = 1350-(2x2530)/4 = 85
A41 = 796 - (2x1566)/4 = 13
A43 = 770 - (2x1566)/4 = -13   r4 = 979,300-(1566x2530)/4
A34 = 770 - (2x1566)/4 = -13   = -11,195
A15 = 675 - (2x1335)/4 = 7.5
A51 = 675 - (2x1335)/4 = 7.5
A35 = 660 - (2x1335)/4 = 7.5   r5 = 836,400-(1335x2530)/4
A53 = 660 - (2x1335)/4 = -7.5   = -7,987.5
A44 = 613,716-(1566x1566)/4 = 627
A55 = 447,925-(1335x1335)/4 = 2368.75
A45 = 523,180-(1335x1566)/4 = 527.5
A54 =   "   "   "   = 527.5

[Z'Z$_A$]   [RHS   ].

1	0	-1	13	7.5		-85
0	0	0	0	0		0
-1	0	1	-13	-7.5		85
13	0	-13	627	527.5		-11,195
7.5	0	-7.5	527	2368.75		-7,987.5

The following HYS class is accumulated into Z'Z$_A$ and RHS

[Z'Z$_A$] [Vect] [RHS   ]:

1+1	0	-1+0	13+401	7.5+380		1		700-85
0	2	0	791	665		2		980+0
-1	0	1	-13	-7.5		0		85
13+401	791-13	-13	627 +475,065	527.5 +417,065		1192		672,560 -11,195
7.5+38	665	-7.5	527.5 +417,065	2368.7 367,025.		1045		595,700.0 -7,987.5

n_{HYS} = [3] y_{HYS} = [1680].

New Z'Z$_A$ and RHS are then calculated with formulas (3) and
(4), i.e.,

A11 = 2 -(1x1)/3 = 1.66

A12 = 0 -(1+2)/3 = -.66

.
.
.

r5 = 5,877,125.5 - (1045x1680)/3 = 2512.5.

The second Z'Z$_A$ and RHS is then obtained:

$$
\begin{bmatrix} & & Z'Z_A & & \end{bmatrix} \begin{bmatrix} RHS \end{bmatrix}.
$$

1.66	-.66	-1	16.66	39.16		55
-.66	.66	0	-3.66	-31.66		-140
-1	0	1	-13	-7.5		85
-16.66	-3.66	-13	2,507.66	2,379.16		6155
39.16	-31.66	-13	2,379.16	5,385.41		2512.5

This example only has three HYS classes to absorb. Thus, the information in the last HYS group is accumulated now.

$$
\begin{bmatrix} & & Z'Z_A & & \end{bmatrix} \begin{bmatrix} Vect \end{bmatrix} \begin{bmatrix} RHS \end{bmatrix}
$$

1+1.66	0-.66	0-1	16.66+304	39.16+280	1	800+55
0-.66	1+.66	0	-3.66+550	-31.66+410	1	910+140
0-1	0	1+1	-13+410	-7.5+3.8	1	580+85
16.66+304	-3.66+550	-13+410	+2507 563016	+2,379 466420	1264	6155 +981500
39.16+280	-31.66+410	-7.5+38	+2379 466420	+5385 390900	1070	2512 +817500

$$n_{HYS} = [3] \qquad Y_{HYS} = [1680].$$

The last $Z'Z_A$ and RHS then are calculated with algorithm formulas (3) and (4), i.e.,

A11 = 2.66 - (1x1)/3 = 2.33

A12 = -.66 - (1x1)/3 = -.99 ~ 1

.

.

r5 = 820,012 - (1070x2290)/3 = 3245.85.

The final absorbed matrix $Z'Z_A$ and vector RHS are:

[		$Z'Z_A$		]	[	RHS	] .
2.33+15	-1	-1.33	-100.67	-37.50		91.67	
-1	1.33+15	-.33	125	21.67		6.67	
-1.33	-.33	1.66+15	-24.33	15.83		-98.33	
100.67	125	-24.33	32958	17972		10491.6	
-37.50	21.67	15.83	17972	14652		3245.83	

Before $Z'Z_A$ is inverted and s and covariate are solved the ratio L was added to the diagonal elements of random or sire components as one can see in the underlined elements of $Z'Z_A$. This is to follow the MME procedures described.

The solutions obtained by these two methods of absorption are the same as those from the unabsorbed matrix. However, no solutions are obtained for the absorbed fixed effects. The solution vector is:

$\hat{s}_1$		7.375
$\hat{s}_2$		-3.462
$\hat{s}_3$	=	-3.913
COV~$_1$		.649
COV~$_2$		-.546

The solutions for HYS can be obtained by back solving. This is using the solutions, the accumulated Vect of each HYS, and nHYS and yHYS. This either requires storing the Vect, nHYS and yHYS information or rereading the data. Generally

the data are reread as this reduces storage requirements significantly. For the present example one has for HYS_1.

Solutions	$Vect_1$	nHYS	$YHYS_1$
7.375	2	[4]	[2580]
-3.462	0		
-3.913	2		
.649	1566		
-.546	1335		

HYS_1 = [2580 - 2(7.375) - 0(7.375) - 0(-3.462) - 2(-3.913)

$\qquad$ -(.649) (1566) - (-.546) (1335)]/4

$\qquad$ = [2235.65]/4 = 558.91.

For HYS_2 similarly;

$Vect_2$	$nHYS_2$	$YHYS_2$
1	[3]	[1680]
2		
0		
1192		
1045		

HYS_2 = [1680 - 1(7.375) - 2(-3.462) -

$\qquad$ .649(1192) - -.546 (1045)]/3

$\qquad$ = [1476.51]/3 = 492.17.

Finally, for HYS_3 we have:

$Vect_3$	$nHYS_3$	$YHYS_3$
1	[3]	[2290]
1		
1		
1264		
1070		

$$HYS_3 = [2290 - (7.375) - (-3.462) - (2.913)$$

$$-.649 (1264) - -.546(1070)]/3$$

$$= [2053.88]/3 = 684.62.$$

In this way the vector of solutions is identical to that obtained with the direct inversion approach. This also could be used to check the two methods.

$$
\begin{array}{|c|}
HYS_1 \\
HYS_2 \\
HYS_3 \\
s1 \\
s2 \\
s3 \\
COV_1 \\
COV_2 \\
\end{array}
=
\begin{array}{|c|}
559 \\
492 \\
684 \\
7.37 \\
-3.46 \\
-3.91 \\
.64 \\
-.54 \\
\end{array}
.
$$

An in-house Fortran program to accomplish this task was developed. The International Mathematical and Statistical Library (IMSL) of Fortran subroutines was used in part to do the matrix manipulations. The matrix inversion algorithm used was developed by Healy (1968). The in-house program is in Appendix A.

III.4 <u>Variance</u> <u>Components</u> <u>Estimation</u>

There are many methods for estimation of variance
components from unbalanced data. These methods vary in
degrees of computational complexity. There is no method
that can be said to be best for all situations (Mao, 1981).
The appropriate method is determined by the statistical
properties and characteristics of the model studied. The
iterative restricted maximum likelihood (REML) procedure
using solutions from mixed model equations (MME) was used to
compute variance components in this study. The most
desirable characteristics of REML are (Mao, 1981):

1) It renders non-negative estimates of variance compo-
nents when MME solutions are used in maximum likelihood
equation.

2) It maximizes the random portion of the likelihood
which is invariant to the fixed effects in a mixed model and
does not require that the fixed effects are known, as in
maximum likelihood (ML).

3) REML can be used in iterative computations.

Let us continue using the sample of data to illustrate
the iterative restricted maximum likelihood procedure.

The solutions from mixed model equation $\hat{b}$ and $\hat{s}$ will be
computed for each lactation by:

$$
\begin{vmatrix} \hat{s} \\ \hat{b} \end{vmatrix} = \begin{vmatrix} Z'Z + G^{-1}L & Z'X \\ X'Z & X'X \end{vmatrix}^{-} \begin{vmatrix} Z'y \\ X'y \end{vmatrix}
$$

where: L is a diagonal matrix of the variance ratio of $\hat{V}e/\hat{V}s$, and G^{-1} is the inverse of the relationship matrix for sires.

Let C be generalized inverse of the coefficient matrix:

$$
\begin{vmatrix} Z'Z + G^{-1} & Z'X \\ X'Z & X'X \end{vmatrix}^{-} = \begin{vmatrix} C_{ss} & C_{sx} \\ C_{xs} & C_{xx} \end{vmatrix}.
$$

The generalized inverse of A absorbed matrix from sample data with the ratio L of 15 is the following:

$$
C^{-} = \begin{vmatrix}
0.0590 & 0.0017 & 0.0054 & 0.0003 & -.0002 \\
0.0017 & 0.0649 & 0.00001 & -0.005 & .0006 \\
0.0054 & 0.00001 & 0.0612 & 0.0002 & -0.0003 \\
0.0003 & -.0005 & 0.0002 & 0.00009 & -.0001 \\
-.0002 & 0.0006 & -.0003 & -.0001 & 0.0002
\end{vmatrix}
$$

$$L_1 = 15.$$

The RHS are:

$$\begin{vmatrix} 91.67 \\ 6.67 \\ -98.33 \\ 1049.16 \\ 3245.83 \end{vmatrix}.$$

The solutions are:

$$\begin{vmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{b}_1 \\ \hat{b}_2 \end{vmatrix} = \begin{vmatrix} 7.376 \\ -3.463 \\ -3.910 \\ .649 \\ -.546 \end{vmatrix}.$$

Where $\hat{s}_i$ corresponds to the random effects, sire effects, and $\hat{b}_i$ are solutions for covariates. The REML estimates of the variance components are:

$$\hat{V}e = (y'y_c - b'X'y - s'Zy - Ls's)/[n- r(x)].$$

Because the absorption of some of the fixed effects HYS_i took place prior to solving the equations, no explicit solutions of those absorbed fixed classes are available. The computation of residual variances required some adjustments. During the absorption process, sums of squares pertaining to the HYS were collected.

These sums of squares are to be subtracted from the total sums of squares, $y'y_c$. Because the program was designed to obtain the HYS_i solutions with the back up procedure previously illustrated, the appropriate sum of

squares was calculated, therefore, with the HYS solutions; both techniques give identical results.

The product of L $s_s's_s$ also had to be substracted for the computation of the residual variance. This corresponds to the L variance ratio added to the random portion of the MME as demonstrated by Thompson (1969).

Because only one random factor is involved as in the present model the sire variance is computed under the assumption:

$$V(\hat{s}_s) = G_{Vs}.$$

Therefore:

$$\hat{V}s = [s'_s s'_s + (trC_{ss})] / q_s$$

where:

trC_{ss} is the trace of the C generalized inverse corresponding to sires

q_s is the number of classes in the random sire effects.

Let us continue with the numerical example.

$$Y'Y_c = Y'Y - SS_{HYS}$$

where:

SS_{HYS} is the sum of squares of absorbed HYS.

$$\mathbf{y'y_c} = 4,616,600 - \frac{2530^2}{4} - \frac{1680^2}{3} - \frac{2290^2}{3}$$

$$= 4,616,60 - 4,289,058$$

$$= 32,7582$$

$$\mathbf{b'x'y} = (0.649)(10491.6) + (-.546)(3246)$$

$$= 5037$$

$$\mathbf{s'Z'y} = (7.37)(91.67) + (-3.46)(6.67) + -3.91$$

$$= 1039$$

$$\mathbf{\hat{s}'\hat{s}} = 7.37^2 + -3.46^2 + -3.91^2$$

$$= 81.8$$

$$\mathbf{L\hat{s}'\hat{s}} = (15)(81.84) = 1228.$$

$$\mathbf{trC_{ss}} = 0.059 + 0.0649 + 0.0612 = 0.1851$$

$$\mathbf{\hat{V}e} = (327,582 - 5037 - 1039 - 1228)/8$$

$$= 40,035$$

$$\mathbf{\hat{V}s} = [81.84 + 90.1851)(40,035)]/3$$

$$= 2,497.$$

With the estimate of $\hat{V}e$ and $\hat{V}s$, a new ratio can be computed.

$$\mathbf{L_2} = \hat{V}e / \hat{V}s = 40,035/2497 = 16$$

As one can see, $\hat{s}$ and $\hat{b}$ depend on L; $\hat{V}e$ relies on $\hat{s}$, $\hat{V}e$, and L; $\hat{V}e$ relies on $\hat{s}$, $\hat{b}$, and L; and $\hat{V}e$ and $\hat{V}s$ are needed to compute a new estimate of L. These relationships lend REML to iterative computation. The heritabilities in the literature determined the first L. The process of solving for $\hat{b}$ and $\hat{s}$, then computing $\hat{V}e$ and $\hat{V}s$ is continued by replacing L with the new ratio of $\hat{V}e$ / $\hat{V}s$ and recomputing the REML estimators until the current and new ratio converge (convergence was defined by when the difference between the current and new ratio was less than.2).

From the first round of iteration in our numerical example, the new L ratio calculated with the variance estimates was 16. This ratio was added to the sire portion of the MME before inversion, and a new process then was started. The statistics for the computation are:

$$
C^- = \begin{vmatrix}
.0755 & .0015 & .0047 & .0002 & -.002 \\
.0015 & .0603 & .000 & -.0005 & .0005 \\
.0047 & .000 & .0570 & .0002 & -.0003 \\
-.0002 & -.0005 & .0002 & .0000 & -.0001 \\
-.0002 & 0.0005 & -.0003 & -.0001 & .0002
\end{vmatrix}
$$

$L_2 = 16.$

The new solutions are

$$
\begin{vmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \tilde{b}_1 \\ \tilde{b}_2 \end{vmatrix} = \begin{vmatrix} 6.92 \\ -3.22 \\ -3.70 \\ 0.645 \\ -.544 \end{vmatrix}
$$

$y'Y_C = 32,7582$ $trC_{ss} = .1928$

$b'xy = 5001$ $\hat{s}'\hat{s} = 85.63$

$s'Zy = 976$ $L\hat{s}'\hat{s} = 1370$

$\hat{V}e = 40,029$

$\hat{V}s = 2,601$

$L_3 = 15.32$.

If one were to continue the new ratio to be used would be 15.32. For the purpose of this example it was chosen to stop here.

III.5 <u>Heritabity</u> <u>Estimates</u>

With the estimates of residual and sire components of
variance an estimate of heritability for milk yield was
calculated. As defined by Lush (1945) heritability in the
"narrow sense" is the proportion of the total variance in a
trait that is attributed to the average of additive effects
of genes. Heritability in the "broad sense" is the fraction
of total variance attributable to genetic variance, which
not only contains the variance due to additive effects but
includes variance due to dominance and epistatic effect.
Heritability in the "narrow sense" is the one usually
referred in the literature and in this study.

Genetic theory indicates that in a randomly mated popu-
lation the half-sibs Vs is equal to one-quarter of the
additive genetic variance (with zero epistatic variance),
and Vs + Ve expresses the total phenotypic variance.
Therefore, heritability in the narrow sense can be expressed
as:

$$h^2 = 4 \; Vs \; / (Vs+Ve).$$

The denominator is the phenotypic variance after the
variance due to named fixed effects in the model have been

removed.

The approximate standard error of the ratio of variance components will be used to compute the standard errors of heritability (Dickerson, 1969).

$$SE(\hat{X}/\hat{Y}) = 4/\hat{Y} \; [\; SQR \; V(\hat{X}) \;].$$

III.6 Ranking of Sires

Sire solutions for each lactation were compared. Spearman's rank correlation analysis was used (Gill, 1978). Spearman's r_s is defined as the sum of the squared differences in the paired ranks for two variables over all cases, divided by a quantity which can be described as what the sum of the squared differences in ranks would have been had the two sets of rankings been totally independent.

$$r_s = 1 - [6 \; SUM(i=1,n) \; d_i^2] \; / \; [\; n^3-n \;].$$

In animal breeding rank correlations are used to compare methods of sire evaluations.

III.7 Methods for the Analysis of the Error Variance-Covariance Structure

Procedures to construct the variance-covariance matrix of the three lactations are outlined in III.2.

Solutions for the model were stored on disk. Estimates of $\hat{y}_{ij}$ were obtained for milk yield of lactation i^{th} in j^{th} cow (i.e., y_{ij}) from the stored solutions.

Individual residuals were estimated by subtracting the estimates from the observations:

$$\hat{a}_{ij} = y_{ij} - \hat{y}_{ij}$$

where:

$\hat{a}_{ij}$ is the estimated residual of lactation i^{th} in the jth cow.

The estimated residuals for each lactation were stored in vectors. The matrix (a vector for each of the three lactations) of residuals was squared and multiplied directly to form a sum of squares and cross products matrix (SSCP). The SSCP matrix was divided by degrees of freedom, which is the number of observations minus the number of fixed effects fitted into the model, that is, covariable days in milk and corresponding herd-year-seasons. Once divided the matrix SSCP becomes the variance-covariance

matrix S. Direct multiplication and squaring was a shortcut to avoid matrix multiplication of the three vectors of the residuals matrix.

III.7.1 Test for Homogeneous Variance

A procedure to examine the assumption of homogeneous variance and common covariance is described by Gill (1978). This procedure was developed to study repeated measurement designs. One can think of the three lactational milk yields per cow as repeat measurements. To better understand the procedure let's follow a formal explanation.

Among r animals measured, one may compute p different variances ($S^2 = S_{ii}$, $i = 1, 2, ..., p$), one for each of the p periods of measurements. One also may compute a covariance ($S_{ii'}$, $i \neq i'$) for any two periods. Because $S_{12} = S_{21}$, $S_{13} = S_{31}$, there are only $p(p-1)/2$ different covariances. The p variances must be estimates of a common population variance "V" and the $p(p-1)/2$ covariances must be estimates of a common population covariance "VV". If the population variances are equal, any population covariance is equal to the common correlation in the population "rho" multiplied by the common variance, i.e., "rhoV". From the sample of data, one may estimate V from the average sample variance, $\bar{s} = SUM\ s^2_i/p$, and rhoV from the average sample covariance, $s_{ij} = SUM\ SUM\ s_{ii'}/(p(p-1)/2)$. The p x p matrix

of common variances and covariances of the population is an "ideal" matrix $SIGM_0$ to which the real population variance-covariance matrix SIGM may or may not conform to the hypothesis $Ho:SIGM=SIGM_0$. The sample variance-covariance matrix S is an estimate of SIGM, and the matrix of averaged sample variance- covariance S_0 is an estimate of $SIGM_0$ if the hypothesis is true:

$$SIGM = \begin{vmatrix} V_1 & VV_{12} & \cdots & VV_{1p} \\ VV_{21} & V_2 & \cdots & VV_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ VV_{p1} & VV_{p2} & & V_3 \end{vmatrix} \quad SIGM_0 = \begin{vmatrix} V & rhoV & \cdots & rhoV \\ rhoV & V & \cdots & rhoV \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ rhoV & rhoV & & V \end{vmatrix} .$$

And the sample matrices would be:

$$S = \begin{vmatrix} s_1 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2 & \cdots & s_{2p} \\ s_{p1} & s_{p2} & \cdots & s_p \end{vmatrix} \quad S_0 = \begin{vmatrix} s_i & s_{ii'} & \cdots & s_{ii'} \\ s_{ii'} & s_i & \cdots & s_{ii'} \\ s_{ii'} & s_{ii'} & \cdots & s_i \end{vmatrix} .$$

The test for the $Ho:SIGM = SIGMA_0$ used in this study was developed by Box (1949, 1950). The statistics for an experiment with r subjects, each measured for y at p times

are:

$$h_1 = [p(p+1)^2 (2p-3)]/[6(r-1)(p-1)(p^2+p-4]$$

$$h_3 = [-(r-1) \log_e (|S|/|S_0|)$$

where $|S|$ and $|S_0|$ are determinants of the matrices S and S_0. For $p \leq$ the test statistics is $q = (1-h_1)h_3$ versus the table Chi Square statistic with v degrees of freedom, $v_1 = (p^2 + p-4)/2$.

The hypothesis is rejected if this statistic exceeds the critical value for alpha, value of the test.

III.7.2 Method to Test for the Hypothesis that Errors Follow a First Order Autoregressive Model

A maximum likelihood procedure was used to test the hypothesis of Mansour, Nordheim, and Rutledge (1985). They stated that the errors of consecutive lactation milk yields follow a first order non-stationary autoregressive process.

The objective of the test is to obtain from the variance-covariance matrix calculated from data S, estimates of Va, Ve, $\emptyset$ the model parameters. The estimates were used recalculate a matrix SIGMAT. This was used to test the hypothesis (i.e., Ho:$SIGM_u$ = SIGMAT) with Chi Square procedure.

The estimate of $\hat{\emptyset}$ the autoregressive parameter and $\hat{R}$ =

$\hat{V}a/\hat{V}e$ were fitted with an iterative routine used to maximize this three statistics simultaneously.

The derivation of the method is as follows; the autoregressive model can be expressed:

Let the errors e_1, e_2, e_3 be independent and each normally distributed with mean zero and variance Ve ~N(0,Ve).

Let the observed residuals be:

$$u_1 = e_1 + a$$

$$u_2 = \emptyset\ u_1 + e_2 + a$$

$$u_3 = \emptyset\ u_2 + e_3 + a = \emptyset^2\ e_1 + \emptyset\ e_2 + e_3 + a$$

where :

    a   is the random effect of the cow and is assumed independently and identically distributed (iid). N(0, Va),

    Va   variance attributed to genetic and permanent environmental factor.

    Ve   variance attributed to temporary environmental factors.

then:

$$u = \begin{vmatrix} u_1 \\ u_2 \\ u_3 \end{vmatrix} \sim N(0, SIGM_u).$$

The vector of residuals corresponding to the three lactation periods is assumed to be multivariate normally distributed with mean zero and variance $SIGM_u$, the variance-covariance assuming first order autoregression of error where:

$$SIGM_u = \begin{vmatrix} Va+Ve & Va+\emptyset Ve & Va+\emptyset^2 Ve \\ Va+\emptyset Ve & Va+Ve+\emptyset^2 Ve & Va+\emptyset Ve+\emptyset^3 Ve \\ Va+\emptyset^2 Ve & Va+\emptyset Ve+\emptyset^3 Ve & Va+\emptyset^2 Ve+\emptyset^4 Ve \end{vmatrix}.$$

Let $R = Va/Ve$, then:

$$SIGM_u = Ve \begin{vmatrix} R+1 & R+\emptyset & R+\emptyset^2 \\ R+\emptyset & R+\emptyset^2+1 & R+\emptyset+\emptyset^3 \\ R+\emptyset^2 & R+\emptyset+\emptyset^3 & R+\emptyset^2+\emptyset^4+1 \end{vmatrix} = VeA_\emptyset.$$

The log of the likelihood function for n independent observations $u_1, \ldots, u_n$ on u is

Under Ho: that Mansour model holds

$$\log L = -1/2N \log |SIGM_u| - 1/2 \ SUM \ u_i \ SIGM^{-1} u_i$$

$$= -1/2N [\log |A_\emptyset| + 3 \log Ve + \{tr(A_\emptyset^{-1} S)\}/Ve]$$

This is maximized for Ve = $\hat{V}e$ = 1/3 tr($A_{\emptyset}^{-1}$ S), where: $\emptyset$
and R are chosen to maximize:

log L = -1/2N [log$|A_{\emptyset}|$ + 3 log {1/3 tr($A_{\emptyset}^{-1}$ S)} +3]

= -1/2N [log$|A_{\emptyset}|$ + 3 log tr ($A_{\emptyset}^{-1}$S)] - 1/2N [-3 log3+3].

Then test with Chi square statistic:

Q = N [$\log_e$ $|A_{\emptyset}|$ + 3 $\log_e$ tr($A_{\emptyset}^{-1}$ S) - ($\log_e$ $|S|$ +

3 $\log_e$ 3)] .

is asymptotically distributed as Chi Square$_{n-3}$ = Chi Square$_3$
under the null hypothesis, Ho: SIGM$_u$ = SIGMAT. The
hypothesis is rejected if the appropiate test statistics
exceeds the critical value for alpha value of the test.

IV RESULTS AND DISCUSSION

IV.1 Basic Parameters of the Model for each Lactation

The first results of this study were those from the fitting of the model to each lactation. Results, reported in Table 4.1, indicate the characteristics of the model for each lactation.

Table 4.1 Basic parameters of the model for each lactation.

Parameter	Lactation		
	1	2	3
Number of records	3989	3985	3990
Number of sires	193	193	193
Number of herd-year-seasons	361	357	371
R^2	.97	.97	.97

Numbers of records used in the analysis of the three lactations differ because observations belonging to a single herd-year-season group were deleted during the program execution. This difference is small, therefore, numbers of

observations were equal for the purpose of this study.

The number of sires in the analysis was the same in each lactation. This indicates that the number of daughters per sire was enough to make up for the loss of those single herd-year-season observations. The number of herd-year-seasons absorbed into the sire and covariate effect was also similar for each lactation. Because the number of observations and classes are similar for each lactation, it was assumed that there was not a major difference in the of the analysis for each of the lactations. Similar size of incidence matrices and equations to be absorbed seemed to support this assumption.

The coefficient of determination R^2 which measures the proportion of reduction of the variation of milk yield achieved by the model was high and identical for each lactation. This indicates the adequacy of the model in describing causes of variation for each lactation.

IV.2 Variance Components and Heritability Estimates

Means and coefficients of variation for milk yield for the three lactations as well as the iterative REML estimates of variance components and heritability are in Table 4.2.

Table 4.2 Variance components, heritabilities, L ratios, means and coefficients of variation for milk yield in three lactations.

	Lactation		
	1	2	3
Mean of unadjusted milk yield (kg)	4,890	5,520	5,520
Coefficient of variation for unadj. milk yield	.27	.26	.34
Sire Variance (Vs)	837	664	703
Error Variance (Ve)	7,416	10,054	11,469
Final L ratio Ve/Vs	8.86	15.56	16.10
Heritability (h^2)	.40	.23	.23
Standard error h^2	.04	.03	.03

Means for unadjusted milk yields were 4,890 kg for first lactation and 5,520 kg for second and third lactation; this 12.5% increment in milk yield between first and later lactation is expected from the normal rate of maturing of the cows. Production is approximately the same for the Holstein dairy herd in Mexico as reported by McDowell et al. (1976b) and Cisneros et al. (1980).

Variability of milk yield expressed as the coefficient of variation was similar for the first and second lactation and higher for the third; this also was expected because as

a cow gets older the probability that environmental factors
affecting its performance is higher. Another explanation
could be that because of the relative higher cost of
registered cattle in Mexico, dairy breeders tend to retain
relatively lower producing cows in the herd for the purpose
of getting an extra valuable offspring out of them. The
latter would explain both the similarity of milk yield from
the second and third lactation, as well as the higher
variability for milk yield in the third one.

Sire variances were lower for second and third lacta-
tions. The explanation that environmental variance
increases throughout the life of a cow overshadowing its
genetic makeup (sire variance), as suggested by Barker and
Robertson (1966) does not hold. Sire variance for the
second lactation is lower than that of third lactation. The
same authors indicated that analyses of Swedish data showed
a decline of sire variance from first to second lactation,
followed by an increase from second to third lactation
(Hansson and During, 1961), in exactly the same manner as in
this study. This would be expected if selection had been
based on factors having a higher environmental than genetic
connection with milk yield; for instance, accidents,
diseases, or low fertility. Another explanation for these
results could be increase of milk yield with age at calving.
Sire differences for rate of maturity and therefore, higher
milk yield in later lactations also could cause sire

variance to increase in later lactations. However, Hansson and During (1961) suggested that environmental disturbances may have a cumulative effect over succeeding lactations causing this difference in sire variances. Differences in sire variance are believed to be caused by the cumulative effect of environment over succeeding lactations.

Heritabilities of milk yield for first, second, and third lactations agree with the literature. Table 4.3 summarizes the heritabilities for three lactations in Holstein cattle. The similarity of heritabilities estimates with those reported is not only in their magnitude but also in their trend. Heritabilities for milk yield in first lactation is higher than those of later lactations. Most authors favor that environmental factors in the lifetime of the cow, such as diseases and management, increase variation of later records. There are other possible explanations for the decline of heritability such as larger effect of genes controlling first lactation than second lactation. The presence of a genetic maternal effect that gradually decreases in importance also might explain the decline. Yet another possible explanation is the presence of constant genetic effect on second lactations which would lower estimates of heritability of second lactations (Butcher and Freeman, 1968).

Recent heritabilities are from maximum likelihood (ML) Rothschild and Henderson (1979), Meyer (1983), and from res-

tricted maximum likelihood (REML) Tong et al. (1979). The study of Powell et al. (1981) uses a set of data similar to that of the present study because the daughter-dam regression procedure used required that both dam and daughter have three lactations.

Heritabilities can not be compared because they are drawn from different populations which are under different climate, management, and selection practices. They also vary in the analytical procedure used to estimate heritabilities. And the trait, i.e., milk yield, was sometimes actual and sometimes adjusted. Adjustments were not always the same. McDowell et al. (1976b) using a subset of the same data used in this study reported a heritability estimate of.12 for milk yield. To estimate this heritability, they used a sire component within herd from Henderson's method 1 (an alternative analytical procedure)

Conclusion from this study is that because one is able to observe an increase of environmental variation, heritabilities would tend to be lower in later lactations. Cumulative effects of injury or mastitis would be consistent with this viewpoint.

Table 4.3   Heritabilities for milk yield of the first three
lactations in Holstein cattle.

Source and Method	Lactation		
	1	2	3
Freeman (1960) daugher-dam regression	.36	.24	.26
Molinuevo and Lush (1964) daughter-dam regression	.30	.09	.08
Barker and Robertson (1966) paternal half sisters	.35	.24	.23
Butcher and Freeman (1968) intrasire regression of daughter on dam	.37	.25	.35
Butcher and Freeman (1969) Fiveprocedures	.17 to .39	.11 to .35	.11 to .40
Maijala and Hanna (1974) paternal half sisters	.26	.20	.17
Rothschild and Henderson (1979b) paternal half sisters	.41	.35	--
Tong et al. (1979) paternal half sisters	.26	.19	.17
Powell et al. (1981) daughter-dam regression on Modified Contemporary deviations	.36	.31	.26
Meyer (1983) paternal half sisters	.33	.34	.28
This study paternal half sisters	.40	.23	.23

IV.3 Effect of Covariable Days in Milk

Data were not adjusted to the customary 305 days lactation but actual days in lactation were fit linearly as a covariate with milk yield. Fitting one regression coefficient per lactation to correct for the systematic effect of lactation length allowed full expression of variation.

Regression coefficients of days in milk on milk yield for first, second, and third lactations were $1.36 \pm .06$, $2.33 \pm .08$ and $2.16 \pm .03$. The magnitude of the coefficients is associated with milk production in each lactation; the coefficients for the second and third lactation are almost a unit greater than that of the first lactation

Meyer (1983) also fitted linearly lactation length as one of the regression coefficients in her model; however, she did not report the regression coefficients.

IV.4 Sire Solutions for Each Lactation

Solutions for the 193 sires were computed for each
lactation. Summary of these solutions is in Table 4.4.
The mean of BLUP sire solutions are zero by definition.

Table 4.4 Range and standard deviation of BLUP sire
          solutions (n = 193) each lactation.

	Lactation		
	1	2	3
Range	-90 to 77	-46 to 42	-46 to 45
Standard deviation	20.88	16.00	16.51

The range of the BLUP sire solutions is wider for first
lactation than for second and third. This variability also
could be appreciated through the standard deviations. The
standard deviation for the BLUPs for first lactation is
almost 25% higher than for the others. Lofgren et al.
(1983) also reported a 5% higher standard deviation of BLUPs
for first lactation.

Results of the Spearman's Rank correlation for the sire
solutions are in Table 4.5. The range of rank correlations
between sires with daughters in different lactations, i.e.,
only first lactation or only second lactation, reported in
the literature goes from .64 (Tomaszewski et al., 1975) to

.95 (Lofgren et al., 1983). Most of the estimates are in the upper part of this range. Rank correlations from this study are low (Table 4.5).

Schneeberger (1982) reported a low rank correlation between first and second lactation BLUP sire values for Holsteins in Venezuela. One of his explanations for this was that when the environmental fluctuations are larger and number of progeny per sire is low, variation of sire values is high. McDowell (1983) also agreed with this explanation.

Because the primary objective of this study was not to estimate sire values, sires with as few as six progeny remained in the data. If the objective of this study had been to estimate sire values accurately, a larger number of progeny per sire would have been required. In the data set, the majority of bulls had between 6 and 14 progeny. The low number of progeny per sire is probably the explanation for the low rank correlation and the large standard deviation and range the sire values.

Table 4.5   Rank correlation* of BLUP sire solutions for
            the three lactations.

	Lactation	
	First	Second
Second Lactation	.58	
Third Lactation	.53	.59

*Spearman's "r"

IV.5 Analysis of the Variance-Covariance Matrix

The variance-covariance structure from this study is:

$$
S = \begin{vmatrix}
7,416 & 3,991 & 3,114 \\
(3989) & (3950) & (3954) \\
3,991 & 10,541 & 4,813 \\
(3950) & (3985) & (3941) \\
3,114 & 4,813 & 11,469 \\
(3985) & (3941) & (3990)
\end{vmatrix}.
$$

Numbers within parenthesis correspond to the number of observations.

As explained earlier, not all cows were in each lactation because of elimination of single herd-year-seasons during the program execution. Therefore, the degrees of freedom were calculated from each model for each element of the S matrix. For the case of diagonal elements, these are the same as for the error variance component, that is, the total number of observations minus the number of fixed effects fitted into each model. In this case, fixed effects are covariable days in milk, and number of herd-year-seasons. For the off-diagonal elements, degrees of freedom were calculated from the number of products minus the average number of herd-year-seasons in the two involved lactations. The difference of degrees of freedom was small. The degrees of freedom were 3,627 for first and second lactations and 3,618 for third lactations. For statistical

testing purposes, it was decided to use 3,985 as the number of observations because it represented a midpoint between the several numbers of observations.

IV.5.1 Test for Homogeneous Variance

The purpose of this procedure was to test the null hypothesis H:SIGM =SIGM$_0$ by the method of Box (1949, 1950), or in this case H: S = S$_0$. The test consisted in obtaining an estimate S$_0$ by separately averaging variances and covariances in S. Then the likelihood ratio criterion,i.e., of the ratio of the determinants of matrices S and S$_0$, L = |S| / |S$_0$|,was used.

The average variance-covariance matrix S$_0$ from our data was:

$$S_0 = \begin{vmatrix} 9,809 & 3,973 & 3,973 \\ 3,973 & 9,809 & 3,973 \\ 3,973 & 3,973 & 9,809 \end{vmatrix}$$

- repeated measurements p = 3

- total of animals r = 3,985 measured

h_1 = [3 (4^2) (3)]/[6 (3984) 2 (8)]

h_1 = .0004

h_3 = [-3984 log 3 {| 5.59 x 10^{11} | / | 6.04 x 10^{11} |}

h_3 = 308

v_1 = (9 + 3 - 4)/2 = 4

$q = (1 - .0004) (308) = 307$

$q = 307$ which by far exceeds

Chi square $_{0.01, 4} = 13.27$.

Therefore, one should conclude that variance covariance matrix S is heterogeneous.

Examining the estimated variance-covariance structure one can note increasing values along principal diagonal (7416, 10541, 11469), increasing values on the first diagonal (3991, 4813), and decreasing values in the first row (7416, 3991, 3114). It appears that the error variance increased with time and that the error covariances are larger in adjacent lactations. This pattern is similar to that of the error variance-covariance matrix of lactational yield of milk fat reported by Butcher and Freeman (1968; 1969) and used by Mansour, Northeim, and Rutledge (1985) in a first order autoregressive model. Other estimates of error variance (Ve) for first, second, and third lactations also show an increase for second and third lactations. Tong et al. (1979) reported first, second, and third lactation (REML) Ve estimates of 543,531; 715,826 and 824,750. This is an increase of 30% for the second and of 50% increase for the third with respect to the first. Rothschild and Henderson (1979a) reported (ML) Ve estimates of 900,581 and 1,032,391 for first and second lactation this 14% increase is the smallest of those reported. Meyer (1983) reported (REML) Ve estimates of 557.31, 928.51 and 1008.67; the

percent increase in variance was 60% from first to second and 80% from first to third. The percent Ve increase in this study was 42% from first to second and 52% from first to third. These increases are larger than those reported by Tong et al. (1979) and smaller than those reported by Meyer (1983). The Ve increase from second to third lactation is less than from first to second lactation. The percentage increase from second lactation to third was 8% from Meyer (1983) and this study, and 15% from Tong (1979). This smaller difference in Ve between second and third lactation compared to Ve between first and later lactations seems to indicate that environment may have a different influence in these lactations than in the first. Deaton and McGilliard (1964) Mansour, Nordheim, and Rutledge (1985), and Nicholson et al. (1978) have suggested treating first lactations differently from later lactations.

The Box test for homogeneous variance was not used to test the other estimates of variance-covariance matrices because they involved different cows and different numbers of cows at each lactation.

Because Rothschild and Henderson (1979a) did not study third lactations and Tong et al. (1979) assumed that no error covariance existed in their multiple trait model it was not possible to compare their results with this study directly. Results could be compared only with those of Meyer(1983). The error covariance was highest between

second and third lactation in both studies, and they are almost the same between first and second, and first and third for Meyer (1983). The present study error covariances are greater for adjacent lactations.

The variance-covariance matrix estimated in this study is heterogeneous. It also appears that other variance-covariance matrices in the literature follow a similar pattern, even those from studies in which the selection effect has been removed (Tong et al., 1979; Rothschild and Henderson, 1979a; Meyer, 1983). Implications of heterogeneous variance-covariances for the three lactations are several: (1) First lactations should be treated differently from later lactations for example by the use of an index weighting lactations by their heritabilities; (2) Later lactations should be used in sire evaluation systems to account fully for error variation of daughters; (3) Transformation of the data to reduce the heterogeneity should be used to be able to use all the records of daughters of a sire. This is important for example, for small populations such as in less developed countries where all records available should be used to improve the accuracy of the sire evaluations.

The pattern followed by the error variance-covariance matrix from this study may induce one to think that the errors follow a non-stationary autoregressive random process as assumed by Mansour, Nordheim, and Rutledge (1985). One

of the could be modeled after this pattern.

IV.5.2 Test for the Hypothesis That Errors Follow A First Order Autoregressive Model

The purpose of this procedure is to test the hypothesis that the errors of consecutive lactational milk yield follow a first order non-stationary autoregressive process.

For:

$$S = \begin{vmatrix} 7,416 & 3,991 & 3,114 \\ 3,991 & 10,541 & 4,813 \\ 3,114 & 4,813 & 11,469 \end{vmatrix}.$$

The tr $(A_\emptyset^{-1}S)$ is minimized for $\emptyset = \hat{\emptyset} = .414$ and $\hat{R} = Va/Ve = .14$

being tr $(A_\emptyset^{-1}S) = 23,377$.

$\hat{V}e = 1/3(\text{tr}(A_\emptyset^{-1}S) = 7,792;$

therefore,

$\hat{V}a = \hat{R}\ \hat{V}e = (.14)\ (7,792) = 1,090.$

$$SIGM = Ve \begin{vmatrix} R+1 & R+\emptyset & R+\emptyset^2 \\ R+\emptyset & R+\emptyset+1 & R+\emptyset+\emptyset^3 \\ R+\emptyset^2 & R+\emptyset+\emptyset^3 & R+\emptyset^2+\emptyset^4+1 \end{vmatrix} = VeA_\emptyset$$

$$A_{\hat{\beta}} = \begin{vmatrix} 1.14 & 0.554 & 0.311 \\ .554 & 1.3113 & 0.624 \\ 0.311 & 0.624 & 1.340 \end{vmatrix}$$

$|A_{\hat{\beta}}| = 1.235.$

In this case the estimate of $SIGM_u$ is SIGMAT $= \hat{V}e \; A_{\hat{\beta}}$.

$$SIGMAT = \begin{vmatrix} 8,883 & 4,317 & 2,426 \\ 4,317 & 10,218 & 4,869 \\ 2,426 & 4,869 & 10,447 \end{vmatrix}.$$

The best fit is the one that makes SIGMAT as close as possible to S.

The Chi square test for our estimates $\hat{V}e$, $\hat{V}a$ and $\hat{\beta}$ is

$$Q = N \; [\log_e |A_{\beta}| + 3 \log_e \text{ tr } (A_{\beta}^{-1}S) - (\log_e |S| + 3 \log_e 3)]$$

$$N = 3985$$

$$= 3985 - [0.211 + 30.17 - (27.05 + 3.295)]$$

$$= 3985 \; [30.38 - 30.28] = 404.47$$

$$Q = 404.47 \text{ for 3 df.}$$

Therefore, reject the Ho : $SIGM_u = SIGMAT$ at any reasonable alpha.

Before rejecting the assumption that errors follow a
first order autoregressive process, one should compare the
error variance-covariance matrix from this study S with that
estimated with $\hat{V}a$, $\hat{V}e$, and $\hat{\emptyset}$ obtained from the likelihood
procedure **SIGMAT**. In general, **SIGMAT** resembles S because
elements of the main diagonal increase with time and
covariances from adjacent lactations are higher than those
from nonadjacent lactations. The difference between these
two matrices is that elements in **SIGMAT** corresponding to
first lactation seem to be higher and elements corresponding
to the third lactation are lower than in S. The test
statistic rejects the hypothesis that errors follow a non-
stationary first order autoregressive process; however,
given the large number of observations it is expected to
reject statistically nearly any hypothesis.

One should look for a model that could explain better
the pattern for the estimated variance-covariance matrix S.
The reason the elements corresponding to the third lactation
were underestimated may have had something to do with a
change of the real ability of the cow in later lactations.
Age affects milk yield. The age or maturing effect usually
is corrected in most sire evaluation programs to enable
comparisons of yields of cows at a common age. It is usual
to express the production of a cow in mature equivalent
units. For this reason, it was decided to incorporate an
adjustment parameter into the model. This parameter θ

which accounts for 1 degree of freedom represents the possible increment in Va caused by the maturing effect of the cow.

The hypothesis to be tested was that errors follow a first order non-stationary autoregressive model and that the Va in later lactations is also affected by θ or the maturing effect of the cow.

Let the observed residuals be:

$$u_1 = e_1 + a$$

$$u_2 = \emptyset u_1 + e_2 + \theta a$$

$$u_3 = \emptyset u_2 + e_3 + \theta^2 a = \emptyset^2 e_1 + \emptyset e_2 + \theta^2 a + e_3$$

where:

θ = parameter associated to the cow effect, representing the increase in milk yield in later lactations or maturing parameter,

$SIGM_\theta$ will have this form:

$$SIGM_\theta = \begin{vmatrix} Va + Ve & \emptyset\,Ve + \theta Va & \emptyset^2\,Ve + \theta^2\,Va \\ & Ve + \theta^2\,Va + \emptyset^2\,Ve & \emptyset^3\,Ve + \emptyset Ve + \theta^3\,Va \\ Sym & & Ve(1+\emptyset^2 + \emptyset^4) + \theta^4 Va \end{vmatrix}$$

Let R = Va/Ve

$$\text{SIGM}_\theta = \text{Ve} \left| R \right| \left| \begin{array}{ccc} 1 & \theta & \theta^2 \\ \theta & 1+\theta^2 & \theta^3 \\ \theta^2 & \theta^3 & 1+\theta^4 \end{array} \right| + \left| \begin{array}{ccc} 0 & \phi & \phi^2 \\ \phi & \phi^2 & \phi+\phi^3 \\ \phi^2 & \phi+\phi^3 & \phi+\phi^4 \end{array} \right| = \text{VeA}_{\phi\theta}$$

The tr $(A^{-1}_{\phi\theta}S)$ is maximized for $\hat\theta = 1.27$, $\hat\phi = .23$, $\hat{R} = .26$

$\hat{V}e = 6,598$, $\hat{V}a = 1,715$.

Estimate of SIGM is $\text{SIGMAT}_\theta = \text{Ve } A_{\phi\theta}$

$$\text{SIGMAT}_\theta = \left| \begin{array}{ccc} 8,313 & 3,696 & 3,116 \\ 3,696 & 9,714 & 5,112 \\ 3,116 & 5,112 & 11,428 \end{array} \right|$$

Q = 95.72 for 2 df.

Reject the Ho: $\text{SIGM}_u = \text{SIGMAT}_\theta$ at any reasonable alpha.

The model that incorporates the θ parameter to account for the maturing effect of the cow seems to model the pattern of the variance-covariance matrix better. The test statistic to reject this hypothesis is much lower than that to reject the first order autoregressive assumption(404.47 vs 95.72). This indicates that SIGMAT from the θ models is considerably closer to S (specially in the covariances),

SIGMAT from θ model has larger estimates for elements corresponding to the third lactation. The matrices S and SIGMAT from both models are in Table 4.6 .

Table 4.6 Sample S and Variance-Covariance matrices from first order autoregressive plus cow's maturing effect models. SIGMAT$_\emptyset$ and SIGMAT$_\theta$.

$$S = \begin{vmatrix} 7{,}416 & 3{,}991 & 3{,}114 \\ 3{,}991 & 10{,}541 & 4{,}813 \\ 3{,}114 & 4{,}813 & 11{,}469 \end{vmatrix}$$

$$SIGMAT_\emptyset = \begin{vmatrix} 8{,}883 & 4{,}317 & 2{,}426 \\ 4{,}317 & 10{,}218 & 4{,}869 \\ 2{,}426 & 4{,}869 & 10{,}447 \end{vmatrix}$$

$$SIGMAT_\theta = \begin{vmatrix} 8{,}313 & 3{,}696 & 3{,}116 \\ 3{,}696 & 9{,}714 & 5{,}112 \\ 3{,}116 & 5{,}112 & 11{,}428 \end{vmatrix}$$

Though, the θ model is a better approximation to the variance-covariance matrix from this study, it is not possible to obtain a perfect fit without incorporating other parameters and saturating the model. These models fit the lactation data better than a model assuming no correlations of errors.

A positive association of errors exists. A biological

justification would be that if a cow produces an unusually good or bad yield initially, it would be natural to think that her manager would have a high or low expectation of her ability and provide sequent care accordingly. Because Ve is large relative to Va, the major impact of the differential management would be reflected by autoregression. Another argument is that effects of accidents or diseases such as an infection could be carried over from one lactation to the next.

The two models presented in this study could be considered adequate, the autoregressive process models the effects for the different lactations; the arbitrary θ model, that included the maturing effect, models the data better but has the disadvantage of the number of unknown parameter (unknown variances and covariances) that would be added to the model.

V SUMMARY AND CONCLUSIONS

Errors from each record of the same cow have been
treated as independent of each other when analyzing later
lactations. Also a new and equal set of environmental
effects affected the performance of the cow during each
productive cycle or lactation.

Need to incorporate more flexibility to these relation
by allowing errors from each lactation to be correlated with
each other and to vary from lactation to lactation in
accordance with a first order non-stationary autoregressive
process has been suggested.

The potential to use recursive estimation techniques
later lactations would be possible if residual errors follow
a first order autoregressive pattern. More accurate
estimates of parameters and real ability of the cow could be
obtained from these relationships.

Purposes of this study were to estimate the error
variance-covariance structure of first, second, and third
lactations; test the assumptions that there is homogeneous
error variance and constant covariance among each pair of
lactations; and to examine the hypothesis that errors might
follow a first order non-stationary autoregressive process.

Milk yield records of cows with three consecutive
lactations were used to estimate the variance-covariance

matrix S. Data were from registered Holstein cows in Mexican
DHI.

The model applied to each lactation included variables
representing random effects of sire, fixed effect of herd-
year-season, and days in lactation fitted as a covariable.
Normal equations for sire and covariate days in milk were
set up while herd-year-seasons were absorbed by a looping
process. The normal equations then were transformed into
mixed model equations by incorporating the ratio L (Ve/Vs)
in the submatrix for sires. Solutions from the mixed model
equations were used to back solve for herd-year-season
solutions. Then all solutions were used to compute
residuals. Sire and fixed effects solutions, the trace of
the sire submatrix, the product of the ratio L times the
squared sire solutions, and the trace from the generalized
inverse were used to obtain restricted maximum likelihood
(REML) variance components.The process was iterated until
the estimates converged. A FORTRAN program to perform this
task is in Appendix A.
With the REML estimates of sire and residual variance
components, milk yield heritabilities were computed.
Heritabilities and standard errors wer .40 $\pm$ .04, .23 $\pm$ .03,
and .23 $\pm$ .03 for first, second, and third lactations.
These results agree with those in the literature in
magnitude and trend. The decline of heritability is due to
the increasing environmental factors in the lifetime of a

cow. This causes an increment in variation that reduces the importance of the genetic make-up of a cow i.e., this reduces heritabilities for later lactation.

Solutions for 193 sires were computed for each lactation. The range and standard deviations of the solutions were -90 to 77 ($\pm$ 21), -46 to 42 ($\pm$ 16), and -46 to 45 ($\pm$ 16) for first, second, and third lactations. Spearman's Rank correlations for sire solutions were computed to compare the value of sires in each lactation. The rank correlation was .58 for sire values in first and second, .53 for sire values of first and third and .59 for sire values of second and third lactations.

There is large variability in sire solutions and a low rank correlation of sire values. This is due to the number of progeny per sire.

The S matrix is not homogeneous ($p < 0.01$) according to the Box test for homogeneous variance. In the S matrix, increasing principal diagonal (7416, 10541, 11469), increasing first diagonal (3991, 4813), and decreasing first row (7416, 3991, 3114) are noted. It appears that the error variance increases with time and that the error covariances are higher in adjacent lactations. This pattern induces one to conclude that the errors might follow a first order non-stationary autoregressive process. A maximum likelihood procedure was developed to test the hypothesis that errors follow a first order autoregressive model. Estimates that

maximized the Log likelihood were $\hat{\beta}$ = .414, $\hat{V}e$ = 7,792, and $\hat{V}a$ = 1,090. These estimates were used to recalculate a matrix SIGMAT$_\beta$. The H: S = SIGMAT$_\beta$ was tested with a Chi square statistic. The hypothesis that errors follow a first order autoregressive process (Mansour, Nordheim, and Rutledge, 1985) was rejected (p < .01). The autoregressive model underestimated the error variance in later lactations. A model that incorporated a parameter θ representing the possible increment in Va caused by the maturing effect of the cow was tested. The test is that the errors follow a first order non-stationary autoregressive model and that the Va (genetic and permanent environmental contribution of the cow) in later lactations is also affected by θ or the maturing effect of the cow.

The maximum likelihood estimates were $\hat{\beta}$ = .23 $\hat{\theta}$ 1.27, $\hat{V}e$ = 6,598, and $\hat{V}a$ = 1,715. The hypothesis H:S = SIGMAT$_\theta$ was also tested. Model was rejected (p < .01). SIGMAT$_\theta$ has larger estimates for the elements corresponding to the third lactation. Though the θ model is a better approximation to the variance-covariance matrix S, it is not possible to obtain a perfect fit.

One may conclude that even though these models were rejected statistically, they fitted S better than a model with no correlation of errors i.e., β=0. A biological justification would be that if a cow produces an unusually good or bad yield initially, it would be natural to think

that her caretaker would have a high or low expectation of her ability and provide sequent husbandry accordingly. Because Ve is large relative to Va, the major impact of the different management would be reflected by autoregression. Another argument is that cumulative effects of accidents or diseases would be carried over from one lactation to the next. These two models could be considered adequate, although not in compliance with the assumptions. The autoregressive process models the effects for the different lactations. The arbitrary θ better describes the data but has the disadvantage of the number of unknown parameters (unknown variances and covariances) that would be added to the model. A first order non-stationary process to model the error structure should be used for analises of repeated lactations to reduce variances of estimates of parameters.

LIST OF REFERENCES

LIST OF REFERENCES

Anderson, C.R. 1981. A biometrical and genetic study of Tribolium egg production curies as a model for lactation curves. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Bar-Anan, R. 1975. Relations between first and second lactation characteristics of progeny groups and effects of tandem selection on yield improvement. Anim. Prod. 21:121.

Barker, J.S.F., and A. Robertson. 1966. Genetic and phenotypic parameters for the first three lactations in Friesian cows. Anim. Prod. 8:221.

Box, G.E.P. 1949. A general distribution theory for a class of likelihood criteria. Biometrica 36: 317-46.

_____. 1950. Problems in the analysis of growth and wear curves. Biometrics 6:362-89.

Box, G.E.P., and G.M. Jenkins. 1970. Time Series Analysis, Holden Day, San Francisco.

Butcher, D.F., and A.E. Freeman. 1968. Heritabilities and repeatabilities of milk fat production by lactations. J. of Dairy Sci. 51: 1387.

_____. 1969. Estimation of heritability and repeatability of milk and milk fat with selection effects removed. J. Dairy Sci. 52: 1259.

Cassell, B.G. and, B.T. McDaniel. 1983. Use of later records in diary sire evaluation: A review. J. Dairy Sci. 66:1.

Cassell, B.G., B.T. McDaniel, and H.D. Norman. 1983a. Modified contemporary comparison sire evaluation from first, all, and later lactations. J. Dairy Sci.66:140.

Cassell, B.G., B.T. McDaniel, and H.D. Norman. 1983b. Impact of culling on modified contemporary comparison sire evaluations. J. Dairy Sci. 66:1359.

Cassell, B.G., J.S. Clay and, H.D. Norman. 1985.
Differences in modified contemporary comparison sire
evaluations from first and later lactation by breed.
J. Dairy Sci. 68:1778.

Christian, L.E., D.O. Everson, and S.L. Davis. 1978.
A statistical method for detection of hormone
secretory spikes. J. Anim. Sci. 46:699.

Cisneros, P.G., G.R. Wiggans and, G.J. King. 1980. USDA
summary of 1980 Mexican cow herds averages. Mimeo
of Animal Improvement Programs Laboratory.
Beltsville, MD.

Curnow, R.N. 1961. The estimation of repeatability and
heritability from records subject to culling.
Biometrics 17:553.

Deaton, O.W., and L.D. McGilliard. 1964. First, second
and third record of a cow to estimate superiority
of her daughters. J. Dairy Sci. 47:1004.

Dickerson, G.E. 1969. Techniques for research in
quantitative animal genetics. Techniques and
procedures in animal science research.
In Monograph of Amer. Soc. Anim. Sci.
Champaign, IL. pp. 36.

Durbin, J. 1960. Estimation of parameters in
time-seriesregression models.
J.R. Statis. Soc., B, 22.

Freeman, A.E. 1960. Genetic relationships among the
firstthree lactations of Holstein cows. J. Dairy
Sci. 43:876.

_____. 1973. Age adjustments of production
records: History and basic problems. J. Dairy Sci.
56:941.

Gallant, A.R., and J.J. Goebel. 1976. Nonlinear
regressionwith autocorrelated errors. J. Am.
Stat. Assoc. 71:961.

Gill, J.L., and H.D. Hafs. 1971. Analysis of repeated
measurements of animals. J. Anim. Sci. 33:331.

_____. 1978. Design and Analysis of Experiments in
the Animal and Medical Sciences. Vols. 1 and 2
Iowa State University Press, Ames, IA.

_____. 1979. Combined significance of non-independent test for repeated measurements. J. Anim. Sci. 48:1979.

_____. 1981. Evolution of statistical design and analysis of experiments. J. Dairy Sci. 64:1494.

Gupta, S.D., and M.D. Perlman. 1974. Power of the noncentral F-test: Effect of additional variables on Hotelling's T2-test. J. Am. Stat. Assoc. 69:174.

Hansson, A., and T. Düring. 1961. Reliability of heifer records in progeny testing. K. Lantbr Högskol Annlr. 27:351.

Hargrove, G.L. 1974. Rate of maturity of dairy females. J. Dairy Sci. 57:328.

Harville, D.A. 1977. Maximum likelihood approaches to variance components estimation and to related problems. J. Am. Stat. Assoc. 72:320.

Harville, D.A. 1979. Recursive estimation using mixed linear models with autoregressive random effects. Proceedings of a conference in honor of C.R. Henderson Variance components and animal breeding. L.D. Van Vleck and S.R. Searle (eds.). 157-179. Ithaca: Cornell University Press.

Healy, M.J.R. 1968. Inversion of a positive semi-definite symmetric matrix. J. Appl. Stat. 17:199.

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

Henderson, C.R., and R.L. Quaas. 1977. A comparison of method 3 and maximum likelihood estimators of heritabilities and genetic correlations. Res. ASAS 1977 Mtg.

Hickman, G.G., and C.R. Henderson. 1955. Components of the relationship between level of production and rate of maturity in dairy cattle. J. Dairy Sci. 38:883.

Hillers, J.K., and A.E. Freeman. 1965. Differences between sires in rate of maturity of their daughters. J. Dairy Sci. 48:1680.

Hinks, C.J. 1966. Selection practices in dairy herds. IIselection patterns in later lactations. Anim. Prod. 8:467.

Keown, J.R., H.D. Norman, and R.L. Powell. 1976. Effects of selection bias on sire evaluation procedures. J. Dairy Sci. 59:1808.

Kesner, J.S., E.M. Convey, and C.R. Anderson. 1981. Evidence that estadiol induces the preovulatory LH surge in cattle by increasing pituitary sensitivity to LHRH and then increasing LHRH release. Endocrinology 108:1386.

Lofgren, D.L., B.G. Cassell, H.D. Norman, and B.T. McDaniel. 1983. Effects of culling on sire evaluations by mixed models. J. Dairy Sci. 66:2418.

Lush, J.L. 1945. Animal Breeding Plans. The Iowa State College Press, Ames, Iowa. pp. 171.

Lush, J.L., and Shrode, R.R. 1950. Changes in milk production with age and milking frequency. J. Dairy Sci. 33:338.

Maijala, K., and M. Hanna. 1974. Reliable phenotypic and genetic parameters in dairy cattle. Pages 541-563 in Proc. 1st World Congr. Genet Appl. to Livestock Prod. Madrid.

Mansour, H., E.V. Nordheim, and J.J. Rutledge. 1985. Maximum likelihood estimation of variance components in repeated measures designs assuming autoregressive errors. Biometrics 41:287.

Mao, I.L. 1981. Lecture notes on multi-factor unbalanced data. Michigan State University, E. Lansing.

_____. 1982. Modeling and data analysis in animal breeding. Notes for an inter-nordic post-graduate course. Swedish University of Agricultural Sciences. Uppsala, Sweden.

Meyer, K. 1983. Maximum likelihood procedures for estimating genetic parameters for later lactations of dairy cattle. J. Dairy Sci. 66:1983.

Mohammad, W.A., M. Grossman, and R.D. Shanks. 1985. Algebraic equivalence of matrix inversion, elimination and absorption for use in animal breeding. Am. Stat. 39:124.

Molinuevo, H.A., and J.L. Lush. 1964. Reliability of first, second, and third records for estimating the breeding value of dairy cows. J. Dairy Sci. 47:890.

Morrison, J.L. 1967. <u>Multivariate Statistical Methods</u>. McGraw-Hill, New York, NY.

McDowell, R.E. 1983. Strategy for improving beef and dairy cattle in the tropics. Cornell Intern.Agric. Mimeo. 100.

McDowell, R.E., J.K. Camoens, L.D. Van Vleck, E. Christensen, and E. Cabello Frias. 1976. Factors affecting performance of Holsteins in subtropical regions of Mexico. J. Dairy Sci. 59:722.

Nicholson, H.H., L.R. Schaeffer, E.B. Burnside, and M.G. Freeman. 1978. Use of later records in dairy sire evaluation. Can. J. Anim. Sci. 58:615.

Nelson, C.R. 1973. <u>Applied Time Series Analysis</u>. Holden-Day, Inc., San Francisco.

Pindyke, R.S. and D.L. Rubinfeld. 1981. <u>Econometric models and economic forecast</u>. McGraw-Hill, Inc. New York.

Powell, R.L., H.D. Norman, and R.M. Elliot. 1981. Different lactations for estimating genetic merit of dairy cows. J. Dairy Sci. 64:321.

Robertson, A., and J.S.F. Barker. 1966. The correlation between first lactation milk production and longevity in dairy cattle. Anim. Prod. 8:241.

Rothschild, M.F., and C.R. Henderson. 1979a. Maximum likelihood estimates of parameters of first and second lactation milk records. J. Dairy Sci. 62:990.

Rothschild, M.F., and C.R. Henderson. 1979b. Effects of selection on variances and covariances of simulated first and second lactations. J. Dairy Sci. 62:996.

Schaeffer, L.R. 1982. Notes on linear model theory and Henderson's mixed model techniques. University of Guelph, Guelph Canada.

Schaeffer, L.R., J.W. Wilton, and R. Thompson. 1978. Simultaneous estimation of variance aand covariance components from multitrait mixed model equations. Biometrics 34:199.

Schneeberger, C.P. 1982. Factors affecting the performance of Holstein herd in Maracay, Venezuela. Ph.D. thesis, Cornell University, Ithaca, NY.

Searle, S.R. 1971. Linear Models. John Wiley, New York.

Sokal, R.R., and F.J. Rohlf. 1969. Biometry. W.H. Freeman and Company, San Francisco.

Thompson, R. 1969. Iterative estimation of variance components for non-orthogonal data. Biometrics 25, 767.

Tomaszewski, M.A., B.T. McDaniel, H.D. Norman, and F.N. Dickinson. 1975. Relations between sire summaries of first and second lactations. J. Dairy Sci. 58:116.

Tong, A.K.W., B.W. Kennedy, and J.E. Moxley. 1979. Heritabilities and genetic correlations for the first three lactations from records subject to culling. J. Dairy Sci. 62:1784.

Ufford, G.R., C.R. Henderson, J.F. Keown, and L.D. Van Vleck. 1979. Accuracy of first lactation versus all lactations sire evaluations by best linear unbiased prediction. J. Dairy Sci. 62:603.

Van Vleck, L.D., and C.R. Henderson. 1963. Bias in sire evaluation due to selection. J. Dairy Sci. 46:976.

Walter, J.P., and I.L. Mao. 1985. Multiple and single trait analyses for estimating genetic parameters in simulated populations under selection. J. Dairy Sci. 68:91.

Ware, J.H. 1985. Linear models for the analysis of longitudinal studies. Am. Stat. 39:95.

Wickham, B.W. and C.R. Henderson. 1977. Sire evaluation by second lactation records of daughters. J. Dairy Sci. 60:96.

Wilson, P.D., J.R. Hebel, and R. Sherwin. 1981. Screening and diagnosis when within-individual observations are Markow-dependent. Biometrics 37:553.

APPENDIX A

TABLE A . FORTRAN program for mixed model analysis.


```
      PROGRAM REML(TAPE6,TAPE3,TAPE4,TAPE5,TAPE7,OUTPUT,TAPE2=OUTPUT)
    C ***RESTRICTED MAXIMUM LIKELIHOOD PROGRAM***
C
C         PROGRAM  ABSORSB HYS EQUATIONS USING THE ROW LINE TECHNIQUE
C
C''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''''
C                        VARIABLE LIST
C
C         VARIABLE NAME                DESCRIPTION
C         SIRE                         SIRE
C         HYS                          HERD-YEAR-SEASON GROUP
C         MY                           MILK YIELD
C         CI                           CALVING  INTERVAL
C         DIM                          DIM
C         LHYS                         LAST HERD-YEAR-SEASON
C         NREC                         DUMMY RECORD COUNTER
C         TMYHYS                       TOTAL MILK YIELD EACH HYS
C         INDICE(I,J)                  FUNCTION THAT CONVERTS ROW, COLUMN
C                                      REFERENCES TO HALF-STORED VECTOR REFS
C         A(INDICE(I,J)                VECTOR CONTAINING HALF-STORED Z'Z MATRI
C         AINV(INDICE(I,J)             VECTOR CONTAINING HALF-STORED Z'Z INVER
C         RHS(I)                       RIGHT HAND SIDE
C         RHS1(I)                      NON-ABSORBED OR ORIGINAL RIGHT HAND SID
C         VECT(I)                      VECTOR OF SIRE PROGENY AND
C                                      TOTAL COVARIATES
C         NHYS                         NUMBER OF RECORDS IN EACH HYS
C         NOHYS                        NUMBER OF HERD-YEAR-SEASONS
C         WK(I)                        WORKING AREA A VARIABLE NEEDED BY SUBRO
C                                      LINV4P.  DIMENSION TO RANK OF MATRIX BE
C                                      INVERTED.
C         D(I)                         STANDARD ERRORS OF SIRE SOLUTIONS
C         C(I)                         SIRE SOLUTIONS
C         TOT(I)                       TOTALS VECTOR OF COVARIATES PLUS DEPEND
C
C                        I.-READ VARIABLES
C
C                            I.A -DECLARE AND DEFINE VARIABLES
C
      INTEGER HERD,YEAR,SEA,SIRE,LHERD,LYEAR,LSEA,NREC,NHYS,EOREC
    + ,COW
      REAL MY,DIM,CI
      DIMENSION C(3000,1),AINV(19110),A(19110),RHS(195,1),RHS1(195,1),
    + VECT(195),WK(195),D(195),TOT(3),TOTBAR(3),CONT(3)
      EQUIVALENCE (MY,CONT(1)),(CI,CONT(3)),(DIM,CONT(2))
C
C                            I.B -INITIALIZING TO ZERO
C
```

```
          DATA A/19110*0.0/
          DATA RHS/195*0.0/
          DATA RHS1/195*0.0/
          DATA C/3000*0.0/
          DATA AINV/19110*0.0/
          DATA VECT/195*0.0/
          DATA D/195*0.0/
          EOREC = 0
C                                      ACCUMULATOR OF TOTAL DEP OF FIXED
C                                      FACTORS BEING ABSORBED
          TMYHYS = 0.0
          NHYS = 0
C                                      TOTAL SUMS OF SQUARES FOR DEPEND
          TYY = 0.0
          NREC = 1
C                                      CORRECTION FACTOR USED TO CALCULA
C                                      HERD-YEAR-SEASON SUMS OF SQUARES
          CORFAC = 0.0
          NOHYS = 0
          SUMSIR = 0.0
C                                      COEFFICIENT OF DETERMINATION
          RSQ = 0.0
C                                      VARIANCE OF DEPENDENT VARIABLE
          VARDEP = 0.0
C
C                                      I.C -PARAMETERS TO BE SET
C
          RATIO1 = 15.6
          SETDIF = 00.2
          NSIRES = 193
          NCOV = 1
C                                      SET UP DUMMY VARIABLE EQUAL TO
C                                      NUMBER OF COV PLUS DEPENDENT
          NN = NCOV + 1
          DO 67 I=1,NN
            TOT(I) = 0.0
67        CONTINUE
          NRANK = NSIRES + NCOV
C
C                                      I.D -VARIABLES
C
          LSIRE = NSIRES*(NSIRES+1)/2
          LHALF = NRANK*(NRANK+1)/2
C
          READ(6,99,END=55)HERD,CID,MY,SIRE,CI,DIM,YEAR,SEA,COW
C
          GO TO 22
C
C                                      I.E -READ INPUT
C
C                                      START OF INFINITE LOOP TO READ
C                                      ALL INPUT DATA
```

```
C
100    READ(6,99,END=55)HERD,CID,MY,SIRE,CI,DIM,YEAR,SEA,COW
C         WRITE(2,99)HERD,MY,SIRE,CI,DIM,YEAR,SEA,COW
99     FORMAT(1X,I6,I4,9X,F5.0,I5,F4.0,F4.0,I3,I2,1X,I5)
       NREC = NREC+1
       IF(HERD.NE.LHERD)GO TO 5
       IF(YEAR.NE.LYEAR)GO TO 5
       IF(SEA.NE.LSEA)GO TO 5
C
C                                   II.-FORM INCIDENCE MATRIX
C
C                                      II.A -ACCUMULATE PROCESS
C
22     A(INDICE(SIRE,SIRE)) = A(INDICE(SIRE,SIRE))+1.0
       A(INDICE(NSIRES+1,SIRE)) = A(INDICE(NSIRES+1,SIRE))+DIM
C      A(INDICE(NSIRES+2,SIRE)) = A(INDICE(NSIRES+2,SIRE))+CI
       A(INDICE(NSIRES+1,NSIRES+1)) = A(INDICE(NSIRES+1,NSIRES+1))+
     +     DIM**2
C      A(INDICE(NSIRES+2,NSIRES+2)) = A(INDICE(NSIRES+2,NSIRES+2))+
C    +     CI**2
C      A(INDICE(NSIRES+2,NSIRES+1)) = A(INDICE(NSIRES+2,NSIRES+1))+
C    +     DIM*CI
       RHS1(SIRE,1) = RHS1(SIRE,1)+MY
       RHS1(NSIRES+1,1) = RHS1(NSIRES+1,1)+MY*DIM
C      RHS1(NSIRES+2,1) = RHS1(NSIRES+2,1)+MY*CI
       RHS(SIRE,1) = RHS(SIRE,1)+MY
       RHS(NSIRES+1,1) = RHS(NSIRES+1,1)+MY*DIM
C      RHS(NSIRES+2,1) = RHS(NSIRES+2,1)+MY*CI
C      WRITE(2,70) A(INDICE(SIRE,SIRE)),A(INDICE(NSIRES+2,NSIRES+2))
C    +     ,RHS(SIRE,1)
70     FORMAT(1X,3E16.5)
       VECT(NSIRES+1) = VECT(NSIRES+1)+DIM
C      VECT(NSIRES+2) = VECT(NSIRES+2)+CI
       VECT(SIRE) = VECT(SIRE)+1.0
       NHYS = NHYS+1
       TMYHYS = TMYHYS+MY
C                                      SETTING FLAGS FOR NEXT READ
       LHERD = HERD
       LYEAR = YEAR
       LSEA = SEA
       TYY = MY**2+TYY
       DO 68 I=1,NN
          TOT(I) = TOT(I) + CONT(I)
68     CONTINUE
       K = NOHYS + 1
C      WRITE(2,91)HERD,MY,SIRE,CI,DIM,YEAR,SEA,COW,K
       WRITE(5,91)HERD,MY,SIRE,CI,DIM,YEAR,SEA,COW,K
91     FORMAT(1X,I6,6X,F6.0,1X,I5,1X,2(F5.0,1X),2(I4),2(I5))
       GO TO 100
C
C                                      II.B -ABSORBING HYS
C
```

```
55        EOREC = 1
5         CONTINUE
C         PRINT*,'          ABSORBING  HYS INTO SIRE AND COVARIATES    '
          DO 1 I = 1,NRANK
             DO 2 J = 1,I
                A(INDICE(I,J)) = A(INDICE(I,J))-(VECT(I)*VECT(J))/NHYS
C               WRITE(2,98)A(INDICE(I,J))
98              FORMAT(10X,E16.5)
2            CONTINUE
          RHS(I,1) = RHS(I,1)-VECT(I)*TMYHYS/NHYS
1         CONTINUE
       CORFAC = CORFAC+((TMYHYS**2)/NHYS)
C
C
C                                    II.C -SAVE VECT NHYS AND TMYHYS T
C                                         COMPUTE SOLUTIONS FOR HYS
C
       WRITE(3,*) NHYS,TMYHYS
3000   FORMAT(3X,I8,1X,E15.5)
C
       WRITE(3,*)(VECT(I),I=1,NRANK)
C
3001   FORMAT(1X,8E15.5)
C
C                                    II.D- ZEROING VECT,NHYS,TMYHYS
C                                         BEFORE START READING NEW
C                                         HYS GROUP
C
       NHYS = 0
       TMYHYS = 0.0
       DO 10 I = 1,NRANK
          VECT(I) = 0.0
10        CONTINUE
C
C                                    II.E- COUNTING THE NUMBER OF HERD
C                                         YEAR SEASONS
C
       NOHYS = NOHYS+1
       IF(EOREC.NE.1)GO TO 22
       PRINT*,'         END OF THE ABSORBING PHASE              '
       PRINT*,'*******************************************************'
C
C                                    II.F- END OF ABSORBING PHASE WRIT
C                                         A AND RHS
C
       PRINT*,'*****NO. OF RECORDS***RECORDS USED***COUNT HYS***'
       WRITE(2,702) NREC,K
       PRINT*,'********NUMBER OF HERD-YEAR-SEASONS*****'
       WRITE(2,702)NOHYS
702    FORMAT(5X,I8)
       TOTBAR(1) = TOT(1)/NREC
       VARDEP = (TYY-TOT(1)*TOT(1)/NREC)/(NREC-1)
       PRINT*,'     ROW VARIANCE     '
```

```fortran
      WRITE(2,*)VARDEP
      DO 69 I=2,NN
       TOTBAR(I)=TOT(I)/NREC
69    CONTINUE
      PRINT*,'       MEANS OF DEPEND  AND  COVAR''S '
      WRITE(2,70) (TOTBAR(I),I=1,NN)
C      PRINT*,'        FINAL Z''Z  MATRIX AFTER ABSORBING HYS   '
C         DO 3 I = 1,NRANK
C         WRITE(2,94)(A(INDICE(I,II)),II=1,NRANK)
C94       FORMAT(1X,8E15.5)
C3        CONTINUE
C      PRINT*,'              FINAL  RHS  AFTER ABSORBING  HYS   '
C         DO 410 I = 1,NRANK
C         WRITE(2,97) RHS(I,1),RHS1(I,1)
C97       FORMAT(3X,2E16.5)
C410      CONTINUE
C
C                              III.-START REML PROCESS
C
C                                 III.A- OBTAIN SOLUTIONS ADD
C                                     RATIO TO SIRE DIAGONALS
C
      NDIAG = 0
      ITERA = 0
C         PRINT*,'       SIRE  DIAGONALS + RATIO      '
130      DO 150 I = 1,NSIRES
          IF(A(INDICE(I,I)).EQ.0.0) THEN
             NDIAG = NDIAG+1
             END IF
           A(INDICE(I,I)) = A(INDICE(I,I))+RATIO1
C           WRITE(2,70) A(INDICE(I,I))
150      CONTINUE
      PRINT*,'**** NUM OF SIRES NO PRESENT**    ',NDIAG
700   CONTINUE
C
C                              III.B- INVERT A INTO AINV
C
C      PRINT*,'     VECTOR OF SYMETRIC MATRIX    '
C         DO 401 I = 1,LHALF
C           WRITE(2,104)A(I)
C104        FORMAT(10X,E16.5)
C401      CONTINUE
C      K = 1
C         DO 735 I = 1,NRANK
C           J = I*(I+1)/2
C           PRINT*,' ROW  ',I
C           PRINT 1004,(A(L),L = K,J)
C           K = J+1
C1004       FORMAT(1X,6E16.5)
C735      CONTINUE
      N = NRANK
      CALL LINV4P(A,N,AINV,WK,IER)
```

```
C         PRINT*,'             ZZ +RATIO INVERTED MATRIX HALF STORED     '
C            DO 134 I = 1,LHALF
C               WRITE (2,34)AINV(I)
34             FORMAT (1X,E16.5)
C134        CONTINUE
C
C                                        III.B- RESTORE ORIGINAL A MATRIX
C
           DO 151 I = 1,NSIRES
              A(INDICE(I,I)) = A(INDICE(I,I))-RATIO1
151        CONTINUE
C
C                                        III.C- MULTIPLY AINV BY RHS AND
C                                               OBTAIN SOLUTIONS
C
        M = 1
        IB = NRANK
        IC = NRANK
C
        CALL VMULSF(AINV,N,RHS,M,IB,C,IC)
C
C          PRINT*,'      SOLUTIONS        '
C            DO 5001 I = 1,NRANK
C               WRITE(2,34)C(I,1)
C5001       CONTINUE
C
C                                        IV.A- CALCULATE ERROR VARIANCE
C
C                                               CALCULATE UPU
C
        UPU = 0.0
           DO 1290 I = 1,NSIRES
           UPU = UPU+C(I,1)**2
1290       CONTINUE
        PRINT*,'             U PRIME U          '
        WRITE(2,290)UPU
290     FORMAT(1X,E16.5)
C
C
C                                               CALCULATE THOMPSON
C                                               FACTOR
        THOFAC=RATIO1*UPU
C
        PRINT*,'          THOMPSON FACTOR     '
        WRITE(2,290)THOFAC
C
C
C                                        IV.A.1- MULTIPLY SOLUTIONS BY RHS
C
        BPXY = 0.0
        BPXY1 = 0.0
           DO 1950 I = 1,NRANK
```

```
          BPXY = BPXY+RHS(I,1)*C(I,1)
          BPXY1 = BPXY1+RHS1(I,1)*C(I,1)
1950      CONTINUE
          PRINT*,'        SOLUTIONS TIMES RHS    BPXY AND BPXY1      '
          WRITE (2,950)BPXY,BPXY1
950       FORMAT(1X,2E16.5)
          ERRVAR = TYY-CORFAC-BPXY-THOFAC
          DFERR = NREC-NCOV-NOHYS
          ERRVAR = ERRVAR/DFERR
          PRINT*,'ERRVAR,TOTSS,BPXY,CORFAC'
          WRITE(2,750)ERRVAR,TYY,BPXY,CORFAC
750       FORMAT(1X,4E16.5)
          RSQ=(CORFAC+BPXY+THOFAC)/TYY
          PRINT*,'      RSQ      '
          WRITE(2,*)RSQ
C
C                                             IV.A.2- CALCULATE SIRE VARIANCE
C                                             IV.A.3- CALCULATE TRACE
C
          TRACE = 0.0
          II = 1
          I = 0
600       I = I+II
          TRACE = TRACE+AINV(I)
          II = II+1
          IF(I.EQ.LSIRE)GO TO 890
          GO TO 600
890       CONTINUE
          PRINT*,'             TRACE       '
          WRITE(2,290)TRACE
C
C                                             IV.A.4- CALCULATE SIRE VARIANCE
C
          SIRVAR = TRACE*ERRVAR
          SIRVAR = SIRVAR+UPU
          SIRVAR = SIRVAR/NSIRES
          PRINT*,'    SIRE VARIANCE       ERROR VARIANCE       '
          WRITE(2,292)SIRVAR,ERRVAR
292       FORMAT(5X,E16.5,5X,E16.5)
          HER = (4*SIRVAR)/(SIRVAR+ERRVAR)
          WRITE(2,291) HER
          PRINT*,'*****************'
291       FORMAT ( ' HERITABILITY ' ,F9.4)
C
C                                             IV.A.5- CALCULATE NEW RATIO
C
          RATIO2 = ERRVAR/SIRVAR
C
C                                             IV.A.6- CALCULATE RATIO DIFFERENC
C
          RATDIF = ABS(RATIO1-RATIO2)
          PRINT*,'      RATIO1    NEW RATIO      RATIO DIFFERENCE    '
```

```
        WRITE(2,768)RATIO1,RATIO2,RATDIF
768     FORMAT(1X,3E16.5)
        IF(RATDIF.LE.SETDIF)GO TO 999
        ITERA = ITERA+1
        PRINT*,'****** END OF ITERATION  ***  ',ITERA
        RATIO1 = RATIO2
        GO TO 130
999     CONTINUE
        PRINT*,'*                                              *'
        PRINT*,'***********************************************'
        PRINT*,' THE ITERATIVE PROCESS HAS ENDED  '
        PRINT*,'***********************************************'
        PRINT*,' '
        PRINT*,'** THE RATIO DIFFERENCE IS LESS THAN**  ',SETDIF
        PRINT*,' '
        PRINT*,'  THERE WERE  ',ITERA,' ROUNDS OF ITERATION  '
        PRINT*,' '
        PRINT*, 'SIRE VARIANCE '
        WRITE(2,290) SIRVAR
        PRINT*,'********************'
        PRINT*,' '
        PRINT*,  'ERROR VARIANCE'
        WRITE(2,290) ERRVAR
        PRINT*,'********************'
        PRINT*,' '
        PRINT*,'HERITABILITY'
        WRITE(2,290) HER
        PRINT*,'********************'
        PRINT*,' '
        PRINT*,' '
          DO 728 I = 1,NCOV
            WRITE(2,827) I,A(INDICE(NSIRES+I,NSIRES+I))
827         FORMAT(5X,'SSCOV ',I5,'= ',E16.5)
            PRINT*,'************************'
728       CONTINUE
C     WRITE(2,729) A(INDICE(NSIRES+2,NSIRES+1))
729   FORMAT(5X,'SCPCOV1 COV2 = ',E16.5)
        PRINT*,'************************'
C
C                                      IV.B- CALCULATE STANDARD ERRORS
C                                              OF SOLUTIONS
C
C                                      IV.B.1- CREATE A VECTORS WITH
C                                              INVERSE DIAGONALS
C
        J = 0
          DO 7355 I = 1,NRANK
            J = J+I
            D(I) = AINV(J)
C
C                                      IV.B.2- MULTIPLY DIAG * ERROR VAR
C
```

```
              D(I) = D(I)*ERRVAR
C
C
C                                              IV.B.3- CALCULATE STANDARDS ERROR
C
              D(I) = SQRT(D(I))
7355      CONTINUE
C
C                                      V.- PRINT FINAL RESULTS
C
      PRINT*, ' SOLUTIONS AND STANDARD ERRORS '
      PRINT*,'*********************************'
          DO 5002 I = 1,NSIRES
             SUMSIR = SUMSIR = SUMSIR+C(I,1)
             WRITE(2,2005)I,C(I,1),D(I)
2005         FORMAT(5X,' SIRE',I5,'= ',F9.4,5X,F9.4)
5002      CONTINUE
          DO 5003 I = NSIRES+1,NSIRES+NCOV
             WRITE(2,3005)I,C(I,1),D(I)
3005         FORMAT(5X, 'COVAR',I5, '= ',F9.4,5X,F9.4)
5003      CONTINUE
      WRITE(2,9090) SUMSIR
9090  FORMAT(5X,'THE SIRE SOLUTIONS SUM TO  ',F12.4)
C
C                                      IV.- START THE BACK UP PROCESS TO
C                                            OBTAIN HYS SOLUTIONS
C
      REWIND 3
      SUM = 0.0
      NSOL = NOHYS+NRANK
      BPHY = 0.0
          DO 1001  J = NRANK+1,NSOL
             READ(3,*) NHYS,TMYHYS
             READ(3,*)(VECT(I),I=1,NRANK)
C            WRITE(2,3000) NHYS,TMYHYS
C            WRITE(2,3001)(VECT(I),I=1,NRANK)
                DO  567 I = 1,NRANK
                   SUM = SUM+(C(I,1)*(-1.0*VECT(I)))
C                  WRITE(2,404)C(I,1),VECT(I)
C404               FORMAT(2X,2E15.5)
567             CONTINUE
             C(J,1) = (TMYHYS+SUM)/NHYS
C            WRITE(2,66)SUM,TMYHYS,C(J,1)
66           FORMAT(5X,3E15.5)
             SUM = 0.0
             BPHY = BPHY+(TMYHYS*C(J,1))
1001      CONTINUE
      ERRVAR2 = TYY-BPHY-BPXY1-THOFAC
      ERRVAR2 = ERRVAR2/DFERR
C
      ERRDIF=ERRVAR2-ERRVAR
C
      PRINT*,'  ERROR VAR HYS SOLUTIONS AND ERROR VAR CORR FACTOR '
```

```
      WRITE(2,750) ERRVAR2 ,ERRVAR,ERRDIF
C
8081  PRINT*,'    ** ALL SOLUTIONS INCLUDING THE HYS **  '
      PRINT*,'      ** WERE WRITEN IN TAPE 4 **  '
      DO 76 I = 1,NSOL
C         WRITE(2,34)C(I,1)
          WRITE(4,*)C(I,1)
76        CONTINUE
      STOP
      END
C     **** END OF REML PROGRAM********
C***************SUBROUTINES*******************
      SUBROUTINE VMULSF(A,N,B,M,IB,C,IC)
      DIMENSION A(1), B(IB,M), C(IC,M)
        DO 51 J = 1,M
        DO 51 I = 1,N
          C(I,J) = 0.
          INDEX = I*(I-1)/2
            DO 52 K = 1,I
              INDEX = INDEX+1
52            C(I,J) = C(I,J)+A(INDEX)*B(K,J)
          IF (I.EQ.N) GOTO 51
          K = I+1
          INDEX = INDEX+I
            DO 53 KK = K,N
              C(I,J) = C(I,J)+A(INDEX)*B(KK,J)
53            INDEX = INDEX+KK
51      CONTINUE
      RETURN
      END
C ALGORITHM AS7---J. APPL. STAT. 17,199---HEALY
      SUBROUTINE LINV4P(A,N,AINV,W,IER)
      DIMENSION A(1),AINV(1),W(1)
      NROW = N
      IER = 130
      IF(NROW.LE.0)GOTO 100
      IER = 0
      CALL LUDECP(A,AINV,NROW,D1,D2,IER)
      IF(IER.NE.0)GOTO 100
      NN = NROW*(NROW+1)/2
      IROW = NROW
      NDIAG = NN
16    IF(AINV(NDIAG).EQ.0)GOTO 11
      L = NDIAG
        DO 10 I = IROW,NROW
          W(I) = AINV(L)
          L = L+I
10      CONTINUE
      ICOL = NROW
      JCOL = NN
      MDIAG = NN
15    L = JCOL
```

```
        X = 0.
        IF(W(IROW).EQ.0)GOTO 99
        IF(ICOL.EQ.IROW) X = 1./W(IROW)
        K=NROW
13      IF(K.EQ.IROW)GOTO 12
        X = X-W(K)*AINV(L)
        K = K-1
        L = L-1
        IF(L.GT.MDIAG)L = L-K+1
        GOTO 13
12      IF(W(IROW).EQ.0)GOTO 99
        AINV(L) = X/W(IROW)
        IF(ICOL.EQ.IROW)GOTO 14
        MDIAG = MDIAG-ICOL
        ICOL = ICOL-1
        JCOL = JCOL-1
        GOTO 15
11      L = NDIAG
          DO 17 J = IROW,NROW
          AINV(L) = 0.
17        L = L+J
14      NDIAG = NDIAG-IROW
        IROW = IROW-1
        IF(IROW.NE.0)GOTO 16
100     RETURN
99      IER = 129
        END
C          -----ALGORITHM AS6---J. APPL. STAT 17,195---HEALY
        SUBROUTINE LUDECP(A,UL,N,D1,D2,IER)
C          -----D1,D2 NOT CURRENTLY IMPLEMENTED
        DIMENSION A(1),UL(1)
        DATA ETA/1.E-9/
        IER = 1
        IF(N.LE.0)GOTO 100
        IER = 129
        J = 1
        K = 0
          DO 10 ICOL = 1,N
          L = 0
              DO 11 IROW = 1,ICOL
              K = K+1
              W = A(K)
              M = J
                  DO 12 I = 1,IROW
                  L = L+1
                  IF(I.EQ.IROW)GOTO 13
                  W = W-UL(L)*UL(M)
12                M = M+1
13              IF(IROW.EQ.ICOL)GOTO 14
                IF(UL(L).EQ.0)GOTO 21
                UL(K) = W/UL(L)
                GOTO 11
```

```
21            UL(K) = 0.
11            CONTINUE
14         IF(ABS(W).LT.ABS(ETA*A(K)))GOTO 20
           IF(W.LT.0.)GOTO 100
           UL(K) = SQRT(W)
           GOTO 15
20         UL(K) = 0.
15         J = J+ICOL
10      CONTINUE
        IER = 0
100     RETURN
        END
C
        FUNCTION INDICE(IROW,ICOL)
        INDICE = IROW*(IROW-1)/2+ICOL
        RETURN
        END
*EOS00  Line=584 Sec=1
OK-
```