Linear Models for the Prediction of Animal Breeding Values
Second Edition

Linear Models for the Prediction of Animal Breeding Values

Second Edition

R.A. Mrode, PhD

Scottish Agricultural College
Sir Stephen Watson Building
Bush Estate
Penicuik
Midlothian EH26 0PH
UK

With a chapter contributed by

R. Thompson, Professor

Rothamsted Research
Harpenden
UK
and
Roslin Institute
Midlothian
UK

CABI Publishing

CABI Publishing is a division of CAB International

CABI Publishing
CAB International
Wallingford
Oxfordshire OX10 8DE
UK

Tel: +44 (0)1491 832111
Fax: +44 (0)1491 833508
E-mail: cabi@cabi.org
Website: www.cabi-publishing.org

CABI Publishing
875 Massachusetts Avenue
7th Floor
Cambridge, MA 02139
USA

Tel: +1 617 395 4056
Fax: +1 617 354 6875
E-mail: cabi-nao@cabi.org

A catalogue record for this book is available from the British Library, London, UK.

Library of Congress Cataloging-in-Publication Data
Mrode, R.A.
Linear models for the prediction of animal breeding values /
R.A. Mrode. -- 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-85199-000-2 (alk. paper)
1. Livestock--Breeding--Mathematical models. 2. Livestock--Breeding--Statistical methods. 3. Livestock--Genetics--Mathematical models.
4. Livestock--Genetics--Statistical methods. I. Title.

SF105.M77 2005
636.08′2′015195--dc22

ISBN 0 85199 000 2

Typeset in Melior by AMA DataSet Ltd, UK.
Printed and bound in the UK by Cromwell Press, Trowbridge.

Contents

Preface

Best linear unbiased prediction (BLUP) has become the most widely accepted method for genetic evaluation of domestic livestock. Since the method was first published by Henderson (1949), it has evolved in terms of its application in models for genetic evaluation, from sire, sire and maternal grandsires models in the early years, followed by univariate and multivariate animal models, and more recently to random regression models for the analysis of longitudinal data. Advances in computational methods and computing power have enhanced this development. Currently, most national genetic evaluation systems for several domestic livestock species are based on animal or random regression models using BLUP.

In spite of these developments and the proliferation of information in the literature, there has been no simple and straightforward text on the application of linear models to the prediction of breeding values. Moreover, in developing countries, where access to journals is limited, there is a basic lack of practical information on the subject area. This book has been written, and fully updated since the first edition was published in 1996, with a good balance of theory and application to fill this gap. It places in the hands of the reader the application of BLUP in modelling several genetic situations in a single text. The book has been compiled from various publications and experience gained from several colleagues in the subject area and from involvement in several national evaluation schemes over the last 14 years. Relevant references are included to indicate sources of some of the materials.

Initially, in Chapter 1, the basic model and assumptions governing genetic evaluation are presented together with simple situations involving prediction of breeding values from the records of an individual. This is followed by the introduction and use of selection indices to predict

genetic merit combining information on several traits and individuals. Then the general framework on the application of BLUP in genetic evaluation in a univariate and multivariate situations is presented in Chapters 3 to 5. The simplification of multivariate evaluations by means of several transformations is also examined, followed by a multi-trait across-country model for the analysis of dairy sires across several countries. Maternal trait models are examined in Chapter 6. Random regression models for the analysis of longitudinal data are discussed in Chapter 7, followed by a chapter on incorporating genetic marker information into genetic evaluations, fitting a random model. Non-additive genetic animal models are discussed, with methods for rapidly computing the inverse of the relationship matrices for dominance and epistasis effects. In Chapter 11, the basic concepts for variance component estimation are introduced, followed by the application of the Gibbs sampler in estimation of genetic parameters and evaluations for univariate and multivariate models. Finally, computing strategies for solving mixed model equations are examined, with a presentation of the several formulae governing iterative procedures on the data. A knowledge of basic matrix algebra is needed to understand the principles of genetic evaluation discussed in the text. For the benefit of those not familiar with matrix algebra, a section on introductory matrix algebra has been incorporated as Appendix A. It is also assumed that the reader is familiar with the basic principles of quantitative genetics.

Several examples have been used to illustrate the various models for genetic evaluation covered in the text and attempts have been made to present formulae which explain how the solutions for random and fixed effects in the models were obtained from the mixed model equations. This illustrates for the reader how the various pieces of information are weighted to obtain the genetic merit of an animal under various models.

Every attempt has been made to ensure accuracy of materials in the text. However, in the event of errors being discovered, please inform the author.

Professor Robin Thompson contributed the chapter on estimation of variance components, and reviewed most of the manuscript. His contribution is immensely appreciated. I am greatly indebted to Professor W.G. Hill and Mr G. Swanson for reviewing the manuscript; their comments and suggestions resulted in substantial improvements in the text. Drs P. Visscher, Sue Brotherstone, Victor Olori, Martin Lidauer and Ismo Stranden read specific chapters or sections; I acknowledge their useful suggestions. The assistance of Mr Mark Paget in preparing the graphs in the text is greatly appreciated. In addition, experience gained from working with the late Professors C. Smith and B.W. Kennedy has been valuable in writing this book. I also wish to express my thanks to Professor R.L. Quaas for permission to use information from his unpublished note on the inbreeding algorithm and in addition to the Animal Genetics and Breeding Unit, University of New England, Australia, for allowing me to use some materials from the *BLUP Handbook* for Chapter 2 of the text.

My sincere gratitude to my wife Doris for her immense support and for typing part of the manuscript. Special thanks to Kevwe, Joshua and Esther for their cooperation, especially when I had to take time off to prepare the manuscript, and to many dear friends who were a great encouragement. Finally, to God be all the glory.

R.A. Mrode
Scottish Agricultural College
Edinburgh

1 Genetic Evaluation with Different Sources of Records

The prediction of breeding values constitutes an integral part of most breeding programmes for genetic improvement. Crucial to the accurate prediction of breeding value is the availability of records. In a population, data available at the initial stages are usually on individual animals, which may or may not be related, and later on offspring and other relatives. Thus, initially, the prediction of breeding values may be based on the records of individuals and few relatives. In this chapter the use of individual records and information from other related sources in the prediction of breeding value is addressed. Also the principles for the calculation of selection indices combining information from different sources and relatives are discussed.

1.1 The Basic Model

Every phenotypic observation on an animal is determined by environmental and genetic factors and may be defined by the following model:

Phenotypic observation = environmental effects
+ genetic effects + residual effects

or:

$$y_{ij} = \mu_i + g_i + e_{ij} \qquad [1.1]$$

where y_{ij} is the record j of the ith animal; μ_i refers to the identifiable non-random (fixed) environmental effects such as herd management, year of birth or sex of the ith animal; g_i is the sum of the additive (g_a), dominance (g_d) and epistatic (g_e) genetic values of the genotype of animal i; and e_{ij} is the sum of random environmental effects affecting animal i.

The additive genetic value in the g term above represents the average additive effects of genes an individual receives from both parents and is termed the breeding value. Each parent contributes a sample half of its genes to its progeny. The average effect of the sample half of genes which a parent passes to its progeny is termed transmitting ability of the parent and corresponds to one-half of its additive genetic value. The breeding value of the progeny therefore is the sum of the transmitting abilities of both parents. Since the additive genetic value is a function of the genes transmitted from parents to progeny, it is the only component that can be selected for and therefore the main component of interest. In most cases, dominance and epistasis, which represent intra-locus and inter-loci interactions, respectively, are assumed to be of little significance and are included in the e_{ij} term of the model as:

$$y_{ij} = \mu_i + g_{ai} + e^*_{ij} \qquad [1.2]$$

with e^*_{ij} being the sum of the random environmental effects, dominance and epistatic genetic values. Equation [1.2] constitutes the linear model usually employed in most problems of breeding value prediction in animal breeding. Usually it is assumed that y follows a multivariate normal distribution, implying that traits are determined by infinitely many additive genes of infinitesimal effect at unlinked loci, the so-called infinitesimal model (Fisher, 1918; Bulmer, 1980). Also it is assumed that var(y), var(g_{ai}) and var(e^*_{ij}) are known and that there is no correlation between g_{ai} and e^*_{ij} (cov(g_{ai}, e^*_{ij}) = 0) nor is there any correlation among mates (cov(e^*_{ij}, e^*_{ik}) = 0). Also, μ, which is used subsequently in this chapter to represent the mean performance of animals in the same management group, for instance animals reared under the same management system, of the same age and sex, is assumed known. From [1.2] the problem of predicting breeding value reduces to that of adjusting phenotypic observations for identifiable non-random environmental effects and appropriately weighting the records of animals and their available relatives.

From the earlier explanation, if a_i is the breeding value of animal i, then

$$a_i = g_{ai} = \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + m_i$$

where a_s and a_d are the breeding values of the sire and dam, respectively, and m_i is the deviation of the breeding value of animal i from the average breeding value for both parents, that is, Mendelian sampling. The sampling nature of inheritance implies that each parent passes only a sample one-half of their genes to their progeny. There is, therefore, genetic variation between offspring of the same parents since all offspring do not receive exactly the same genes. Mendelian sampling could be regarded as the deviation of the average effects of additive genes an individual receives from both parents from the average effects of genes from the parents common to all offspring.

The accurate prediction of breeding value constitutes an important component of any breeding programme since genetic improvement

through selection depends on correctly identifying individuals with the highest true breeding value. The method employed for the prediction of breeding value depends on the type and amount of information available on candidates for selection. In the subsequent section prediction of breeding value using different sources of information is discussed. It should be noted that many applications of genetic evaluation deal with the prediction of transmitting ability, usually referred to as predicted transmitting ability (PTA) or estimated transmitting ability (ETA), which is one-half of the predicted breeding value.

1.2 Breeding Value Prediction from the Animal's Own Performance

1.2.1 Single record

When one phenotypic record is the only available information on each animal, the estimated breeding value (a_i) for animal i can be calculated as:

$$\hat{a}_i = b(y_i - \mu) \qquad [1.3]$$

where b is the regression of true breeding value on phenotypic performance and μ, as indicated earlier, is the mean performance of animals in the same management group and is assumed to be known. Thus:

$$\begin{aligned} b &= \text{cov}(a, y)/\text{var}(y) = \text{cov}(a, a + e)/\text{var}(y) \\ &= \sigma_a^2/\sigma_y^2 \\ &= h^2 \end{aligned}$$

The prediction is simply the adjusted record multiplied by the heritability (h^2). The correlation between the selection criterion, in this case the phenotypic value, and the true breeding value is known as the accuracy of prediction. It provides a means of evaluating different selection criteria because the higher the correlation, the better the criterion as a predictor of breeding value. In some cases, the accuracy of evaluations is reported in terms of reliability or repeatability (r^2), which is the squared correlation between the selection criterion and the true breeding value. With a single record per animal, the accuracy is:

$$\begin{aligned} r_{a,y} &= \text{cov}(a, y)/(\sigma_a \sigma_y) \\ &= \sigma_a^2/(\sigma_a \sigma_y) \\ &= h \end{aligned}$$

and reliability equals h^2.

Expected response (R) to selection on the basis of a single record per individual (Falconer and MacKay, 1996) is:

$$R = i r_{a,y}^2 \sigma_y = i h^2 \sigma_y$$

where i, the intensity of selection, refers to the superiority of selected individuals above population average expressed in phenotypic standard deviation.

The variance of estimated breeding value (var($\hat{a}_i$)) is:

$$\begin{aligned}\text{var}(\hat{a}_i) &= \text{var}(by) = \text{var}(h^2 y) \\ &= h^4 \sigma_y^2 \\ &= r_{a,y}^2 h^2 \sigma_y^2 = r_{a,y}^2 \sigma_a^2\end{aligned} \quad [1.4]$$

Example 1.1
Given that the yearling weight of a heifer is 320 kg in a herd with a mean of 250 kg, predict her breeding value and its accuracy if the heritability of yearling weight is 0.45.

From equation [1.3]:

$$\hat{a} = 0.45(320 - 250) = 31.50 \text{ kg}$$

and:

$$r_{a,y} = \sqrt{0.45} = 0.67$$

1.2.2 Repeated records

When multiple measurements on the same trait, such as milk yield in successive lactations, are recorded on an animal, its breeding value may be predicted from the mean of these records. With repeated measurements it is assumed that there is additional resemblance between records of an individual due to environmental factors or circumstances that affect the records of the individual permanently. In other words, there is an additional covariance between records of an individual due to non-genetic permanent environmental effects. Thus the between-individual variance is partly genetic and partly environmental (permanent environmental effect). The within-individual variance is attributed to differences between successive measurements of the individual arising from temporary environmental variations from one parity to the other. The variance of observations (var(y)) could therefore be partitioned as:

$$\text{var}(y) = \text{var}(g) + \text{var}(pe) + \text{var}(te)$$

where var(g) = genetic variance including additive and non-additive, var(pe) = variance due to permanent environmental effect, and var(te) = variance due to random temporary environmental effect.

The intra-class correlation (t), which is the ratio of the between-individual variance to the phenotypic:

$$t = (\text{var}(g) + \text{var}(pe))/\text{var}(y) \quad [1.5]$$

is usually called the repeatability and measures the correlation between the records of an individual. From [1.5]:

$$\text{var}(te)/\text{var}(y) = 1 - t \quad [1.6]$$

With this model, it is usually assumed that the repeated records on the individual measure the same trait, that is, there is a genetic correlation of one between all pairs of records. Also it is assumed that all records have equal variance and that the environmental correlations between all pairs of records are equal. Let $\tilde{y}$ represent the mean of n records on animal i. The breeding value may be predicted as:

$$\hat{a}_i = b(\tilde{y}_i - \mu) \qquad [1.7]$$

where:

$$b = \text{cov}(a, \tilde{y})/\text{var}(\tilde{y})$$

Now:

$$\text{cov}(a, \tilde{y}) = \text{cov}(a, g + pe + \sum te/n) = \sigma_a^2$$

and:

$$\text{var}(\tilde{y}) = \text{var}(g) + \text{var}(pe) + \text{var}(te)/n$$

Expressing the items in terms of equations [1.5] and [1.6]:

$$\text{var}(\tilde{y}) = [t + (1-t)/n]\sigma_y^2$$

Therefore:

$$\begin{aligned} b &= \sigma_a^2/[t + (1-t)/n]\sigma_y^2 \\ &= nh^2/[1 + (n-1)t] \end{aligned}$$

Note that b now depends on heritability, repeatability and the number of records.

As mentioned earlier, the difference between repeated records of an individual is assumed to be due to temporary environmental differences between successive performances. However, if successive records are known to be affected by factors which influence performance, these must be corrected for. For instance, differences in age at calving in first and second lactations may influence milk yield in first and second lactation. Such age differences should be adjusted for before using the means of both lactations for breeding value prediction.

The accuracy of the estimated breeding value is:

$$\begin{aligned} r_{a,y} &= \text{cov}(a, \tilde{y})/(\sigma_a \sigma_y) \\ &= \sigma_a^2/(\sigma_a \sqrt{[t + (1-t)/n]}\sigma_y^2) \\ &= h\sqrt{[n/(1 + (n-1)t)]} \\ &= \sqrt{[nh^2/(1 + (n-1)t)]} = \sqrt{b} \end{aligned}$$

Compared with single records, there is a gain in the accuracy of prediction with repeated records from the above equation, which is dependent on the value of repeatability and the number of records. This gain in accuracy results mainly from the reduction in temporary environmental variance (within individual variance) as the number of records increases.

Table 1.1. Percentage increase in accuracy of prediction with repeated records compared with single records at a heritability of 0.35.

	Number of records				
t values	2	4	6	8	10
0.2	29	58	73	83	89
0.5	15	26	31	33	35
0.8	5	8	10	10	10

When t is low, this gain is substantial as the number of records increases. When t is high, there is little gain in accuracy with repeated records compared with using only single records. The gain in accuracy from repeated records compared with selection on single records can be obtained as the ratio of accuracy from repeated records (r_n) to that from single records (r_k):

$$\frac{r_n}{r_k} = \frac{\sqrt{\dfrac{h^2}{t + \dfrac{(1-t)}{n}}}}{h} = \sqrt{\frac{1}{t + \dfrac{(1-t)}{n}}}$$

Using the above equation, the gain in accuracy from repeated records compared with selection on single records is given in Table 1.1. The increase in accuracy with four measurements at a low t value of 0.2 was 58% but this dropped to only 8% when t equalled 0.8. In general, the rate of increase dropped rapidly as the number of records exceeded four, and it is seldom necessary to record more than four measurements.

Expected response to selection on the basis of mean of repeated records is:

$$R = ih^2_{a,y}\sigma_y\sqrt{[t + (1-t)/n]}$$

Example 1.2

Assume that a cow has a mean yield of 8000 kg of milk for first and second lactations. If the phenotypic standard deviation and heritability of milk yield in the first two lactations are 600 kg and 0.30, respectively, and the correlation between first and second lactation yields is 0.5, predict the breeding value of the cow for milk yield in the first two lactations and its accuracy. Assume that the herd mean for first and second lactations is 6000 kg.

From equation [1.7]:

$$\hat{a}_{cow} = b(8000 - 6000)$$

with:

$b = 2(0.3)/(1 + (2 - 1)0.5) = 0.4$

Therefore:

$\hat{a}_{cow} = 0.4(8000 - 6000) = 800$ kg

and:

$r_{a,\tilde{y}} = \sqrt{0.4} = 0.632$

1.3 Breeding Value Prediction from Progeny Records

For traits where records can be obtained only on females, the prediction of breeding values for sires is usually based on the mean of their progeny. This is typical of the dairy cattle situation, where bulls are evaluated on the basis of their daughters. Let $\bar{y}_i$ be the mean of single records of n progeny of sire i with the assumption that the progeny are only related through the sire (paternal half-sibs), the breeding value of sire i is:

$$\hat{a}_i = b(\bar{y}_i - \mu) \qquad [1.8]$$

where:

$b = \text{cov}(a, \bar{y})/\text{var}(\bar{y})$

Now:

$\text{cov}(a, \bar{y}) = \text{cov}(a, \frac{1}{2}a_s + \frac{1}{2}a_d + \Sigma_e/n)$

where a_s is the sire breeding value and a_d represents the breeding value for the dams. Therefore

$\text{cov}(a, \bar{y}) = \frac{1}{2}\text{cov}(a, a_s) = \frac{1}{2}\sigma_a^2$

Using the same principles as in Section 1.2.2:

$\text{var}(\bar{y}) = [t + (1 - t)/n]\sigma_y^2$

assuming there is no environmental covariance between the half-sib records and t the intra-class correlation between half-sibs is $\frac{1}{4}\sigma_a^2/\sigma_y^2 = \frac{1}{4}h^2$.

Therefore:

$$\begin{aligned} b &= \tfrac{1}{2}\sigma_a^2/[t + (1 - t)/n]\sigma_y^2 \\ &= \tfrac{1}{2}h^2\sigma_y^2/[\tfrac{1}{4}h^2 + (1 - \tfrac{1}{4}h^2)/n]\sigma_y^2 \\ &= 2nh^2/(nh^2 + (4 - h^2)) \\ &= 2n/(n + (4 - h^2)/h^2) \\ &= 2n/n + k \end{aligned}$$

with:

$$k = (4 - h^2)/h^2$$

The term k is constant for any assumed heritability. The weight (b) depends on the heritability and number of progeny and approaches two as the number of daughters increases.

The accuracy of the estimated breeding value is:

$$r_{a,\bar{y}} = \mathrm{cov}(a, \bar{y})/\sqrt{(\mathrm{var}(a)\,\mathrm{var}(\bar{y}))}$$

From the above calculations, this could be expressed as

$$r_{a,\bar{y}} = \frac{\frac{1}{2}h^2\sigma_{\bar{y}}^2}{\sqrt{h^2\sigma_{\bar{y}}^2\left(\frac{1}{4}h^2 + \frac{(1-\frac{1}{4}h^2)}{n}\right)\sigma_{\bar{y}}^2}} = \frac{\frac{1}{2}h}{\sqrt{\frac{1}{4}h^2 + \frac{(1-\frac{1}{4}h^2)}{n}}}$$

$$= \sqrt{\frac{nh^2}{nh^2 + (4 - h^2)}}$$

$$= \sqrt{\frac{n}{n+k}}$$

which approaches unity (one) as the number of daughters becomes large. Reliability of the predicted breeding value therefore equals $n/(n + k)$.

The equation for expected response when selection is based on the mean of half-sibs is the same as that given in Section 1.2.2 for the mean of repeated records but with t now referring to the intra-class correlation between half-sibs.

The performance of any future daughters of the sire can be predicted from the mean performance of the present daughters. The breeding value of a future daughter ($\hat{a}_{daugh.}$) of the sire can be predicted as:

$$\hat{a}_{daugh.} = b(\bar{y} - \mu)$$

with $\bar{y}$ and μ as defined in equation [1.8] and [1.3], respectively, and:

$$b = \mathrm{cov}(a_{daugh.}, \bar{y})/\mathrm{var}(\bar{y})$$

Now:

$$\mathrm{cov}(a_{daugh.}, \bar{y}) = \mathrm{cov}(\tfrac{1}{2}a_s + \tfrac{1}{2}a_{d^*}, \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + \Sigma e/n)$$

where the subscript d^* refers to the dam of the future daughter, which is assumed to be unrelated to dams (d) of present daughters. Therefore:

$$\mathrm{cov}(a_{daugh.}, \bar{y}) = \mathrm{cov}(\tfrac{1}{2}a_s, \tfrac{1}{2}a_s) = \tfrac{1}{4}\mathrm{cov}(a_s, a_s) = \tfrac{1}{4}\sigma_a^2$$

Therefore:

$$b = \tfrac{1}{4}\sigma_a^2/[t + (1-t)/n]\sigma_{\bar{y}}^2$$

Using the same calculations for obtaining b in [1.8]:

$$b = n/n + k$$

The b value is half of the value of b in equation [1.8]; thus the predicted breeding value of a future daughter of the sire is equal to half the estimated breeding value of the sire. The performance of a future daughter of the sire can be predicted as:

$$y = M + \hat{a}_{daugh.}$$

where M is the management mean.

The accuracy of the predicted breeding value of the future daughter is:

$$r_{a,\bar{y}} = \mathrm{cov}(a_{daugh.}, \bar{y})/\sqrt{(\mathrm{var}(a)\,\mathrm{var}(\bar{y}))}$$

This could be expressed as

$$r_{a_{daugh.},\bar{y}} = \frac{\frac{1}{4}h^2\sigma_{\bar{y}}^2}{\sqrt{h^2\sigma_{\bar{y}}^2\left(\frac{1}{4}h^2 + \frac{(1-\frac{1}{4}h^2)}{n}\right)\sigma_{\bar{y}}^2}} = \frac{\frac{1}{4}h}{\sqrt{\frac{1}{4}h^2 + \frac{(1-\frac{1}{4}h^2)}{n}}}$$

$$= \frac{1}{2}\sqrt{\frac{n}{n+k}}$$

which is equal to half of the accuracy of the predicted breeding value of the sire. Reliability of the predicted breeding value equals $1/4n/(n + k)$, which is one-quarter of the reliability of the bull proof.

Example 1.3

Suppose the fat yield of 25 half-sib progeny of a bull averaged 250 kg in the first lactation. Assuming a heritability of 0.30 and herd mean of 200 kg, predict the breeding value of the bull for fat yield and its accuracy. Also predict the performance of a future daughter of this bull for fat yield in this herd.

From [1.8]:

$$\hat{a}_{bull} = b(250 - 200)$$

with:

$$b = 2n/(n + (4 - h^2)/h^2) = 2(25)/(25 + (4 - 0.3)/0.3) = 1.34$$

$$\hat{a}_{bull} = 1.34(250 - 200) = 67\,kg$$

$$r_{a,\bar{y}} = \sqrt{(n/(n + k))} = \sqrt{[25/(25 + (4 - 0.3)/0.3)]} = 0.82$$

The future performance of the daughter of the bull is

$$y = (0.5)a_{bull} + \text{herd mean}$$

$$= 33.5 + 200 = 233.5 \text{ kg}$$

1.4 Breeding Value Prediction from Pedigree

When an animal has no record, its breeding value can be predicted from the evaluations of its sire (s) and dam (d). Each parent contributes half of its genes to their progeny; therefore the predicted breeding value of progeny (o) is:

$$\hat{a}_o = (\hat{a}_s + \hat{a}_d)/2 \qquad [1.9]$$

Let $f = (\hat{a}_s + \hat{a}_d)/2$; then the accuracy of the predicted breeding value is:

$$r_{\hat{a}_o, f} = \frac{\text{cov}(a_o, \frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d)}{\sqrt{\sigma_a^2 \text{ var}(\frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d)}}$$

Now:

$$\begin{aligned}\text{cov}(a_o, \tfrac{1}{2}\hat{a}_s + \tfrac{1}{2}\hat{a}_d) &= \text{cov}(a_o, \tfrac{1}{2}\hat{a}_s) + \text{cov}(a_o, \tfrac{1}{2}\hat{a}_d) \\ &= \text{cov}(\tfrac{1}{2}a_s + \tfrac{1}{2}a_d, \tfrac{1}{2}\hat{a}_s) + \text{cov}(\tfrac{1}{2}a_s + \tfrac{1}{2}a_d, \tfrac{1}{2}\hat{a}_d)\end{aligned}$$

Assuming sire and dam are unrelated:

$$\begin{aligned}\text{cov}(a_o, \tfrac{1}{2}\hat{a}_s + \tfrac{1}{2}\hat{a}_d) &= \tfrac{1}{4}\text{cov}(a_s, \hat{a}_s) + \tfrac{1}{4}\text{cov}(a_d, \hat{a}_d) \\ &= \tfrac{1}{4}\text{var}(\hat{a}_s) + \tfrac{1}{4}\text{var}(\hat{a}_d)\end{aligned}$$

Substituting the solution for the variance of estimated breeding value in equation [1.4]:

$$\text{cov}(a_o, \tfrac{1}{2}\hat{a}_s + \tfrac{1}{2}\hat{a}_d) = \tfrac{1}{4}(r_s^2 + r_d^2)\sigma_a^2$$

From the calculation above, the term $\text{var}(\frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d)$ in the denominator of equation [1.9] is also equal to $\frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2$, assuming random mating and the absence of joint information in the sire and dam proofs. Therefore:

$$r_{\hat{a}_o, f} = \frac{\frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2}{\sqrt{\sigma_a^2 \frac{1}{4}(r_s^2 + r_d^2)\sigma_a^2}} = \frac{\sigma_f}{\sigma_a} = \tfrac{1}{2}\sqrt{(r_s^2 + r_d^2)}$$

where $\sigma_f = \sqrt{[\text{var}(\frac{1}{2}\hat{a}_s + \frac{1}{2}\hat{a}_d)]}$. From the above equation, the upper limit for r when prediction is from pedigree is $\frac{1}{2}\sqrt{2} = 0.7$; that is, the accuracy of the proof of each parent is unity. Note that when the prediction is only from the sire proof, for instance, then $r_{\hat{a}_o, \frac{1}{2}\hat{a}_s} = \frac{1}{2}\sqrt{r_s^2} = \frac{1}{2}\sqrt{n/n+k}$, the accuracy of the predicted breeding value of a future daughter of the sire as shown in Section 1.3.

Expected response to selection on the basis of average proof of parents is:

$$R = ir_{\hat{a}_o, f}\sigma_a$$

Substituting σ_f/σ_a for r:

$$R = i\sigma_f$$

Example 1.4
Suppose that the estimated breeding values for the sire and dam of a heifer are 180 and 70 kg for yearling body weight, respectively. Given that the accuracy of the proofs are 0.97 for the sire and 0.77 for the dam, predict the breeding value of the heifer and its accuracy for body weight at 12 months of age.

From equation [1.9]:

$$\hat{a}_{heifer} = 0.5(180 + 70) = 125\,\text{kg}$$

The accuracy is:

$$r_{\hat{a},a} = 0.5\sqrt{(0.97^2 + 0.77^2)} = 0.62$$

1.5 Breeding Value Prediction for One Trait from Another

The breeding value of one trait may be predicted from the observation on another trait if the traits are genetically correlated. If y is the observation on animal i from one trait, its breeding value for another trait x is:

$$\hat{a}_{ix} = b(y - \mu) \qquad [1.10]$$

with:

$$b = \text{cov}(a_x, \text{measurement on } y)/\text{var}(\text{measurement on } y) \qquad [1.11]$$

The genetic correlation between traits x and y (r_{axy}) is:

$$r_{axy} = \text{cov}(a_x, a_y)/(\sigma_{ax}\sigma_{ay})$$

Therefore:

$$\text{cov}(a_x, a_y) = r_{axy}\sigma_{ax}\sigma_{ay} \qquad [1.12]$$

Substituting equation [1.12] into [1.11]:

$$b = r_{axy}\sigma_{ay}\sigma_{ax}/\sigma_y^2 \qquad [1.13]$$

If the additive genetic standard deviations for x and y in equation [1.13] are expressed as the product of the square root of their individual heritabilities and phenotypic variances, then:

$$\begin{aligned} b &= r_{axy}\sigma_y\sigma_x h_x h_y/\sigma_y^2 \\ &= r_{axy}h_x h_y\sigma_x/\sigma_y \end{aligned} \qquad [1.14]$$

The weight depends on the genetic correlation between the two traits, their heritabilities and phenotypic standard deviations.

The accuracy of the predicted breeding value is:

$$\begin{aligned} r_{ax,ay} &= \text{cov}(a_x, \text{measurement on } y)/\sigma_{ax}\sigma_y \\ &= r_{axy}\sigma_{ay}\sigma_{ax}/(\sigma_{ax}\sigma_y) \\ &= r_{axy}h_y \end{aligned}$$

The accuracy depends on the genetic correlation between the two traits and heritability of the recorded trait.

Correlated response (CR) in trait x as a result of direct selection on y (Falconer and Mckay, 1996) is:

$$\text{CR}x = ih_x h_y r_{axy}\sigma_y$$

Example 1.5

Suppose the standard deviation for growth rate (GR) (g/day) to 400 days in a population of beef cattle was 80, with a heritability of 0.43. The standard deviation for lean growth rate (LGR) (g/day) for the same population was 32, with a heritability of 0.45. If the genetic correlation between both traits is 0.95 and the population mean for growth rate is 887 g/day, predict the breeding value for lean growth rate for an animal with a growth rate of 945 g/day.

Using equation [1.10]:

$$\hat{a}_{\text{LGR}} = b(945 - 887)$$

with:

$$b = \text{cov(GR, LGR)/var(GR)}$$

From [1.13]:

$$b = (0.95(0.656)(0.671)(32))/80 = 0.167$$
$$\hat{a}_{\text{LGR}} = 0.167(945 - 887) = 9.686$$

The accuracy of the prediction is:

$$r = 0.9(\sqrt{0.43}) = 0.623$$

1.6 Selection Index

The selection index is a method for estimating the breeding value of an animal combining all information available on the animal and its relatives. It is the best linear prediction of an individual breeding value. The numerical value obtained for each animal is referred to as the index (I) and it is the basis on which animals are ranked for selection. Suppose y_1, y_2 and y_3 are

phenotypic values for animal *i* and its sire and dam, the index for this animal using this information would be:

$$I_i = \hat{a}_i = b_i(y_1 - \mu_1) + b_2(y_2 - \mu_2) + b_3(y_3 - \mu_3) \quad [1.15]$$

where b_1, b_2, b_3 are the factors by which each measurement is weighted. The determination of the appropriate weights for the several sources of information is the main concern of the selection index procedure. In the above equation, the index is an estimate of the true breeding value of animal *i*.

Properties of a selection index are:

1. It minimizes the average square prediction error, that is, it minimizes the average of all $(a_i - \hat{a}_i)^2$.
2. It maximizes the correlation $(r_{a,\hat{a}})$ between the true breeding value and the index. The correlation is often called the accuracy of prediction.
3. The probability of correctly ranking pairs of animals on their breeding value is maximized.

The b values in equation [1.15] are obtained by minimizing $(a - I)^2$, which is equivalent to maximizing r_{aI}. This is the same procedure employed in obtaining the regression coefficients in multiple linear regression. Thus the b values could be regarded as partial regression coefficients of the individual's breeding value on each measurement. The minimization results in a set of simultaneous equations similar to the normal equations of multiple linear regression, which are solved to obtain the b values. The set of equations to be solved for the b values is:

$$\begin{aligned} b_1p_{11} + b_2p_{12} + \cdots + b_mp_{1m} &= g_{11} \\ b_2p_{21} + b_2p_{22} + \cdots + b_mp_{2m} &= g_{12} \\ \vdots \qquad\qquad & \vdots \\ b_1p_{m1} + b_2p_{m2} + \cdots + b_mp_{mm} &= g_{1m} \end{aligned} \quad [1.16]$$

where p_{mm} and g_{mm} are the phenotypic and genetic variances, respectively, for individual or trait m; and p_{mn} and g_{mn} are the phenotypic and genetic covariances, respectively, between individuals or traits m and n.

In matrix form, equation [1.16] is:

$$\mathbf{Pb} = \mathbf{G}$$

and:

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

where **P** is the variance and covariance matrix for observations and **G** is the covariance matrix between observations and breeding value to be predicted.

Therefore the selection index equation is:

$$I = \hat{a} = \mathbf{P}^{-1}\mathbf{G}(\mathbf{y} - \boldsymbol{\mu}) \quad [1.17]$$

$$= \mathbf{b}'(\mathbf{y} - \boldsymbol{\mu}) \quad [1.18]$$

where **μ** refers to estimates of environmental influences on observations, assumed to be known without error. The application of the selection index to some data therefore involves setting up equation [1.17]. From equation [1.18] it is obvious that the previous methods for prediction of breeding values discussed in Sections 1.2 to 1.5 are not different from a selection index and they could be expressed as in equation [1.17].

1.6.1 Accuracy of index

As before, the accuracy ($r_{a,I}$) of an index is the correlation between the true breeding value and the index. The higher the correlation, the better the index as a predictor of breeding value. It provides a means of evaluating different indices based on different observations, to find out, for instance, whether a particular observation is worth including in an index or not.

From the definition above:

$$r_{a,I} = \text{cov}(a, I)/(\sigma_a \sigma_I)$$

First we need to calculate σ_I^2 and cov(a, I) in the above equation. Using the formula for the variance of predicted breeding value in Section 1.2.1:

$$\begin{aligned}\sigma_I^2 &= \text{var}(b_1 y_1) + \text{var}(b_2 y_2) + \cdots + 2b_1 b_2 \,\text{cov}(y_1, y_2) + \cdots \\ &= b_1^2 \,\text{var}(y_1) + b_2^2 \,\text{var}(y_2) + \cdots + 2b_1 b_2 \,\text{cov}(y_1, y_2) + \cdots \\ \sigma_I^2 &= b_1^2 p_{11} + b_2^2 p_{22} + \cdots + 2b_1 b_2 p_{12} + \cdots\end{aligned}$$

or in general:

$$\sigma_I^2 = \sum_{i=1}^{m} b_1^2 p_{ii} + \left(\sum_{i=1}^{m} \sum_{j=1}^{m} b_i b_j p_{ij};\ i \neq j \right)$$

where m is the number of traits or individuals in the index.

In matrix notation:

$$\sigma_I^2 = \mathbf{b'Pb}$$

Now $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}$; substituting this value for **b**:

$$\sigma_I^2 = \mathbf{G'P}^{-1}\mathbf{G} \qquad [1.19]$$

The covariance between the true breeding value for trait or individual i and index is:

$$\begin{aligned}\text{cov}(a_i, I) &= \text{cov}(a_i, b_1 y_1) + \text{cov}(a_i, b_2 y_2) + \cdots + \text{cov}(a_i, b_j y_j) \\ &= b_1 \text{cov}(a_i, y_1) + b_2 \text{cov}(a_i, y_2) + \cdots + b_j \text{cov}(a_i, y_j)\end{aligned}$$

or in general:

$$\text{cov}(a_i, I) = \sum_{j=1}^{m} b_j g_{ij} \qquad [1.20]$$

where g_{ij} is the genetic covariance between traits or individuals i and j, and m is the number of traits or individuals in the index.

In matrix notation:

$$\text{cov}(a_i, I) = \mathbf{b}'\mathbf{G}$$

Substituting $\mathbf{P}^{-1}\ \mathbf{G}$ for $\mathbf{b}$:

$$\text{cov}(a_i, I) = \mathbf{G}'\mathbf{P}^{-1}\mathbf{G}$$
$$= \sigma_I^2$$

Thus, as previously, the regression of breeding value on predicted breeding values is unity. Therefore:

$$r_{a,I} = \sigma_I^2/(\sigma_a\sigma_I) = \sigma_I/\sigma_a$$

For calculation purposes, r is better expressed as:

$$r_{a,I} = \sqrt{\frac{\sum_{j=1}^{m} b_j g_{ij}}{\sigma_a^2}} \qquad [1.21]$$

Response to selection on the basis of an index is:

$$R = ir_{a,I}\sigma_a$$
$$= i\sigma_I$$

1.6.2 Examples of selection indices using different sources of information

Data available on correlated traits

Example 1.6

Assume the following parameters were obtained for average daily gain (ADG) from birth to 400 days and lean per cent (LP) at the same age in a group of beef calves:

	Heritability	Standard deviation
ADG (g/day)	0.43	80.0
LP (%)	0.30	7.2

If the genetic and phenotypic correlations (r_g and r_p) between ADG and LP are 0.30 and –0.10, respectively, construct an index to improve growth rate in the beef calves. Assuming ADG as trait 1 and LP as trait 2, then, from the given parameters:

$$p_{11} = 80^2 = 6400$$
$$p_{22} = 7.2^2 = 51.84$$
$$p_{12} = r_p\sqrt{(p_{11})(p_{22})} = -0.1\sqrt{(6400)(51.84)} = -57.6$$
$$p_{12} = p_{21} = \text{cov}(y_1, y_2) = 1/2\sigma^2{}_a = 1/2(2752) = 1376$$
$$g_{11} = h^2(p_{11}) = 0.43(6400) = 2752$$

$$g_{22} = h^2(p_{22}) = 0.30(51.84) = 15.552$$

$$g_{12} = g_{21} = r_g\sqrt{(g_{11})(g_{22})} = 62.064$$

The index equations to be solved are:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}^{-1} \begin{bmatrix} g_{11} \\ g_{21} \end{bmatrix}$$

Inserting appropriate values gives:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 6400.00 & -57.60 \\ -57.60 & 51.84 \end{bmatrix}^{-1} \begin{bmatrix} 2752.000 \\ 62.064 \end{bmatrix}$$

The solutions are $b_1 = 0.445$ and $b_2 = 1.692$.

The index therefore is:

$$I = 0.445(\text{ADG} - \mu_{\text{ADG}}) + 1.692(\text{LP} - \mu_{\text{LP}})$$

where μ_{ADG} and μ_{LP} are herd averages for ADG and LP. Using equation [1.21]:

$$r = \sqrt{[(0.445(2752) + 1.692(62.064))/2752]} = 0.695$$

Using single records on individual and relatives

Example 1.7

Suppose the average daily gain (ADG) for a bull calf (y_1) is 900 g/day and the ADG for his sire (y_2) and dam (y_3) are 800 g/day and 450 g/day, respectively. Assuming all observations were obtained in the same herd and using the same parameters as in Example 1.6, predict the breeding value of the bull calf for ADG and its accuracy.

From the parameters given:

$$p_{11} = p_{22} = p_{33} = \sigma_y^2 = 6400$$

$$p_{12} = p_{21} = \text{cov}(y_1, y_2) = \tfrac{1}{2}\sigma_a^2 = \tfrac{1}{2}(2752) = 1376$$

$$p_{13} = p_{12} = 1376$$

$$p_{23} = 0$$

$$g_{11} = \sigma_a^2 = 2752$$

$$g_{12} = g_{13} = \tfrac{1}{2}\sigma_a^2 = 1376$$

The index equations are:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 6400 & 1376 & 1376 \\ 1376 & 6400 & 0000 \\ 1376 & 0000 & 6400 \end{bmatrix}^{-1} = \begin{bmatrix} 2752 \\ 1376 \\ 1376 \end{bmatrix}$$

Solutions to the above equations are $b_1 = 0.372$, $b_2 = 0.135$ and $b_3 = 0.135$.

The index is:

$$I = 0.372(900 - \mu) + 0.135(800 - \mu) + 0.135(450 - \mu)$$

where μ is the herd average. The accuracy is:

$$r = \sqrt{[(0.372(2752) + 0.135(176) + 0.135(176))/2752]} = 0.712$$

The high accuracy is due to the inclusion of information from both parents.

Using means of records from animal and relatives

Example 1.8

It is given that average protein yield for the first two lactations for a cow ($\tilde{y}_1$) called Zena is 230 kg and the mean protein yield of five other cows ($\tilde{y}_2$), each with two lactations, is 300 kg, if all cows are all daughters of the same bull and no other relationship exists among them, predict the breeding value of Zena, assuming a heritability of 0.25, a repeatability (t) of 0.5, standard deviation of 34 kg and herd average of 250 kg for protein yield in the first two lactations.

From the given parameters:

$$g_{11} = \sigma_a^2 = h^2\sigma_{\tilde{y}}^2 = 0.25(34^2) = 289$$

and:

$$g_{12} = \text{covariance between half-sibs} = \tfrac{1}{4}(\sigma_a^2) = \tfrac{1}{4}(289) = 72.25$$

From calculations in Section 1.2.2:

$$p_{11} = \text{var}(\tilde{y}_1) = \left(t + \frac{(1-t)}{n}\right)\sigma_{\tilde{y}}^2$$
$$= (0.5 + (1 - 0.5)/2)34^2 = 867$$

Using similar arguments:

$$p_{22} = \text{var}(\tilde{y}_2) = \sigma_B^2 + 1/n(\sigma_W^2)$$

where σ_B^2 is the between-cow variance and $1/n(\sigma_W^2)$ is the mean of the within-cow variance. From Section 1.3, since cows are half-sibs:

$$\sigma_B^2 = \tfrac{1}{4}\sigma_a^2$$

and for cow i in the group of five cows:

$$\sigma_W^2 = \text{var}(\tilde{y}_{2i}) - \sigma_B^2$$

where $\tilde{y}_{2i}$ is the mean of the first two lactations for cow i. Since all five cows each have two records like Zena, var $(\tilde{y}_{2i}) = P_{11}$, therefore:

$$\sigma_W^2 = (p_{11} - \tfrac{1}{4}\sigma_a^2)$$

and:

$$1/n(\sigma_W^2) = 1/n(p_{11} - \tfrac{1}{4}\sigma_a^2)$$

Therefore:

$$p_{22} = \tfrac{1}{4}\sigma_a^2 + 1/n(p_{11} - \tfrac{1}{4}\sigma_a^2)$$
$$= \tfrac{1}{4}(289) + (\tfrac{1}{5})(867 - \tfrac{1}{4}(289)) = 231.2$$

The index equations are:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 867 & 72.25 \\ 72.25 & 231.2 \end{bmatrix}^{-1} = \begin{bmatrix} 289 \\ 72.25 \end{bmatrix}$$

The solutions are $b_1 = 0.316$ and $b_2 = 0.213$ and the index is:

$$I = 0.316(230 - 250) + 0.213(300 - 250)$$

The accuracy of the index is:

$$r = \sqrt{[(0.316(289) + 0.213(72.5))/289]} = 0.608$$

1.6.3 Prediction of aggregate genotype

At times, the aim is to predict not just the breeding value of a single trait but that of a composite of several traits evaluated in economic terms. The aggregate breeding value (H) or merit for such several traits can be represented as:

$$H = w_1a_1 + w_2a_2 + \cdots + w_ma_m$$

where a_i is the breeding value of the ith trait and w_i the weighting factor which expresses the relative economic importance associated with the ith trait. The construction of an index to predict or improve H is based on the same principles as those discussed earlier except that it includes the relative economic weight for each trait.

Thus

$$I = \mathbf{P}^{-1}\mathbf{Gw}(\mathbf{y} - \boldsymbol{\mu}) \qquad [1.22]$$

where $\mathbf{w}$ is the vector of economic weights and all other terms are as defined in [1.17]. The equations to be solved to get the weights (b values) to be used in the index are:

$$\begin{aligned} b_1p_{11} + b_2p_{12} + \cdots + b_mp_{1m} &= w_1g_{11} + w_2g_{12} + \cdots + w_mg_{1m} \\ b_2p_{21} + b_2p_{22} + \cdots + b_mp_{2m} &= w_1g_{12} + w_2g_{12} + \cdots + w_mg_{2m} \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots & \\ b_1p_{m1} + b_2p_{m2} + \cdots + b_mp_{mm} &= w_1g_{m1} + w_2g_{m2} + \cdots + w_mg_{mm} \end{aligned}$$

In matrix notation these equations are:

$$\mathbf{Pb} = \mathbf{Gw}$$

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Gw}$$

It should be noted that it is possible that there are some traits in the index which are not in the aggregate breeding value but may be correlated with other traits in *H*. Conversely, some traits in the aggregate breeding value may be difficult to measure or occur late in life and therefore may not be in the index. Such traits may be replaced in the index with other highly correlated traits which are easily measurable or occur early in life. Consequently the vector of economic weights may not necessarily be of the same dimension as traits in the index, as indicated in the equations for *b* above. Each trait in the index is weighted by the economic weight relevant to the breeding value of the trait it is predicting in the aggregate breeding value.

The index calculated using [1.22] implies that the same economic weights are applied to the traits in the aggregate genotype across the whole population. A change in the economic weight for one of the traits would imply recalculating the index. An alternative formulation of [1.22] involves calculating a sub-index for each trait in *H* without the economic weights. The final index [1.23] is obtained by summing the sub-indices for each trait weighted by their respective economic weights. Thus:

$$I = \sum_{i=1}^{m} I_i w_i \qquad [1.23]$$

where $I_i = \mathbf{P}^{-1}\mathbf{G}_i(\mathbf{y} - \boldsymbol{\mu})$, the sub-index for trait *i* in *H* and w_i = economic weight for trait *i*.

With [1.23], a change in the economic weights of any of the traits in the index can easily be implemented without recalculating the index.

To demonstrate that [1.22] and [1.23] are equivalent, assume that there are two traits in *H*; then [1.23] becomes:

$$\begin{aligned} I &= I_1 w_1 + I_2 w_2 \\ &= \mathbf{P}^{-1}\mathbf{G}_1 w_1(\mathbf{y} - \boldsymbol{\mu}) + \mathbf{P}^{-1}\mathbf{G}_2 w_2(\mathbf{y} - \boldsymbol{\mu}) \end{aligned}$$

where $\mathbf{G}_i$ is the covariance matrix between trait *i* and all traits in the index. Thus:

$$\begin{aligned} I &= \mathbf{P}^{-1}(\mathbf{G}_1 w_1 + \mathbf{G}_2 w_2)(\mathbf{y} - \boldsymbol{\mu}) \\ &= \mathbf{P}^{-1}\mathbf{G}\mathbf{w}(\mathbf{y} - \boldsymbol{\mu}) \end{aligned}$$

which is the same as [1.22].

Example 1.9

Assume the economic weights for ADG and LP are £1.5 and £0.5 per each increase of 1 kg in ADG and 1% increase in LP, respectively. Using the genetic parameters in Example 1.6, construct an index to select fast-growing lean beef calves using equation [1.22]. Repeat the analysis using equation [1.23].

Using [1.22], index equations are:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}^{-1} \begin{bmatrix} w_1 g_{11} + w_2 g_{12} \\ w_1 g_{21} + w_2 g_{22} \end{bmatrix}$$

Inserting the appropriate values:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 6400.00 & -57.60 \\ -57.60 & 51.84 \end{bmatrix}^{-1} \begin{bmatrix} 1.5(2752) + 0.5(62.064) \\ 1.5(62.064) + 0.5(15.552) \end{bmatrix}$$

Solutions for b_1 and b_2 from the above equations are 0.674 and 2.695, respectively. The index therefore is:

$$I = 0.674(\text{ADG} - \mu_{\text{ADG}}) + 2.694(\text{LP} - \mu_{\text{LP}})$$

Applying equation [1.23], the sub-index for ADG is the same as that calculated in Example 1.6, with $b_1 = 0.445$ and $b_2 = 1.692$. The sub-index for LP is:

$$b_1 p_{11} + b_2 p_{12} = g_{12}$$
$$b_1 p_{21} + b_2 p_{22} = g_{22}$$

which gives:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 6400.00 & -57.60 \\ -57.60 & 51.84 \end{bmatrix}^{-1} \begin{bmatrix} 62.064 \\ 15.552 \end{bmatrix}$$

The solutions are $b_1 = 0.0125$ and $b_2 = 0.314$. Multiplying the sub-indices by their respective weights gives:

$$\begin{aligned} I_{\text{ADG}} &= 0.445(1.5)(\text{ADG} - \mu_{\text{ADG}}) + 1.692(1.5)(\text{LP} - \mu_{\text{LP}}) \\ &= 0.668(\text{ADG} - \mu_{\text{ADG}}) + 2.538(\text{LP} - \mu_{\text{LP}}) \end{aligned}$$

and:

$$\begin{aligned} I_{\text{LP}} &= 0.0125(0.5)(\text{ADG} - \mu_{\text{ADG}}) + 0.314(0.5)(\text{LP} - \mu_{\text{LP}}) \\ &= 0.006(\text{ADG} - \mu_{\text{ADG}}) + 0.157(\text{LP} - \mu_{\text{LP}}) \end{aligned}$$

Summing the b terms from the two sub-indices, the final b terms are:

$$b_1 = 0.668 + 0.006 = 0.674$$
$$b_2 = 2.538 + 0.157 = 2.695$$

Therefore the final index is:

$$I = 0.675(\text{ADG} - \mu_{\text{ADG}}) + 2.695(\text{LP} - \mu_{\text{LP}})$$

which is the same as that calculated using equation [1.22].

1.6.4 Overall economic indices using predicted genetic merit

Overall economic indices that combine PTAs or estimated breeding values (BVs) calculated by best linear unbiased prediction (BLUP, see Chapter 3) have become very popular in the last decade. In addition to the recognition that more than one trait contributes to profitability, the broadening of selection goals has also been due to the need to incorporate health and welfare traits to accommodate public concerns. Examples of indices constructed with PTAs or BVs of several traits and used in genetic improvement of dairy cattle include production index (PIN), combining PTAs for

milk, fat and protein, and profitable life index (PLI), which is PIN plus PTAs for longevity and somatic cell count, in the UK; and, in The Netherlands, net profit index for milk (INET), which combines BVs for milk, fat and protein, and durable performance sum (DPS), which is INET plus durability (Interbull, 2000). The principles for calculating these indices are similar to those outlined in previous sections. Given that the PTAs or BVs are from a complete multivariate analysis, the optimal index weights (**b**) are the sum of the partial regression coefficients of each goal trait on each index trait, weighted by the economic value of the goal trait (Veerkamp *et al.*, 1995). Thus, given *m* traits in the selection goal and *n* traits in the index, the partial regressions can be calculated as:

$$\mathbf{R} = \mathbf{G}^{-1}\mathbf{G}_{ig}$$

and:

$$\mathbf{b} = \mathbf{R}\mathbf{w}$$

where **R** is a matrix of partial genetic regression, $\mathbf{G}_{ig}$ is the matrix of genetic covariance between *m* goal and *n* index traits, **G** is the genetic covariance matrix between the index traits, and **w** is the vector of economic weights. It is obvious that when goal and index traits are the same, $\mathbf{G}_{ig} = \mathbf{G}$ and $\mathbf{b} = \mathbf{w}$. In the case where the index and goal traits are not the same, **R** can be estimated directly from a regression of phenotype on the estimated breeding values for the index traits (Brotherstone and Hill, 1991). However, if PTAs or BVs are from a univariate analysis, rather than from a multivariate analysis, the use of **b** above results only in minimal loss of efficiency in the index (Veerkamp *et al.*, 1995).

Selection based on breeding values from BLUP is usually associated with an increased rate of inbreeding as it favours the selection of closely related individuals. Quadratic indices can be used to optimize the rate of genetic gain and inbreeding. This does not fall within the main subject area of this text and interested readers should see the work by Meuwissen (1997) and Grundy *et al.* (1998).

1.6.5 Restricted selection index

Restricted selection index is used when the aim is to maximize selection for a given aggregate genotype, subject to the restriction that no genetic change is desired in one or more of the traits in the index for *H*. This is achieved by the usual index procedure and setting the covariance between the index and the breeding value ($\mathrm{cov}(I, a_i)$) for the *i*th trait specified not to change to zero. It was Kempthorne and Nordskog (1959) who introduced the idea of imposing restrictions on the general index procedure.

For instance, consider the aggregate genotype composed of two traits:

$$H = w_1 a_1 + w_2 a_2$$

However, it is desired that there should be no genetic change in trait 2; thus effectively:

$$H = w_1 a_1$$

and the index to predict H is:

$$I = b_1 y_1 + b_2 y_2$$

To ensure that there is no genetic change in trait 2, $\text{cov}(I, a_2)$ must be equal to zero. From equation [1.20]:

$$\text{cov}(I, a_2) = b_1 g_{12} + b_2 g_{22} = 0$$

This is included as an extra equation to the normal equations for the b values, and a dummy unknown, the so-called Lagrange multiplier, is added to the vector of solutions for the index weights (Ronningen and Van Vleck, 1985). The equations for the index therefore are:

$$\begin{bmatrix} b_1 \\ b_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & g_{12} \\ p_{21} & p_{22} & g_{22} \\ g_{12} & g_{22} & 0 \end{bmatrix}^{-1} \begin{bmatrix} g_{11} \\ g_{12} \\ 0 \end{bmatrix} \qquad [1.24]$$

Example 1.10

Using the same data and parameters as in Example 1.6, construct an index to improve the aggregate genotype for fast-growing lean cattle using an index consisting of growth rate and lean per cent but with no genetic change in lean per cent.

From [1.24], the index equations are:

$$\begin{bmatrix} 6400.000 & -57.600 & 62.064 \\ 57.600 & 51.800 & 15.552 \\ 62.064 & 15.552 & 0.000 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 2752.000 \\ 62.064 \\ 0 \end{bmatrix}$$

The solutions for b_1 and b_2 from solving the above equations are 0.325 and −1.303. Therefore the index is:

$$I = 0.325(\text{ADG} - \mu_{\text{ADG}}) + (-1.303(\text{LP} - \mu_{\text{LP}}))$$

The accuracy of this index (equation [1.21]) is

$$r = \sqrt{(0.325(2752) + (-1.303(62.064)))/2752} = 0.544$$

which is lower than the accuracy for the equivalent index in Example 1.6, but with no restriction on LP, and is also lower than the accuracy of prediction of breeding value for ADG on the basis of single records. The imposition of a restriction on any trait in the index will never increase the efficiency of the index but usually reduces it unless $I_i = 0$ for the constrained trait.

1.6.6 Index combining breeding values from phenotype and genetic marker information

Consider a situation in which one or more genes affecting a trait with a large impact on profit have been identified to be linked to a genetic marker (see Chapter 8). If genetic prediction based only on marker information is available in addition to the conventional BV estimated without marker information, then both sources of information can be combined into an index (Goddard, 1999). It is also possible that the conventional BV is based on a subset of traits in the breeding goal and marker information is available on other traits that are not routinely measured, such as meat quality traits.

A selection index could be used to combine both sources of information and the increase in accuracy from including marker information could be computed (Goddard, 1999). Given r as the accuracy of the conventional breeding BV and d as the proportion of genetic variance explained by the marker information, then the covariance between the two sources of information is dr^2. If m is the BV based on marker information and a the BV from phenotypic information, then:

$$\text{var}\begin{pmatrix} m \\ a \end{pmatrix} = \begin{pmatrix} d & dr^2 \\ dr^2 & r^2 \end{pmatrix}$$

Let g be the true breeding value to be predicted, then cov(g, m) = d and cov(g, a) = r^2. The normal index equations are:

$$\begin{pmatrix} b_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} d & dr^2 \\ dr^2 & r^2 \end{pmatrix}^{-1} \begin{pmatrix} d \\ r^2 \end{pmatrix}$$

Solving the above equations gives the following index weights:

$$b_1 = 1 - r^2/(1 - dr^2) \quad \text{and} \quad b_2 = 1 - d/(1 - dr^2)$$

The variance of the index = reliability r_I^2 is:

$$r_I^2 = [(1 - r^2)d + (1 - d)r^2]/(1 - dr^2)$$

The increase in reliability (r_{inc}^2) from incorporating marker information therefore is:

$$r_{inc}^2 = (r_I^2 - r^2) = d/(1 - dr^2)[(1 - r^2)^2]$$

For example, given that r^2 of the conventional BV is 0.34 and marker information accounts for 25% of the genetic variance, then r_I^2 is 0.49, an increase in reliability of 0.15. However, if r^2 is 0.81, then r_I^2 is 0.83 and r_{inc}^2 is only 0.02. Thus the usefulness of marker information is greater when reliability is low, such as in traits of low heritability and also traits which cannot be measured in young animals, such as carcass traits (Goddard, 1999).

2 Genetic Covariance Between Relatives

Of fundamental importance in the prediction of breeding values is the genetic relationship among individuals. From Chapter 1, it is noticed that the use of the selection index to predict breeding values requires genetic covariance between individuals to construct the genetic covariance matrix. Genetic evaluation using best linear unbiased prediction (BLUP), the subject of the next chapter, is heavily dependent on the genetic covariance among individuals, both for higher accuracy and for unbiased results. The genetic covariance among individuals is comprised of three components: the additive genetic variance, the dominance variance and the epistatic variance. This chapter addresses the calculation of the additive genetic relationship among individuals and how to determine the level of inbreeding. Dominance and epistasis genetic relationships are considered in Chapter 9, which deals with non-additive models.

2.1 The Numerator Relationship Matrix

The probability of identical genes by descent occurring in two individuals is termed the coancestry or the coefficient of kinship (Falconer and MacKay, 1996) and the additive genetic relationship between two individuals is twice their coancestry. The matrix which indicates the additive genetic relationship among individuals is called the numerator relationship matrix (**A**). It is symmetric and its diagonal element for animal i (a_{ii}) is equal to $1 + F_i$, where F_i is the inbreeding coefficient of animal i (Wright, 1922). The diagonal element represents twice the probability that two gametes taken at random from animal i will carry identical alleles by descent. The off-diagonal element, a_{ij}, equals the numerator of the coefficient of relationship (Wright, 1922) between animals i and j. When multiplied by the additive genetic variance (σ_u^2), $\mathbf{A}\sigma_u^2$ is the covariance among breeding values. Thus if u_i is the breeding value for animal i,

$\text{var}(u_i) = a_{ii}\sigma_u^2 = (1 + F_i)\sigma_u^2$. The matrix **A** can be computed using path coefficients but a recursive method which is suitable for computerization was described by Henderson (1976). Initially, animals in the pedigree are coded 1 to n and ordered such that parents precede their progeny. The following rules are then employed to compute **A**.

If both parents (s and d) of animal i are known:

$a_{ji} = a_{ij} = 0.5(a_{js} + a_{jd})$; $j = 1$ to $(i - 1)$
$a_{ii} = 1 + 0.5(a_{sd})$

If only one parent s is known and assumed unrelated to the mate:

$a_{ji} = a_{ij} = 0.5(a_{js})$; $j = 1$ to $(i - 1)$
$a_{ii} = 1$

If both parents are unknown and are assumed unrelated:

$a_{ji} = a_{ij} = 0$; $j = 1$ to $(i - 1)$
$a_{ii} = 1$

For example, assume that the data in Table 2.1 are the pedigree for six animals. The numerator relationship matrix for the example pedigree is:

	1	2	3	4	5	6
1	1.00	0.0	0.50	0.50	0.50	0.25
2	0.00	1.0	0.50	0.00	0.25	0.625
3	0.50	0.50	1.00	0.25	0.625	0.563
4	0.50	0.00	0.25	1.00	0.625	0.313
5	0.50	0.25	0.625	0.625	1.125	0.688
6	0.25	0.625	0.563	0.313	0.688	1.125

For instance:

$a_{11} = 1 + 0 = 1$
$a_{12} = 0.5(0 + 0) = 0 = a_{21}$
$a_{22} = 1 + 0 = 1$
$a_{13} = 0.5(a_{11} + a_{12}) = 0.5(1.0 + 0) = 0.5 = a_{31}$
$a_{23} = 0.5(a_{12} + a_{22}) = 0.5(0 + 1.0) = 0.5 = a_{32}$
$\vdots$
$a_{34} = 0.5(a_{13}) = 0.5(0.5 + 0) = 0.25 = a_{43}$
$\vdots$
$a_{66} = 1 + 0.5(a_{52}) = 1 + 0.5(0.25) = 1.125$

From the above calculation, the inbreeding coefficient for calf 6 is 0.125.

Table 2.1. Pedigree for six animals.

Calf	Sire	Dam
3	1	2
4	1	Unknown
5	4	3
6	5	2

2.2 Decomposing the Relationship Matrix

The relationship matrix can be expressed (Thompson, 1977a) as:

$$\mathbf{A} = \mathbf{TDT}' \qquad [2.1]$$

where **T** is a lower triangular matrix and **D** is a diagonal matrix. This relationship has been used to develop rules for obtaining the inverse of **A**. A non-zero element of the matrix **T**, say t_{ij}, is the coefficient of relationship between animals i and j, if i and j are direct relatives or $i = j$ and it is assumed that there is no inbreeding. Thus the matrix **T** traces the flow of genes from one generation to the other; in other words, it accounts only for direct (parent–offspring) relationships. It can easily be computed applying the following rules.

For the ith animal:

$t_{ii} = 1$

If both parents (s and d) are known:

$t_{ij} = 0.5(t_{sj} + t_{dj})$

If only one parent (s) is known:

$t_{ij} = 0.5(t_{sj})$

If neither parent is known:

$t_{ij} = 0$

The variance of the diagonal matrix **D** is the variance and covariance matrix for Mendelian sampling. The Mendelian sampling (m) for an animal i with breeding value u_i and u_s and u_d as breeding values for its sire and dam, respectively, is:

$$m_i = u_i - 0.5(u_s + u_d) \qquad [2.2]$$

D has a simple structure and can easily be calculated. From [2.2], if both parents of animal i are known, then:

$$\begin{aligned}\text{var}(m_i) &= \text{var}(u_i) - \text{var}(0.5u_s + 0.5u_d) \\ &= \text{var}(u_i) - \text{var}(0.5u_s) - \text{var}(0.5u_d) + 2\text{cov}(0.5u_s, 0.5u_d) \\ &= (1 + F_i)\sigma_u^2 - 0.25a_{ss}\,\sigma_u^2 - 0.25a_{dd}\,\sigma_u^2 - 0.5a_{sd}\,\sigma_u^2\end{aligned}$$

where a_{ss}, a_{dd} and a_{sd} are elements of the relationship matrix **A** and F_i is the inbreeding coefficient of animal i.

$$\text{var}(m_i)/\sigma_u^2 = d_{ii} = (1 + F_i) - 0.25a_{ss} - 0.25a_{dd} - 0.5a_{sd}$$

Since $F_i = 0.5a_{sd}$:

$$\begin{aligned} d_{ii} &= 1 - 0.25(1 + F_s) - 0.25(1 + F_d) \\ &= 0.5 - 0.25(F_s + F_d) \end{aligned}$$

where F_s and F_d are the inbreeding coefficients of animal i's sire and dam, respectively. If only one parent (s) is known, the diagonal element is:

$$\begin{aligned} d_{ii} &= 1 - 0.25(1 + F_s) \\ &= 0.75 - 0.25(F_s) \end{aligned}$$

and if no parent is known:

$$d_{ii} = 1$$

For the pedigree in Table 2.1, the matrix **T** is:

	1	2	3	4	5	6
1	1.0	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0	0.0
3	0.5	0.5	1.0	0.0	0.0	0.0
4	0.5	0.0	0.0	1.0	0.0	0.0
5	0.5	0.25	0.5	0.5	1.0	0.0
6	0.25	0.625	0.25	0.25	0.5	1.0

and **D** is:

$$\mathbf{D} = \text{diag}(1.0, 1.0, 0.5, 0.75, 0.5, 0.469)$$

For instance, animal 3 has only the sire known, which is not inbred; therefore:

$$d_{33} = 0.75 - 0 = 0.75$$

and:

$$d_{66} = 0.5 - 0.25(0.125 + 0) = 0.469$$

because both parents are known and the sire has an inbreeding coefficient of 0.125.

2.3 Computing the Inverse of the Relationship Matrix

The prediction of breeding value requires the inverse of the relationship matrix, $\mathbf{A}^{-1}$. This could be obtained by setting up **A** by the recursive method and inverting it. This, however, is not computationally feasible when evaluating a large number of animals. In 1976, Henderson presented

a simple procedure for calculating $\mathbf{A}^{-1}$ without setting up **A**. The procedure and its principles are described below.

From equation [2.1], the inverse of **A** can be written as:

$$\mathbf{A}^{-1} = (\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1} \qquad [2.3]$$

The matrix $\mathbf{D}^{-1}$ is easy to obtain because **D** is a diagonal matrix. The diagonal elements of $\mathbf{D}^{-1}$ are simply the reciprocals of the diagonal elements of **D** computed in Section 2.2. $\mathbf{T}^{-1}$ is a lower triangular matrix with ones in the diagonal, and the only non-zero elements to the left of the diagonal in the row for the animal *i* are –0.5 for columns corresponding to the known parents. It can be derived as $\mathbf{I} - \mathbf{M}$, where **I** is an identity matrix of the order of animals on the pedigree and **M** is a matrix of the contribution of gametes from parents to progeny (Kennedy, 1989). Since progeny *i* receives half of its genes from each parent, the only non-zero elements in row *i* of **M** are 0.5 corresponding to columns of known parents. Thus, if both parents of progeny *i* are unknown, all elements of row *i* are zero. For the pedigree in Table 2.1, $\mathbf{T}^{-1}$ can be calculated as:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{(\mathbf{I})} - \underbrace{\begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.0 & 0.5 & 0.0 \end{bmatrix}}_{(\mathbf{M})}$$

$$= \underbrace{\begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & -0.5 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.5 & -0.5 & 1.0 & 0.0 \\ 0.0 & -0.5 & 0.0 & 0.0 & -0.5 & 1.0 \end{bmatrix}}_{(\mathbf{T}^{-1})}$$

and:

$$\mathbf{D}^{-1} = \text{diag}(1, 1, 2, 1.333, 2, 2.133)$$

2.3.1 Inverse of the numerator relationship matrix ignoring inbreeding

The relation shown in equation [2.3] was used by Henderson (1976) to derive simple rules for obtaining $\mathbf{A}^{-1}$ without accounting for inbreeding. With inbreeding ignored, the diagonal elements of $\mathbf{D}^{-1}$ are either 2 or $\frac{4}{3}$ or 1 if both or one or no parents are known, respectively. Let α_i represent the diagonal element of $\mathbf{D}^{-1}$ for animal *i*. Initially set $\mathbf{A}^{-1}$ to zero and apply the following rules.

If both parents of the ith animal are known, add:

α_i to the (i,i) element
$-\alpha_i/2$ to the (s,i), (i,s), (d,i) and (i,d) elements
$\alpha_i/4$ to the (s,s), (s,d), (d,s) and (d,d) elements

If only one parent (s) of the ith animal is known, add:

α_i to the (i,i) element
$-\alpha_i/2$ to the (s,i) and (i,s) elements
$\alpha_i/4$ to the (s,s) element

If neither parent of the ith animal is known, add:

α_i to the (i,i) element

As an illustration, the inverse of the relationship matrix in Section 2.1 can be calculated as below. Initially, list all animals in the pedigree:

Calf	Sire	Dam
1	Unknown	Unknown
2	Unknown	Unknown
3	1	2
4	1	Unknown
5	4	3
6	5	2

Then set up a 6×6 table for the animals. For animals 1 and 2, both parents are unknown; therefore $\alpha_1 = \alpha_2 = 1$. Add 1 to their diagonal elements (1,1 and 2,2). For animal 3, both parents are known; therefore $\alpha_3 = 2$. Add 2 to the 3,3 element, –1 to the (3,1), (1,3), (3,2) and (2,3) elements and $\frac{1}{2}$ to the (1,1), (1,2), (2,1) and (2,2) elements. For animal 4, only one parent is known; therefore $\alpha_4 = \frac{4}{3}$. Add $\frac{4}{3}$ to the (4,4) element, $-\frac{2}{3}$ to the (4,1) and (1,4) elements and $\frac{1}{3}$ to the (1,1) element. After the first four animals, the table is:

	1	2	3	4	5	6
1	$1+\frac{1}{2}+\frac{1}{3}$	$\frac{1}{2}$	–1	$-\frac{2}{3}$		
2	$\frac{1}{2}$	$1+\frac{1}{2}$	–1			
3	–1	–1	2			
4	$-\frac{2}{3}$			$\frac{4}{3}$		
5						
6						

After applying the relevant rules to animals 5 and 6, the inverse of **A** then is:

	1	2	3	4	5	6
1	1.83	0.5	–1.0	–0.667	0.0	0.0
2	0.5	2.0	–1.0	0.0	0.5	–1.0
3	–1.0	–1.0	2.50	0.5	–1.0	0.0
4	–0.67	0.0	0.5	1.833	–1.0	0.0
5	0.0	0.5	–1.0	–1.0	2.50	–1.0
6	0.0	–1.0	0.0	0.0	–1.0	2.0

Using equation [2.3], the inverse of **A** can be calculated directly. If inbreeding is ignored, **D** for the pedigree is:

$\mathbf{D}$ = diag(1.0, 1.0, 0.5, 0.75, 0.5, 0.5)

and:

$\mathbf{D}^{-1}$ = diag(1, 1, 2, 1.33, 2, 2)

Therefore the inverse of the relationship matrix using [2.3] is:

$$\underset{(\mathbf{T}^{-1})'}{\begin{bmatrix} 1.0 & 0.0 & -0.5 & -0.5 & 0.0 & 0.0 \\ 0.0 & 1.0 & -0.5 & 0.0 & 0.0 & -0.5 \\ 0.0 & 0.0 & 1.0 & 0.0 & -0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & -0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & -0.5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}} \underset{\mathbf{D}^{-1}}{\begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 2.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.33 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 2.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 2.00 \end{bmatrix}}$$

$$\underset{(\mathbf{T}^{-1})}{\begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ -0.5 & -0.5 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.5 & -0.5 & 1.0 & 0.0 \\ 0.0 & -0.5 & 0.0 & 0.0 & -0.5 & 1.0 \end{bmatrix}}$$

$$= \underset{\mathbf{A}^{-1}}{\begin{bmatrix} 1.83 & 0.50 & -1.00 & -0.67 & 0.00 & 0.00 \\ 0.50 & 2.00 & -1.00 & 0.00 & 0.50 & -1.00 \\ -1.00 & -1.00 & 2.50 & 0.50 & -1.00 & 0.00 \\ -0.67 & 0.00 & 0.50 & 1.83 & -1.00 & 0.00 \\ 0.00 & 0.50 & -1.00 & -1.00 & 2.50 & -1.00 \\ 0.00 & -1.00 & 0.00 & 0.00 & -1.00 & 2.00 \end{bmatrix}}$$

which is the same inverse as that obtained previously by applying the rules.

2.3.2 Inverse of the numerator relationship matrix accounting for inbreeding

The calculation of $\mathbf{A}^{-1}$ with inbreeding accounted for involves the application of the same rules outlined in Section 2.3.1 but **D** and therefore $\mathbf{D}^{-1}$ in equation [2.3] are calculated using the inbreeding coefficients of sires and dams (see Section 2.2). This implies that the diagonal elements of the relationship matrix are needed for $\mathbf{A}^{-1}$ to be properly calculated. This could be achieved by initially calculating the **A** for the group of animals and writing the diagonal elements to a file. The diagonal elements could be read from the file while computing $\mathbf{A}^{-1}$. For a large pedigree, this approach would require a large amount of memory for storage and be computationally demanding. However, Quaas (1976) presented a strategy for obtaining the diagonal elements of **A** while computing $\mathbf{A}^{-1}$ without setting up the relationship matrix.

Recall from Section 2.2 that **A** can be expressed as:

$$\mathbf{A} = \mathbf{TDT}'$$

If $\mathbf{L} = \mathbf{T}\sqrt{\mathbf{D}}$:

$$\mathbf{A} = \mathbf{LL}' \qquad [2.4]$$

where **L** is a lower triangular matrix and, since **D** is diagonal, $\sqrt{\mathbf{D}}$ refers to a matrix obtained by calculating the square root of the diagonal elements of **D**. Equation [2.4] implies that the diagonal element of **A** for animal *i* is:

$$a_{ii} = \sum_{m=1}^{i} l_{im}^2 \qquad [2.5]$$

Thus for a pedigree consisting of *m* animals:

$$\begin{aligned}
a_{11} &= l_{11}^2 \\
a_{22} &= l_{21}^2 + l_{22}^2 \\
a_{33} &= l_{31}^2 + l_{32}^2 + l_{33}^2 \\
&\vdots \\
a_{mm} &= l_{m1}^2 + l_{m2}^2 + l_{m3}^2 + \cdots + l_{mm}^2
\end{aligned}$$

From the above all the diagonal elements of **A** can be computed by calculating **L** one column at a time (Quaas, 1984). Only two vectors of dimension equal to the number of animals for storage will be required, one to store the column of **L** being computed and the second to accumulate the sum of squares of the elements of **L** for each animal. The matrices **L** and $\mathbf{A}^{-1}$ can be computed using the following procedure.

From equation [2.4] the diagonal element of **L** for animal i is:

$$l_{ii} = \sqrt{d_i}$$
$$l_{ii} = \sqrt{[0.5 - 0.25(F_s + F_d)]}$$
$$l_{ii} = \sqrt{[1.0 - 0.25(a_{ss} + a_{dd})]}; \quad \text{with } a_{ss} = 1 + F_{ss} \text{ and } a_{dd} = 1 + F_{dd}$$

Using equation [2.5]:

$$l_{ii} = \sqrt{\left[1.0 - 0.25\left(\sum_{m=1}^{s} l_{sm}^2 + \sum_{m=1}^{d} l_{dm}^2\right)\right]}$$

To set up $\mathbf{A}^{-1}$ at the same time, calculate the diagonal element of $\mathbf{D}^{-1}(\alpha_i)$ for animal i as $\alpha_i = 1/l_{ii}^2$. Then compute the contribution of animal i to $\mathbf{A}^{-1}$, applying the usual rules for computing $\mathbf{A}^{-1}$ (see Section 2.3.1).

The off-diagonal elements of **L** to the left of the diagonal for animal i are calculated as:

$$l_{ij} = 0.5(l_{sj} + l_{dj}); \quad s \text{ and } d \text{ equal to or greater than } j$$

For the example pedigree used in Section 2.3.1, the **L** matrix is:

	1	2	3	4	5	6
1	1.0	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0	0.0
3	0.5	0.5	0.707	0.0	0.0	0.0
4	0.5	0.0	0.0	0.866	0.0	0.0
5	0.5	0.25	0.354	0.433	0.707	0.0
6	0.25	0.625	0.177	0.217	0.354	0.685

and $\mathbf{A}^{-1}$ with inbreeding accounted for is:

	1	2	3	4	5	6
1	1.833	0.5	–1.0	–0.667	0.0	0.0
2	0.5	2.033	–1.0	0.0	0.533	–1.067
3	–1.0	–1.0	2.50	0.5	–1.0	0.0
4	–0.667	0.0	0.5	1.833	–1.0	0.0
5	0.0	0.533	–1.0	–1.0	2.533	–1.067
6	0.0	1.067	0.0	0.0	–1.067	2.133

The calculation columns of **L** and α_i for the first three animals are illustrated below:

$l_{11} = \sqrt{[1 - 0.25(0 + 0)]} = 1$
$\alpha_1 = 1$ and its contribution to $\mathbf{A}^{-1}$ is computed using the rules in Section 2.3.1
$l_{21} = 0$
$l_{31} = 0.5(l_{11} + l_{21}) = 0.5(1 + 0) = 0.5$
$l_{41} = 0.5(l_{11}) = 0.5$
$l_{51} = 0.5(l_{41} + l_{31}) = 0.5(0.5 + 0.5) = 0.5$
$l_{61} = 0.5(l_{51} + l_{21}) = 0.5(0.5 + 0) = 0.25$

$l_{22} = \sqrt{[1 - 0.25(0 + 0)]} = 1$
$\alpha_2 = 1$ and its contribution to $\mathbf{A}^{-1}$ is computed using the rules in Section 2.3.1
$l_{32} = 0.5(l_{12} + l_{22}) = 0.5(0 + 1) = 0.5$
$l_{42} = 0.5(l_{12}) = 0.5(0) = 0$
$l_{52} = 0.5(l_{42} + l_{32}) = 0.5(0 + 0.5) = 0.25$
$l_{62} = 0.5(l_{52} + l_{22}) = 0.5(0.25 + 1.0) = 0.625$

$$l_{33} = \sqrt{[1 - 0.25(l_{11}^2) - 0.25(l_{21}^2 + l_{22}^2)]}$$
$$= \sqrt{[1 - 0.25(1) - 0.25(0 + 1)]} = 0.707$$

$\alpha_3 = 1/(0.707)^2 = 2.0$ and its contribution to $\mathbf{A}^{-1}$ is computed using the usual rules
$l_{43} = 0.5(l_{13}) = 0.5(0) = 0$
$l_{53} = 0.5(l_{43} + l_{33}) = 0.5(0 + 0.707) = 0.354$
$l_{63} = 0.5(l_{53} + l_{23}) = 0.5(0.354 + 0) = 0.177$

Faster algorithms for computing the inverse of **A** accounting for inbreeding based on the **L** matrix have been published by Meuwissen and Luo (1992) and Quaas (1995), and these are presented in Appendix B.

2.4 Inverse of the Relationship Matrix for Sires and Maternal Grandsires

In some cases, the prediction of breeding value is only for sires and maternal grandsires, the so-called sire and maternal grandsire (MGS) model. In such cases, the $\mathbf{A}^{-1}$ to be incorporated in the mixed model equations (MME) involves only sire and maternal grandsires and the rules for calculating $\mathbf{A}^{-1}$ are different from those discussed in the previous sections relating to pedigrees with individuals, sires and dams. With the MGS model, the relationship matrix **A** required pertains to males and can be approximated (Quaas, 1984) as:

$$a_{ii} = 1 + 0.25a_{sk} \quad [2.6]$$
$$a_{ij} = 0.5a_{sj} + 0.25a_{kj} \quad [2.7]$$

where s and k are the sires and maternal grandsires, respectively, for sire i. When all maternal grandams are unrelated (base animals) and there are no maternal half-sibs, the above will yield the exact **A**.

The inverse of approximate **A** can be calculated from a list of sires and maternal grandsires, applying equation [2.6]. In this case, $\mathbf{T}^{-1}$ is a lower triangular matrix with ones in the diagonal, and the only non-zero elements to the left of the diagonal in the row for the ith animal are –0.5 and –0.25 for the columns corresponding to the sire and maternal grandsire, respectively. The elements of **D** and therefore $\mathbf{D}^{-1}$ can be calculated in a manner similar to that described in Sections 2.2 and 2.3. The diagonal elements of **D** (d_{ii}) for animal i are calculated by the following rules.

If both sire (s) and maternal grandsire (k) are known:

$$d_{ii} = \left[\text{var}(u_i) - \text{var}\left(\tfrac{1}{2}u_s + \tfrac{1}{4}u_k\right)\right]/\sigma_u^2$$

where the u terms are breeding values. Following the same arguments as in Section 2.2:

$$d_{ii} = \tfrac{11}{16} - \tfrac{1}{4}F_s - \tfrac{1}{16}F_k$$

where F_s and F_k are inbreeding coefficients for sire and maternal grandsire, respectively.

When only the maternal grandsire is known:

$$d_{ii} = \left[\text{var}(u_i) - \text{var}\left(\tfrac{1}{4}u_k\right)\right]/\sigma_u^2$$
$$d_{ii} = \tfrac{15}{16} - \tfrac{1}{16}F_k$$

When only the sire is known or no parents are known, d_{ii} is as calculated in Section 2.2.

The elements of $\mathbf{D}^{-1}$ are reciprocals of **D**, calculated above. Using equation [2.3], $\mathbf{A}^{-1}$ can be calculated on the basis of $\mathbf{T}^{-1}$ and $\mathbf{D}^{-1}$, defined above, as follows.

Initially, set $\mathbf{A}^{-1}$ to zero.

If both sire (s) and maternal grandsire (k) of animal i are known, add:

d_{ii}^{-1} to the (i,i) element
$-d_{ii}^{-1}/2$ to the (s,i) and (i,s) elements
$-d_{ii}^{-1}/4$ to the (k,i) and (i,k) elements
$d_{ii}^{-1}/4$ to the (s,s) element
$d_{ii}^{-1}/8$ to the (s,k) and (k,s) elements
$d_{ii}^{-1}/16$ to the (k,k) element

Without inbreeding, $d_{ii}^{-1} = \frac{16}{11}$.

If only the maternal grandsire (k) of animal i is known, add:

d_{ii}^{-1} to the (i,i) element
$-d_{ii}^{-1}/4$ to the (k,i) and (i,k) elements
$d_{ii}^{-1}/16$ to the (k,k) element

Without inbreeding, $d_{ii}^{-1} = \frac{16}{15}$.

If only the sire (s) of animal i is known, add:

d_{ii}^{-1} to the (i,i) element

$-d_{ii}^{-1}/2$ to the (s,i) and (i,s) elements

$d_{ii}^{-1}/4$ to the (s,s) element

Without inbreeding, $d_{ii}^{-1} = \frac{4}{3}$ in this situation, as in Section 2.3.1.

When s and k are unknown, add:

d_{ii}^{-1} *to the* (i,i) element

and $d_{ii}^{-1} = 1$.

2.4.1 An example of the inverse of the relationship matrix for sires and maternal grandsires

A pedigree consisting of sires and maternal grandsires is set up from the pedigree in Table 2.1:

Sire	Sire of sire	Maternal grandsire of sire
1	Unknown	Unknown
4	1	Unknown
5	4	1

Recoding sires 1 to n, the pedigree becomes:

1	Unknown	Unknown
2	1	Unknown
3	2	1

Using equations [2.6] and [2.7], **A** is:

$$\mathbf{A} = \begin{bmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.625 \\ 0.5 & 0.625 & 1.125 \end{bmatrix}$$

Note that the relationship among sires is the same as in **A** calculated from the full pedigree in Section 2.1.

The $\mathbf{T}^{-1}$ matrix for the pedigree is:

$$\mathbf{T}^{-1} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ -0.5 & 1.0 & 0.0 \\ -0.25 & -0.5 & 1.0 \end{bmatrix}$$

and:

$\mathbf{D} = \text{diag}(1, \frac{4}{3}, \frac{16}{11})$

Applying equation [2.3], $\mathbf{A}^{-1}$ is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.0 & -0.5 & -0.25 \\ 0.0 & 1.0 & -0.5 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{3} & 0 \\ 0 & 0 & \frac{16}{11} \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ -0.5 & 1.0 & 0.0 \\ -0.25 & -0.5 & 1.0 \end{bmatrix}$$

$$= \begin{bmatrix} 1.424 & -0.485 & -0.364 \\ -0.485 & 1.697 & -0.727 \\ -0.364 & -0.727 & 1.455 \end{bmatrix}$$

To calculate the inverse of the sire and maternal grandsire relationship matrix, applying the rules given earlier, initially set $\mathbf{A}^{-1}$ to zero. The elements of $\mathbf{D}^{-1}$ have already been given above. Processing the first animal, add 1 (d_{11}^{-1}) to the diagonal element (1,1) of $\mathbf{A}^{-1}$. For the second animal, added $\frac{4}{3}$ (d_{22}^{-1}) to the diagonal element (2,2) of $\mathbf{A}^{-1}$, $\frac{1}{3}$ to (1,1) element and $-\frac{2}{3}$ to the (1,2) and (2,1) elements. Finally, processing the third animal, add $\frac{16}{11}$ (d_{33}^{-1}) to the (3,3) element of $\mathbf{A}^{-1}$, $-\frac{16}{22}$ to the (3,4) and (4,3) elements, $-\frac{16}{44}$ to the (1,3) and (3,1) elements, $\frac{16}{44}$ to the (4,4) element, $\frac{16}{88}$ to the (1,4) and (4,1) elements and $\frac{16}{176}$ to the (1,1) element. This gives the same $\mathbf{A}^{-1}$ as previously calculated using equation [2.3].

In the next chapter, the incorporation of $\mathbf{A}^{-1}$ in the mixed model equations for the prediction of breeding value using BLUP is addressed.

3 Best Linear Unbiased Prediction of Breeding Value: Univariate Models with One Random Effect

In Chapter 1, the use of the selection index (best linear prediction) for genetic evaluation was examined; however, it is associated with some major disadvantages. First, records may have to be pre-adjusted for fixed or environmental factors and these are assumed to be known. These are not usually known, especially when no prior data exist for new subclasses of fixed effect or new environmental factors. Secondly, solutions to the index equations require the inverse of the covariance matrix for observations and this may not be computationally feasible for large data sets.

Henderson (1949) developed a methodology called best linear unbiased prediction (BLUP), by which fixed effects and breeding values can be simultaneously estimated. The properties of the methodology are similar to those of a selection index and the methodology reduces to selection indices when no adjustments for environmental factors are needed. The properties of BLUP are more or less incorporated in the name:

- Best – means it maximizes the correlation between true (a) and predicted breeding value ($\hat{a}$) or minimizes prediction error variance (PEV) (var($a - \hat{a}$)).
- Linear – predictors are linear functions of observations.
- Unbiased – estimation of realized values for a random variable, such as animal breeding values, and of estimable functions of fixed effects are unbiased (E($a = \hat{a}$)).
- Prediction – involves prediction of true breeding value.

BLUP has found widespread usage in genetic evaluation of domestic animals because of its desirable statistical properties. This has been enhanced by the steady increase in computing power and has evolved in terms of its application to simple models, such as the sire model, in its early years, and to more complex models, such as the animal, maternal, multivariate and random regression models, in recent years. Several general

purpose computer packages for BLUP evaluations, such as PEST (Groeneveld *et al.*, 1990), BREEDPLAN and a host of others, have been written and made available. In this chapter, BLUP's theoretical background is briefly presented below, considering a univariate animal model, and its application to several univariate models in genetic evaluation is illustrated.

3.1 Brief Theoretical Background

Consider the following equation for a mixed linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \qquad [3.1]$$

where:

- $\mathbf{y} = n \times 1$ vector of observations; n = number of records
- $\mathbf{b} = p \times 1$ vector of fixed effects; p = number of levels for fixed effects
- $\mathbf{a} = q \times 1$ vector of random animal effects; q = number of levels for random effects
- $\mathbf{e} = n \times 1$ vector of random residual effects
- $\mathbf{X}$ = design matrix of order $n \times p$, which relates records to fixed effects
- $\mathbf{Z}$ = design matrix of order $n \times q$, which relates records to random animal effects

Both $\mathbf{X}$ and $\mathbf{Z}$ are termed incidence matrices.

It is assumed that the expectations (E) of the variables are:

$$E(\mathbf{y}) = \mathbf{Xb};\ E(\mathbf{a}) = E(\mathbf{e}) = 0$$

and it is assumed that residual effects, which include random environmental and non-additive genetic effects, are independently distributed with variance σ_e^2; therefore, $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}$; $\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2 = \mathbf{G}$ and $\text{cov}(\mathbf{a}, \mathbf{e}) = \text{cov}(\mathbf{e}, \mathbf{a}) = 0$, where $\mathbf{A}$ is the numerator relationship matrix.

Then:

$$\begin{aligned}\text{var}(\mathbf{y}) = \mathbf{V} &= \text{var}(\mathbf{Za} + \mathbf{e}) \\ &= \mathbf{Z}\,\text{var}(\mathbf{a})\mathbf{Z}' + \text{var}(\mathbf{e}) + \text{cov}(\mathbf{Za}, \mathbf{e}) + \text{cov}(\mathbf{e}, \mathbf{Za}) \\ &= \mathbf{ZGZ}' + \mathbf{R} + \mathbf{Z}\text{cov}(\mathbf{a}, \mathbf{e}) + \text{cov}(\mathbf{e}, \mathbf{a})\mathbf{Z}'\end{aligned}$$

Since $\text{cov}(\mathbf{a}, \mathbf{e}) = \text{cov}(\mathbf{e}, \mathbf{a}) = 0$, then:

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R} \qquad [3.2]$$

$$\begin{aligned}\text{cov}(\mathbf{y}, \mathbf{a}) &= \text{cov}(\mathbf{Za} + \mathbf{e}, \mathbf{a}) \\ &= \text{cov}(\mathbf{Za}, \mathbf{a}) + \text{cov}(\mathbf{e}, \mathbf{a}) \\ &= \mathbf{Z}\,\text{cov}(\mathbf{a}, \mathbf{a}) \\ &= \mathbf{ZG}\end{aligned}$$

and:

$$\begin{aligned}\text{cov}(\mathbf{y}, \mathbf{e}) &= \text{cov}(\mathbf{Za} + \mathbf{e}, \mathbf{e}) \\ &= \text{cov}(\mathbf{Za}, \mathbf{e}) + \text{cov}(\mathbf{e}, \mathbf{e}) \\ &= \mathbf{Z}\,\text{cov}(\mathbf{a}, \mathbf{e}) + \text{cov}(\mathbf{e}, \mathbf{e}) \\ &= \mathbf{R}\end{aligned}$$

The general problem with respect to [3.1] is to predict a linear function of **b** and **a**, that is, **k′b** + **a** (predictand), using a linear function of **y**, say **L′y** (predictor), given that **k′b** is estimable. The predictor **L′y** is chosen such that it is unbiased (i.e. its expected value is equal to the expected value of the predictand) and PEV is minimized. This minimization leads to the BLUP of **a** (Henderson, 1973) as:

$$\hat{\mathbf{a}} = \text{BLUP}(\mathbf{a}) = \mathbf{GZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \qquad [3.3]$$

and:

$$\mathbf{L'y} = \mathbf{k'}\hat{\mathbf{b}} + \mathbf{GZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

where $\hat{\mathbf{b}} = (\mathbf{X'V}^{-1}\mathbf{X})\mathbf{X'V}^{-1}\mathbf{y}$, the generalized least-square solution (GLS) for **b**, and $\mathbf{k'}\hat{\mathbf{b}}$ is the best linear unbiased estimator (BLUE) of **k′b**, given that **k′b** is estimable. BLUE is similar in meaning and properties to BLUP but relates to estimates of linear functions of fixed effects. It is an estimator of the estimable functions of fixed effects that has minimum sampling variance, is unbiased and is based on the linear function of the data (Henderson, 1984). An outline for the derivation of equation [3.3] and the equation for **L′y** above are given in Appendix C.1.

As mentioned in Appendix C.1, BLUP is equivalent to the selection index with the GLS of $\hat{\mathbf{b}}$ substituted for **b** in equation [3.3]. Alternatively, this could simply be illustrated (W.G. Hill, Edinburgh, 1995, personal communication) by considering the index to compute breeding values for a group of individuals with relationship matrix **A**, which have records with known mean. From equation [1.17], the relevant matrices are then:

$$\mathbf{P} = \mathbf{I}\sigma_e^2 + \mathbf{A}\sigma_a^2 \quad \text{and} \quad \mathbf{G} = \mathbf{A}\sigma_a^2$$

with:

$$\alpha = \sigma_e^2/\sigma_a^2 \quad \text{or} \quad (1 - h^2)/h^2$$

Hence:

$$I = \mathbf{P}^{-1}\mathbf{Gy} = (\mathbf{I} + \alpha\mathbf{A}^{-1})^{-1}\mathbf{y}$$

which is similar to the BLUP equation [3.3], assuming fixed effects are absent and with $\mathbf{Z} = \mathbf{I}$.

The solutions for **a** and **b** in [3.3] require $\mathbf{V}^{-1}$, which is not always computationally feasible. However, Henderson (1950) presented the mixed model equations (MME) to estimate solutions **b** (fixed effects solutions) and to predict solutions for random effects (**a**) simultaneously without the need for computing $\mathbf{V}^{-1}$. The proof that solutions for **b** and **a** from MME are the GLS of **b** and the BLUP of **a** is given in Appendix C.2. The MME for [3.1] are:

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

assuming that **R** and **G** are non-singular. Since $\mathbf{R}^{-1}$ is an identity matrix from the earlier definition of **R** in this section, it can be factorized from

both sides of the equation to give:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \quad [3.4]$$

Note that the MME may not be of full rank, usually due to dependency in the coefficient matrix for fixed environmental effects. It may be necessary to set certain levels of fixed effects to zero when there is dependency to obtain solutions to the mixed model equations (see Section 3.5). However, the equations for **a** [3.3] are usually of full rank since **V** is usually positive definite and **Xb** is invariant to the choice of constraint.

Some of the basic assumptions of the linear model for the prediction of breeding value were given in Section 1.2. The solutions to the MME give the BLUE of **k'b** and BLUP of **a** under certain assumptions, especially when data span several generations and may be subject to selection. These assumptions are:

1. Distributions of **y**, **u** and **e** are assumed to be multivariate normal, implying that traits are determined by many additive genes of infinitesimal effects at many infinitely unlinked loci (infinitesimal model, see Section 1.1). With the infinitesimal model, changes in genetic variance resulting from selection, such as gametic disequilibrium (negative covariance between frequencies of genes at different loci), or from inbreeding and genetic drift are accounted for in the MME through the inclusion of the relationship matrix (Sorensen and Kennedy, 1983), as well as assortative mating (Kemp, 1985).
2. The variances and covariances (**R** and **G**) for the base population are assumed to be known or at least known to proportionality. In practice, variances and covariances of the base population are never known exactly but, assuming the infinitesimal model, these can be estimated by restricted maximium likelihood (REML) if data include information on which selection is based.
3. The MME can account for selection if based on a linear function of **y** (Henderson, 1975) and there is no selection on information not included in the data.

The use of these MME for the prediction of breeding values and estimation of fixed effects under an animal model is presented in the next section.

3.2 A Model for an Animal Evaluation (Animal Model)

Example 3.1
Consider the following data set (Table 3.1) for the pre-weaning gain (WWG) of beef calves.

The objective is to estimate the effects of sex and predict breeding values for all animals. Assume that $\sigma_a^2 = 20$ and $\sigma_e^2 = 40$; therefore $\alpha = \frac{40}{20} = 2$.

Table 3.1. Pre-weaning gain (kg) for five beef calves.

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	Unknown	4.5
5	Female	3	2	2.9
6	Female	1	2	3.9
7	Male	4	5	3.5
8	Male	3	6	5.0

The model to describe the observations is:

$$y_{ij} = p_i + a_j + e_{ij}$$

where: y_{ij} = the WWG of the jth calf of the ith sex, p_i = the fixed effect of the ith sex, a_j = random effect of the jth calf, and e_{ij} = random error effect. In matrix notation the model is the same as that described in equation [3.1].

3.2.1 Constructing the mixed model equations

The matrix **X** in the MME relates records to fixed (sex) effects. For the example data set, its transpose is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

The first row indicates that the first, fourth and fifth observations are from male calves and the second row shows that the second and third records are from female calves.

The **Z** matrix relates records to all animals – those with or without yield records. In this case, animals 1 to 3 are parents with no records and animals 4 to 8 are recorded. Thus, for the example data, **Z** is:

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that the first three columns of **Z** are zeros and these correspond to the animals 1 to 3, which are parents without records.

The vector **y** is simply the vector of the observations. For the data set under consideration, it is:

$$\mathbf{y}' = [4.5\ 2.9\ 3.9\ 3.5\ 5.0]$$

Having set up the matrices **X**, **Z** and **y**, the other matrices in the MME, such as **X′Z**, **Z′X**, **X′y** and **Z′y**, are easily obtained by matrix multiplication. In practice, these matrices are not calculated through multiplication from the

design matrices and vector of observations but are usually set up or computed directly. However, for the example data set, these matrices are:

$$\mathbf{X'Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$\mathbf{Z'X}$ is the transpose of $\mathbf{X'Z}$

$$\mathbf{X'y} = \begin{pmatrix} 13.0 \\ 6.8 \end{pmatrix}$$ and the transpose of $\mathbf{Z'y}$ is (0 0 0 4.5 2.9 3.9 3.5 5.0)

The matrix $\mathbf{Z'Z}$ is a diagonal matrix, with the first three diagonal elements zeros and the next five elements all ones.

The various matrices in the MME have been calculated except $\mathbf{A}^{-1}\alpha$. With these matrices, we can set up what are known as the least-square equations (LSE) as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix}$$

For the example data set, the LSE are:

$$\begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 13.0 \\ 6.8 \\ 0.0 \\ 0.0 \\ 0.0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{bmatrix}$$

The addition of $\mathbf{A}^{-1}\alpha$ to $\mathbf{Z'Z}$ in the least-square equations yields the MME. Using the rules outlined in Chapter 2, Section 2.3.1, $\mathbf{A}^{-1}$ for the example data is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.833 & 0.500 & 0.000 & -0.667 & 0.000 & -1.000 & 0.000 & 0.000 \\ 0.500 & 2.000 & 0.500 & 0.000 & -1.000 & -1.000 & 0.000 & 0.000 \\ 0.000 & 0.500 & 2.000 & 0.000 & -1.000 & 0.500 & 0.000 & -1.000 \\ -0.667 & 0.000 & 0.000 & 1.833 & 0.500 & 0.000 & -1.000 & 0.000 \\ 0.000 & -1.000 & -1.000 & 0.500 & 2.500 & 0.000 & -1.000 & 0.000 \\ -1.000 & -1.000 & 0.500 & 0.000 & 0.000 & 2.500 & 0.000 & -1.000 \\ 0.000 & 0.000 & 0.000 & -1.000 & -1.000 & 0.000 & 2.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & -1.000 & 0.000 & 2.000 \end{bmatrix}$$

and $\mathbf{A}^{-1}\alpha$ is easily obtained by multiplying every element of $\mathbf{A}^{-1}$ by 2, the value of α. Adding $\mathbf{A}^{-1}\alpha$ to $\mathbf{Z'Z}$, the MME for the example data are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 3.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 4.000 & 0.000 & -2.000 & 1.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & -1.333 & 0.000 & 0.000 & 4.667 & 1.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & 0.000 & -2.000 & -2.000 & 1.000 & 6.000 & 0.000 & -2.000 & 0.000 \\ 0.000 & 1.000 & -2.000 & -2.000 & 1.000 & 0.000 & 0.000 & 6.000 & 0.000 & -2.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & -2.000 & -2.000 & 0.000 & 5.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & -2.000 & 0.000 & 0.000 & -2.000 & 0.000 & 5.000 \end{bmatrix}^{-1} \begin{bmatrix} 13.0 \\ 6.8 \\ 0.0 \\ 0.0 \\ 0.0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{bmatrix}$$

Solving the MME by direct inversion of the coefficient matrix gives the following solutions:

Effects	Solutions
*Sex**	
1	4.358
2	3.404
Animal	
1	0.098
2	–0.019
3	–0.041
4	–0.009
5	–0.186
6	0.177
7	–0.249
8	0.183

*1 = male, 2 = female (throughout chapter)

The solutions indicate that male calves have a higher rate of gain up to weaning than female calves, which is consistent with the raw averages for males and females. From the first row in the MME (equation [3.4]), the equations for sex effect are:

$$\mathbf{(X'X)\hat{b} = X'y - (X'Z)\hat{a}}$$
$$\mathbf{\hat{b} = (X'X)^{-1} X'(y - Z\hat{a})}$$

Thus the solution for the *i*th level of sex effect may be written as:

$$\hat{b}_i = \left(\sum_j y_{ij} - \sum_j \hat{a}_{ij} \right) \Big/ \text{diag}_i \qquad [3.5]$$

where y_{ij} is the record and $\hat{a}_{ij}$ is the solution of the *j*th animals within the sex subclass *i* and diag_i is the sum of observations for the sex subclass *i*.

For instance, the solution for male calves is:

$$b_1 = [(4.5 + 3.5 + 5.0) - (-0.009 + -0.249 + 0.183)]/3 = 4.358$$

The equations for animal effects from the second row of equation [3.4] are:

$$(\mathbf{Z'Z} + \mathbf{A}^{-1}\alpha)\hat{\mathbf{a}} = \mathbf{Z'y} - (\mathbf{Z'X})\hat{\mathbf{b}}$$
$$(\mathbf{Z'Z} + \mathbf{A}^{-1}\alpha)\hat{\mathbf{a}} = \mathbf{Z'}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$
$$(\mathbf{Z'Z} + \mathbf{A}^{-1}\alpha)\hat{\mathbf{a}} = (\mathbf{Z'Z})\mathbf{YD} \qquad [3.6]$$

with $\mathbf{YD} = (\mathbf{Z'Z})^{-1}\ \mathbf{Z'}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$, where **YD** is the vector of yield deviations and represents the yields of the animal adjusted for all effects other than genetic merit and error. The matrix $\mathbf{A}^{-1}$ has non-zero off-diagonals only for the animal's parents, progeny and mates (see Chapter 2, Section 2.3); transferring off-diagonal terms to the right-hand side of [3.6] gives the equation for animal *i* with *k* progeny as:

$$(\mathbf{Z'Z} + u_{ii}\,\alpha)\hat{a}_i = \alpha u_{ip}(\hat{a}_s + \hat{a}_d) + (\mathbf{Z'Z})YD + \alpha\sum_k u_{im}(\hat{a}_{anim} - 0.5\hat{a}_m)$$

where u_{ip} is the element of the $\mathbf{A}^{-1}$ between animal *i* and its parents with the sign reversed and u_{im} is the element of $\mathbf{A}^{-1}$ between the animal and the dam of the *k*th progeny.

Therefore:

$$(\mathbf{Z'Z} + u_{ii}\,\alpha)\hat{a}_i = \alpha u_{par}(PA) + (\mathbf{Z'Z})YD + 0.5\alpha\sum_k u_{prog}(2\hat{a}_{anim} - \hat{a}_m) \qquad [3.7]$$

where *PA* is the parent average, $u_{par} = 2(u_{ip})$, with u_{ip} equals 1, $\frac{2}{3}$ or $\frac{1}{2}$ if both, one or neither parents are known and $u_{prog} = 2(u_{im})$, with u_{im} equals 1 when the mate of animal *i* is known or $\frac{2}{3}$ when the mate is not known.

Multiplying both sides of the equation by $(\mathbf{Z'Z} + u_{ii}\,\alpha)^{-1}$ (VanRaden and Wiggans, 1991) gives:

$$a_i = n_1(PA) + n_2(YD) + n_3(PC) \qquad [3.8]$$

where $PC = \sum_k u_{prog}\ (2\hat{a}_{anim} - \hat{a}_m)/\sum_k u_{prog}$ is regarded as the progeny contribution and n_1, n_2 and n_3 are weights that sum to one. The derivation of the equation for *PC* is given in Appendix C.3. The numerators of n_1, n_2 and n_3 are αu_{par}, $\mathbf{Z'Z}$ (number of records the animal has) and $0.5\alpha\Sigma_k u_{prog}$, respectively. The denominator of all three *n* terms is the sum of the three numerators.

From equation [3.8], the breeding value for an animal is dependent on the amount of information available on that animal. For base animals, *YD* in the equation does not exist and $\hat{a}_s$ and $\hat{a}_d$ are zeros with no genetic groups in the model; therefore, the solutions for these animals are a function of the contributions from their progeny breeding values adjusted for the mate solutions (*PC*). For instance, the proof for sire 1 in Example 3.1 can be calculated from the contributions from its progeny (calves 4 and 6) using equation [3.8] as:

$$\hat{a}_1 = n_1(0) + n_3[(\tfrac{2}{3})(2\hat{a}_4) + (1)(2\hat{a}_6 - \hat{a}_2)]/(\tfrac{2}{3} + 1)$$
$$\hat{a}_1 = n_1(0) + n_3[(\tfrac{2}{3})(-0.018) + (1)(0.354 - (-0.019))]/(\tfrac{2}{3} + 1)$$
$$\hat{a}_1 = n_3\,(0.2166) = 0.098$$

with $n_1 = \frac{\alpha}{3.667}$ and $n_3 = 0.5\alpha(\frac{2}{3} + 1)/3.667$, where 3.667 is the sum of the numerators of n_1 and n_3. The higher breeding value for sire 1 compared with sire 3 is due to the fact that the progeny of sire 1 have higher proofs after correcting for the solutions of the mates.

The solution for an animal with a record but with no progeny depend on the average contributions from its parents and its yield deviation. Equation [3.8] reduces to:

$$a_i = n_1(PA) + n_2(YD)$$

Thus, for progeny 8, its estimated breeding value can be calculated as:

$$\begin{aligned} a &= n_1(\hat{a}_3 + \hat{a}_6)/2 + n_2(y_8 - b_1) \\ &= n_1(0.068) + n_2(5.0 - 4.358) = 0.183 \end{aligned}$$

with $n_1 = \frac{2\alpha}{5}$, $n_2 = \frac{1}{5}$ and 5 is the sum of the numerators of n_1 and n_2.

It can also be demonstrated that for an animal with a record but with no progeny its solution is a function of an estimate of Mendelian sampling (m) and parent average. From equation [c.8] in Appendix C.3, the solution for calf i with sire s and dam d can be written as:

$$(1 + u_{ii}\alpha)\hat{a}_c + \alpha u_{cs}\hat{a}_s + \alpha u_{cd}\hat{a}_d = y_i$$

Therefore:

$$\hat{a}_c = (1 + u_{ii}\alpha)^{-1}[y_i - \alpha u_{is}\hat{a}_s - \alpha u_{id}\hat{a}_d]$$

If there is no inbreeding, $u_{is} = u_{id} = -0.5u_{ii}$. Therefore:

$$\begin{aligned} \hat{a}_i &= (1 + u_{ii}\alpha)^{-1}[y_i + 0.5u_{ii}\alpha(\hat{a}_s + \hat{a}_d)] \\ &= (1 + u_{ii}\alpha)^{-1}[(y_i - 0.5(\hat{a}_s + \hat{a}_d)) + 0.5(1 + u_{ii}\alpha)(\hat{a}_s + \hat{a}_d)] \\ &= (1 + u_{ii}\alpha)^{-1}[(y_i - 0.5(\hat{a}_s + \hat{a}_d)) + 0.5(\hat{a}_s + \hat{a}_d)] \\ \hat{a}_i &= 0.5(\hat{a}_s + \hat{a}_d) + m_i \end{aligned} \qquad [3.9]$$

where $m_i = k(y_i - 0.5\hat{a}_s - 0.5\hat{a}_d)$, is an estimate of Mendelian sampling and $k = 1/(1 + d^{-1}\alpha)$, with $d = \frac{1}{2}$ if both parents of animal i are known or $\frac{3}{4}$ if only one parent is known. Alternatively, the weight (k) can also be derived as:

$$k = \text{cov}(m, y_c)/\text{var}(y_c) = \text{cov}(m, m + e)/(\text{var}(m) + \text{var}(e))$$

where y_c is the yield record corrected for fixed effects and parent average.

$$\begin{aligned} k &= \text{var}(m)/(\text{var}(m) + \text{var}(e)) \\ &= d\sigma_a^2/(d\sigma_a^2 + \sigma_e^2) \\ &= dh^2/(dh^2 + (1 - h^2)) \end{aligned}$$

where d, as defined earlier, equals $\frac{1}{2}$, $\frac{3}{4}$ or 1 if both, one or no parents are known, respectively. Using the parameters for Example 3.1 and assuming both parents are known, $k = 10/(10 + 40) = 0.2$.

Thus for progeny 8, its estimated breeding value using [3.9], can be calculated as:

$$\begin{aligned}\hat{a}_8 &= 0.5(\hat{a}_3 + \hat{a}_6) + k(y_3 - b_1 - 0.5(\hat{a}_3 + \hat{a}_6)) \\ &= 0.5(-0.041 + 0.177) + 0.2(5.0 - 4.358 - 0.5(-0.041 + 0.177)) \\ &= 0.183\end{aligned}$$

Compared with calf 7, the proof of calf 8 is higher because it has a higher parent average solution and higher estimate of Mendelian sampling.

In the case of an animal with a record and having progeny, there is an additional contribution from its offspring to its breeding value. Thus the breeding values of progeny 4 and 6 using [3.8] are:

$$\begin{aligned}\hat{a}_4 &= n_1(\hat{a}_1/2) + n_2(y_4 - b_1) + n_3((2(\hat{a}_7) - \hat{a}_5)) \\ &= n_1(0.098/2) + n_2(4.5 - 4.358) + n_3((2(-0.249) - (-0.186)) = -0.009\end{aligned}$$

with $n_1 = 2\alpha(\frac{2}{3})/4.667$, $n_2 = 1/4.667$ and $n_3 = 0.5\alpha/4.667$; 4.676 = the sum of the numerators of n_1, n_2 and n_3; and:

$$\begin{aligned}\hat{a}_6 &= n_1((\hat{a}_1 + \hat{a}_2)/2) + n_2(y_6 - b_2) + n_3(2(\hat{a}_8) - \hat{a}_3) \\ &= n_1((0.098 + -0.019)/2) + n_2(3.9 - 3.404) + n_3(2(0.183) - (-0.041)) \\ &= 0.177\end{aligned}$$

with $n_1 = \frac{2\alpha}{6}$, $n_2 = \frac{1}{6}$ and $n_3 = \frac{0.5\alpha}{6}$; 6 = the sum of the numerators of n_1, n_2 and n_3.

Although contributions from parent average to both calves are similar, differences in progeny contributions resulted in a higher breeding value for calf 6, accounting for about 75% of the difference in the predicted breeding values between both calves.

3.2.2 Progeny (daughter) yield deviation

The yield deviation of a progeny contributes indirectly to the breeding value of its sire after it has been combined with information from parents and the offspring of the progeny (see equation [3.8]). Thus progeny contribution is a regressed measure and it is not an independent measure of progeny performance as information from parents and the progeny's offspring is included. VanRaden and Wiggans (1991) indicated that a more independent and unregressed measure of progeny performance is progeny yield deviation (PYD). However, they called it daughter yield deviation (DYD) as they were dealing with the dairy cattle situation and records were only available for daughters of bulls. PYD or DYD can simply be defined as a weighted average of corrected YD of all progeny of a sire; the correction is for all fixed effects and the breeding values of the mates of the sire.

DYD has been used for various purposes in dairy cattle evaluation and research. It was used in the early 1990s for the calculation of conversion equations to convert bull evaluations across several countries (Goddard, 1985). It was initially the variable of choice for international evaluations

of dairy bulls by Interbull but, due to the inability of several countries to calculate DYD, deregressed proofs were used (Sigurdsson and Banos, 1995). In addition, Interbull methods for the validation of genetic trends in national evaluations prior to acceptance for international evaluations utilize DYDs (Boichard *et al.*, 1995). DYDs are also commonly employed in dairy cattle studies aimed at detecting quantitative trait loci using the granddaughter design (Weller, 2001). The equation for calculating DYD from univariate animal model evaluations was presented by VanRaden and Wiggans (1991) and its derivation is briefly outlined.

For the progeny (*prog*) of a bull *i* that has no offspring of her own, equation [3.8] becomes:

$$\hat{a}_{prog} = n_{1prog}\text{PA} + n_{2prog}\text{YD} \qquad [3.10]$$

Substituting [3.10] into the equation for PC in [3.8] gives:

$$\begin{aligned}\text{PC} &= \sum_k u_{prog}\left[2(n_{1prog}\text{PA} + n_{2prog}\text{YD}) - \hat{a}_{mi}\right] \Big/ \sum_k u_{prog} \\ &= \sum_k u_{prog}\left[n_{1prog}(\hat{a}_i + \hat{a}_{mi}) + n_{2prog}2\text{YD} - \hat{a}_{mi}\right] \Big/ \sum_k u_{prog}\end{aligned}$$

where n_{1prog} and n_{2prog} are the n_1 and n_2 of progeny. Since these progeny have no offspring of their own, n_{3prog} equals zero; therefore n_{1prog} equals $1 - n_{2prog}$. Then:

$$\begin{aligned}\text{PC} &= \sum_k u_{prog}\left[(1 - n_{2prog})(\hat{a}_i + \hat{a}_{mi}) + n_{2prog}2\text{YD} - \hat{a}_{mi}\right] \Big/ \sum_k u_{prog} \\ &= \sum_k u_{prog}\left[(1 - n_{2prog})\hat{a}_i + n_{2prog}(2\text{YD} - \hat{a}_{mi})\right] \Big/ \sum_k u_{prog} \\ &= \hat{a}_i + \sum_k u_{prog}\left[n_{2prog}(-\hat{a}_i + 2\text{YD} - \hat{a}_{mi})\right] \Big/ \sum_k u_{prog} \qquad [3.11]\end{aligned}$$

Substituting [3.11] into [3.8] and accumulating all terms involving $\hat{a}_i$ to the left side gives:

$$\begin{aligned}&\hat{a}_i - n_3\hat{a}_i + n_3 \sum_k u_{prog} n_{2prog}\hat{a}_i \Big/ \sum_k u_{prog} \\ &\quad = n_1\text{PA} + n_2\text{YD} + n_3 \sum_k u_{prog} n_{2prog}(2\text{YD} - \hat{a}_{mi}) \Big/ \sum_k u_{prog}\end{aligned}$$

Therefore:

$$\begin{aligned}&\left((1 - n_3 + n_3 \sum_k u_{prog} n_{2prog} \Big/ \sum_k u_{prog}\right)\hat{a}_i \\ &\quad = n_1\text{PA} + n_2\text{YD} + n_3 \sum_k u_{prog} n_{2prog}(2\text{YD} - \hat{a}_{mi}) \Big/ \sum_k u_{prog}\end{aligned}$$

Substituting $(n_1 + n_2)$ for $1 - n_3$ and removing the common denominator of the *n* terms from both sides of the equation, with DYD as:

$$\text{DYD or PYD} = \sum_k u_{prog} n_{2prog}(2\text{YD} - \hat{a}_m) \Big/ \sum_k u_{prog} n_{2prog} \qquad [3.12]$$

the breeding value of animal *i* can be expressed as:

$$\hat{a}_i = w_1\text{PA} + w_2\text{YD} + w_3\text{DYD} \qquad [3.13]$$

where the weights w_1, w_2 and w_3 sum to unity. The numerators of w_1 and w_2 are equal to those of n_1 and n_2 in equation [3.8]. The numerator of $w_3 = 0.5\alpha\Sigma_k u_{prog}\, n_{2prog}$, which is derived as n_3 times $\Sigma_k u_{prog}\, n_{2prog}/\Sigma_k u_{prog}$. As VanRaden and Wiggans (1991) indicated, w_3 is always less than unity and therefore less than n_3, which reflects the fact that PYD or DYD is an unregressed measure of progeny performance. Note that, for bulls with granddaughters, PYD or DYD does not include information from these granddaughters. Also, in the dairy cattle situation, the information from sons is not included in the calculation of DYD.

Illustrating the calculation of PYD or DYD

The computation of DYD is usually carried out in dairy cattle evaluations and it is illustrated later for a dairy data set in Example 4.1. Using the beef data in Example 3.1, the calculation of PYD is briefly illustrated for animal 3, using information on both female and male progeny since observations are available on both sexes.

First, the YDs for both progeny of sire 3 are calculated:

$$\text{YD}_5 = (y_5 - b_2) = (2.9 - 3.404) = -0.504$$
$$\text{YD}_8 = (y_8 - b_1) = (5.0 - 4.358) = 0.642$$

Therefore, using [3.12]:

$$\begin{aligned}\text{PYD}_3 &= n_{2(5)}u_{(5)}(2\text{YD}_5 - \hat{a}_2) + n_{2(8)}u_{(8)}(2\text{YD}_8 - \hat{a}_6)/(n_{2(5)}u_{(5)} + n_{2(8)}u_{(8)}) \\ &= 0.2(1)(-1.008 - (-0.019) + 0.2(1)(1.284 - 0.177)/(0.2(1) + 0.2(1)) \\ &= 0.059\end{aligned}$$

where $n_{2(j)}$ and $u_{(j)}$ are the n_2 and u for the jth progeny. Note that, in calculating $n_{2(j)}$, it has been assumed that progeny j has no offspring. Thus $n_{2(5)} = 1/(1 + 2\alpha(1)) = 0.2$.

Using [3.12] to calculate the breeding value of sire 3 gives the value of 0.0098, with $w_1 = 0.833$ and $w_2 = 0.167$. This is different from the breeding value reported from solving the MME as the granddaughter information (calf 7) has not been included.

3.2.3 Accuracy of evaluations

The accuracy (r) of predictions is the correlation between true and predicted breeding values. However, in dairy cattle evaluations, the accuracy of evaluations is usually expressed in terms of reliability, with the squared correlation between true and predicted breeding values (r^2). The calculations for r or r^2 require the diagonal elements of the inverse of the MME, as shown by Henderson (1975).

If the coefficient matrix of the MME in [3.4] is represented as:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

and a generalized inverse of the coefficient matrix as:

$$\begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$$

Henderson (1975) showed that the PEV is:

$$\text{PEV} = \text{var}(a - \hat{a}) = C^{22}\sigma_e^2 \qquad [3.14]$$

Thus the diagonal elements of the coefficient matrix for animal equations are needed to calculate PEV for animals. The PEV could be regarded as the fraction of additive genetic variance not accounted for by the prediction. Therefore it could be expressed as:

$$\text{PEV} = C^{22}\sigma_e^2 = (1 - r^2)\sigma_a^2$$

where r^2 is the squared correlation between the true and estimated breeding values.

Thus, for animal i:

$$d_i\sigma_e^2 = (1 - r^2)\sigma_a^2$$

where d_i is the ith diagonal element of C^{22}.

$$d_i\sigma_e^2/\sigma_a^2 = 1 - r^2$$
$$r^2 = 1 - d_i\alpha$$

and the accuracy (r) is just the square root of reliability.

From [3.14], the standard error of prediction (SEP) is:

$$\text{SEP} = \sqrt{\text{var}(a - \hat{a})}$$
$$= \sqrt{d_i\sigma_e^2} \text{ for animal } i$$

The inverse of the coefficient matrix for Example 3.1 is:

$$\begin{bmatrix}
0.596 & 0.157 & -0.164 & -0.084 & -0.131 & -0.265 & -0.148 & -0.166 & -0.284 & -0.238 \\
0.157 & 0.802 & -0.133 & -0.241 & -0.112 & -0.087 & -0.299 & -0.306 & -0.186 & -0.199 \\
-0.164 & -0.133 & 0.471 & 0.007 & 0.033 & 0.220 & 0.045 & 0.221 & 0.139 & 0.134 \\
-0.084 & -0.241 & 0.007 & 0.492 & -0.010 & 0.020 & 0.237 & 0.245 & 0.120 & 0.111 \\
-0.131 & -0.112 & 0.033 & -0.010 & 0.456 & 0.048 & 0.201 & 0.023 & 0.126 & 0.218 \\
-0.265 & -0.087 & 0.220 & 0.020 & 0.048 & 0.428 & 0.047 & 0.128 & 0.243 & 0.123 \\
-0.148 & -0.299 & 0.045 & 0.237 & 0.201 & 0.047 & 0.428 & 0.170 & 0.220 & 0.178 \\
-0.166 & -0.306 & 0.221 & 0.245 & 0.023 & 0.128 & 0.170 & 0.442 & 0.152 & 0.219 \\
-0.284 & -0.186 & 0.139 & 0.120 & 0.126 & 0.243 & 0.220 & 0.152 & 0.442 & 0.168 \\
-0.238 & -0.199 & 0.134 & 0.111 & 0.218 & 0.123 & 0.178 & 0.219 & 0.168 & 0.422
\end{bmatrix}$$

The r^2, r and SEP for animals in Example 3.1 are:

Animal	Diagonals of inverse	r^2	r	SEP
1	0.471	0.058	0.241	4.341
2	0.492	0.016	0.126	4.436
3	0.456	0.088	0.297	4.271
4	0.428	0.144	0.379	4.138
5	0.428	0.144	0.379	4.138
6	0.442	0.116	0.341	4.205
7	0.442	0.116	0.341	4.205
8	0.422	0.156	0.395	4.109

In the example, the reliabilities of animals with records are generally higher than those of ancestors since each has only two progeny. The two calves in the female sex subclass are progeny of dam 2 and this may explain the very low reliability for this ancestor as the effective number of daughters is reduced. The amount of information on calves 4 and 5 is very similar. Each has a record, a common sire and parents of the same progeny: hence they have the same reliability. Calf 8 has the highest reliability and this is due to the information from the parents (its sire has another progeny and the dam has both parents known) and its record. The standard errors are large due to the small size of the data set but follow the same pattern as the reliabilities.

In practice, obtaining the inverse of the MME for large populations is not feasible and various methods have been used to approximate the diagonal element of the inverse. A methodology published by Meyer (1989) and used in the national dairy evaluation programme in Canada (Wiggans *et al.*, 1992) in the 1990s is presented in Appendix D.

3.3 A Sire Model

The application of a sire model implies that only sires are being evaluated using progeny records. Most early applications of BLUP for the prediction of breeding values, especially in dairy cattle, were based on a sire model. The main advantage with a sire model is that the number of equations is reduced compared with an animal model since only sires are evaluated. However, with a sire model, the genetic merit of the mate (dam of progeny) is not accounted for. It is assumed that all mates are of similar genetic merit and this can result in bias in the predicted breeding values if there is preferential mating.

The sire model in matrix notation is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e} \qquad [3.15]$$

All terms in [3.15] are as defined for [3.1] and **s** is the vector of random sire effects, **Z** now relates records to sires and:

$$\text{var}(\mathbf{s}) = \mathbf{A}\sigma_s^2$$
$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma_s^2 + \mathbf{R}$$

where **A** is the numerator relationship matrix for sires, $\sigma_s^2 = 0.25\sigma_a^2$ and $\mathbf{R} = I\sigma_e^2$. The MMEs are exactly the same as in [3.4] except that $\alpha = \sigma_e^2/\sigma_s^2 = (4 - h^2)/h^2$.

3.3.1 An illustration

Example 3.2
An application of a sire model is illustrated below using the same data as for the animal model evaluation in Table 3.1. Assigning records to sires and including the pedigree for sires, the data can be presented as follows.

Sex of progeny	Sire of sire	Dam of sire	Dam	WWG (kg)
Male	1	–	–	4.5
Female	3	–	–	2.9
Female	1	–	–	3.9
Male	4	1	–	3.5
Male	3	–	–	5.0

The objective is to estimate sex effects and predict breeding values for sires 1, 3 and 4. Using the same parameters as in Section 3.2.1, $\sigma_s^2 = 0.25(20) = 5$ and $\sigma_e^2 = 60 - 5 = 55$; therefore $\alpha = \frac{55}{5} = 11$.

Setting the design matrices and MME

The design matrix **X** relating records to sex is as defined in Section 3.2.1. However, **Z** is different and its transpose is:

$$\mathbf{Z}' = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

indicating that sires 1 and 3 have two records each while sire 4 has only one record. The vector of observations **y** is as defined in Section 3.1. The matrices **X′X**, **X′Z**, **Z′X**, **Z′Z**, **X′y** and **Z′y** in the MME can easily be calculated through matrix multiplication. Thus:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{X}'\mathbf{Z} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$
$$\mathbf{Z}'\mathbf{Z} = \text{diag}(2, 2, 1)$$

$\mathbf{X'y}$ is as in Section 3.1 and the transpose of $\mathbf{Z'y}$ is:

$(\mathbf{Z'y})' = [8.4\ 7.9\ 3.5]$

The least-square equations are:

$$\begin{bmatrix} 3 & 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{s}_1 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \begin{bmatrix} 13.00 \\ 6.80 \\ 8.40 \\ 7.90 \\ 3.50 \end{bmatrix}$$

Apart from the fact that sire 4 is the son of sire 1, no other relationships exist among the three sires. Therefore $\mathbf{A}^{-1}$ for the three sires is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.333 & 0.0 & -0.667 \\ 0.000 & 1.0 & 0.000 \\ -0.667 & 0.0 & 1.333 \end{bmatrix}$$

The MME obtained after adding $\mathbf{A}^{-1}\alpha$ to $\mathbf{Z'Z}$ in the LSEs are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{s}_1 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} \begin{bmatrix} 3.000 & 0.000 & 1.000 & 1.000 & 1.000 \\ 0.000 & 2.000 & 1.000 & 1.000 & 0.000 \\ 1.000 & 1.000 & 16.666 & 0.000 & -7.334 \\ 1.000 & 1.000 & 0.000 & 13.000 & 0.000 \\ 1.000 & 0.000 & -7.334 & 0.000 & 15.666 \end{bmatrix}^{-1} = \begin{bmatrix} 13.00 \\ 6.80 \\ 8.40 \\ 7.90 \\ 3.50 \end{bmatrix}$$

The solutions to the MME by direct inversion of the coefficient matrix are:

Effects	Solutions
Sex	
1	4.336
2	3.382
Sire	
1	0.022
3	0.014
4	–0.043

The difference between solutions for sex subclasses, $\mathbf{L'b}$, where $\mathbf{L}$ is [1 – 1], is the same as in the animal model. However, sire proofs and differences between sire proofs $(\hat{s}_i - \hat{s}_j)$ are different from those in the animal model, although the ranking for the three sires is the same in both models. The differences in the proofs are due to the lack of adjustment for breeding values of mates in the sire model and differences in progeny contributions under both models. In this example, most of the differences in sire solutions under both models are due to differences in

progeny contributions. The proofs for these sires under the animal model are based on their progeny contributions, since their parents are unknown. This contribution from progeny includes information from progeny yields and those of grand-offspring of the sires. However, in the sire model, progeny contributions include information from only male grand-offspring of the sires in addition to progeny yields. The effect of this difference on sire proofs under the two models is illustrated for two bulls below.

From the calculations in Section 3.2.1, the proportionate contributions of calves 4 and 6 to the proof of sire 1 in the animal model are –0.003 and 0.102, respectively. Using equation [3.8], the contributions from the different yield records to sire 1 under the sire model are as follows.

Contributions (CONT) from yields for calves 4 and 6 are:

$$\text{CONT}_4 = n_2(0.082) = 0.010$$
$$\text{CONT}_6 = n_2(0.259) = 0.031$$

where $n_2 = 2/16.667$.

Contribution from yield record for male grand-progeny (calf 7) through animal 4 (progeny) is:

$$\text{CONT}_7 = n_3(-0.086) = -0.019$$

where $n_3 = 3.667/16.667$.

Therefore:

$$\hat{s}_i = \text{CONT}_4 + \text{CONT}_6 + \text{CONT}_7 = 0.022$$

In the sire model the sum of CONT_4 and CONT_7 is equivalent to the contribution from calf 4 to the sire proof in the animal model. Thus the main difference in the proof for sire 1 in the two models is due largely to the lower contribution of calf 6 in the sire model. This lower contribution arises from the fact that the contribution is only from the yield record in the sire model while it is from the yield and the progeny of calf 6 in the animal model.

Similar calculations for sire 3 indicate that the proportionate contributions from its progeny are –0.088 for calf 5 and 0.047 for calf 8 in the animal model. However, in the sire model the contributions are –0.037 and 0.051, respectively, from the yield of these calves. Again, the major difference here is due to the contribution from calf 5, which contains information from her offspring (calf 7) in the animal model. The similarity of the contributions of calf 8 to the proof of sire 3 in both models is because it is a non-parent and the contribution is slightly higher under the sire model due to the lack of adjustment for the breeding value of the mate.

3.4 Reduced Animal Model

In Section 3.1, the best linear unbiased prediction of breeding value involved setting up equations for every animal, that is, all parents and

progeny. Thus the order of the animal equations was equal to the number of animals being evaluated. If equations were set up only for parents, this would greatly reduce the number of equations to be solved, especially since the number of parents is usually less than the number of progeny in most data sets. Breeding values of progeny can be obtained by back-solving from the predicted parental breeding values. Quaas and Pollak (1980) developed the reduced animal model (RAM), which allowed equations to be set up only for parents in the MME, and breeding values of progeny are obtained by back-solving from the predicted parental breeding values. This section presents the theoretical background for the RAM and illustrates its use for the prediction of breeding values.

3.4.1 Defining the model

The application of a RAM involves setting up animal equations for parents only and representing the breeding values of non-parents in terms of parental breeding value. Thus, for the non-parent i, its breeding value can be expressed as:

$$a_i = \tfrac{1}{2}(a_s + a_d) + m_i \qquad [3.16]$$

where a_s and a_d are the breeding values of sire and dam and m_i is the Mendelian sampling. It was shown in Chapter 2, Section 2.2 that:

$$\text{var}(m_i) = (0.5 - 0.25(F_s + F_d))\sigma_a^2$$

Let $F = (F_s + F_d)/2$, then:

$$\begin{aligned}\text{var}(m_i) &= (0.5 - 0.5(F))\sigma_a^2 \\ &= 0.5(1 - F)\sigma_a^2 \end{aligned} \qquad [3.17]$$

The animal model applied in Section 3.2 was:

$$y_{ij} = p_i + a_j + e_{ij} \qquad [3.18]$$

In matrix notation:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \qquad [3.19]$$

The terms in the above equations have been defined in Section 3.2.

Using [3.16], equation [3.18] can be expressed as:

$$y_{ij} = p_i + \tfrac{1}{2}a_s + \tfrac{1}{2}a_d + m_j + e_{ij} \qquad [3.20]$$

For non-parents, the terms m_j and e_{ijk} can be combined to form a single residual term e^*_{ijk} as:

$$e^*_{ijk} = m_j + e_{ijk} \qquad [3.21]$$

and:

$$\text{var}(e^*_{ijk}) = \text{var}(m_j) + \text{var}(e_{ijk})$$

Using [3.17]:

$$\text{var}(e^*_{ijk}) = \tfrac{1}{2}(1 - F)\sigma_a^2 + \sigma_e^2$$

In general:

$$\text{var}(m_j) = d_j(1 - F_j)\sigma_a^2 \qquad [3.22]$$

where d_j equals $\frac{1}{2}$ or $\frac{3}{4}$ or 1 if both, one or no parents are known, respectively, and, F_j is the average inbreeding for both parents or, if only one parent is known, it is the inbreeding coefficient of the known parent. F_j equals zero when no parent is known. Ignoring inbreeding:

$$\text{var}(e^*_{ijk}) = \sigma_e^2 + d_j\sigma_a^2 = (1 + d_j\alpha^{-1})\sigma_e^2$$

Equation [3.20] can be expressed in matrix notation as:

$$\mathbf{y} = \mathbf{X}_n\mathbf{b} + \mathbf{Z}_1\mathbf{a}_p + \mathbf{e}^* \qquad [3.23]$$

where $\mathbf{X}_n$ is the incidence matrix that relates non-parents' records to fixed effects, $\mathbf{Z}_1$ is an incidence matrix of zeros and halves identifying the parents of animals and $\mathbf{a}_p$ is a vector of breeding values of parents.

The application of RAM involves applying the model:

$$\mathbf{y}_p = \mathbf{X}_p\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

for parents and the model:

$$\mathbf{y}_n = \mathbf{X}_n\mathbf{b} + \mathbf{Z}_1\mathbf{a}_p + \mathbf{e}^*$$

for non-parents.

From the above two equations, the model for RAM analysis can be written as:

$$\begin{bmatrix}\mathbf{y}_p\\ \mathbf{y}_n\end{bmatrix} = \begin{bmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{bmatrix}\mathbf{b} + \begin{bmatrix}\mathbf{Z}\\ \mathbf{Z}_1\end{bmatrix}\mathbf{a}_p + \begin{bmatrix}\mathbf{e}\\ \mathbf{e}^*\end{bmatrix}$$

If:

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix}\mathbf{Z}\\ \mathbf{Z}_1\end{bmatrix} \quad \text{and} \quad \mathbf{R} = \begin{bmatrix}\mathbf{R}_p\\ \mathbf{R}_n\end{bmatrix} = \begin{bmatrix}\mathbf{I}\sigma_e^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_e^{2*}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I} + \mathbf{D}\alpha^{-1}\end{bmatrix}\sigma_e^2$$

then:

$$\text{var}(\mathbf{y}) = \mathbf{W}\mathbf{A}_p\mathbf{W}'\sigma_e^2 + \mathbf{R}$$
$$\text{var}(\mathbf{a}_p) = \mathbf{A}_p\sigma_a^2$$

where $\mathbf{A}_p$ is the relationship matrix among parents and $\mathbf{D}$ above is a diagonal matrix with elements as defined for d_j in equation [3.22].

The mixed model equations to be solved are:

$$\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{a}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W}\\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{A}^{-1}1/\sigma_a^2\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix} \qquad [3.24]$$

Equation [3.24] can also be written as

$$\begin{bmatrix} \mathbf{X}'_p\mathbf{R}_p^{-1}\mathbf{X}_p + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{X}'_p\mathbf{R}_p^{-1}\mathbf{Z} + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{Z}_1 \\ \mathbf{Z}'\mathbf{R}_p^{-1}\mathbf{X}_p + \mathbf{Z}'_1\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}'\mathbf{R}_p^{-1}\mathbf{Z} + \mathbf{Z}'_1\mathbf{R}_n^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1}\,1/\sigma_a^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_p\mathbf{R}_p^{-1}\mathbf{y}_p + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}'\mathbf{R}_p^{-1}\mathbf{y}_p + \mathbf{Z}'_1\mathbf{R}_n^{-1}\mathbf{y}_n \end{bmatrix}$$

Multiplying the equations above by $\mathbf{R}_p$ gives:

$$\begin{bmatrix} \mathbf{X}'_p\mathbf{X}_p + \mathbf{X}'_n\mathbf{R}_v^{-1}\mathbf{X}_n & \mathbf{X}'_p\mathbf{Z} + \mathbf{X}'_n\mathbf{R}_v^{-1}\mathbf{Z}_1 \\ \mathbf{Z}'\mathbf{X}_p + \mathbf{Z}'_1\mathbf{R}_v^{-1}\mathbf{X}_n & \mathbf{Z}'\mathbf{Z} + \mathbf{Z}'_1\mathbf{R}_v^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_p\mathbf{y}_p + \mathbf{X}'_n\mathbf{R}_v^{-1}\mathbf{y}_n \\ \mathbf{Z}'\mathbf{y}_p + \mathbf{Z}'_1\mathbf{R}_v^{-1}\mathbf{y}_n \end{bmatrix} \quad [3.25]$$

where $\mathbf{R}_v^{-1} = 1/(1 + \mathbf{D}\alpha^{-1})$.

3.4.2 An illustration

Example 3.3

The application of RAM using equation [3.24] for the prediction of breeding values is illustrated below with the same data set (Table 3.1) as in Example 3.1 for the animal model evaluation. The genetic parameters are $\sigma_a^2 = 20.0$ and $\sigma_e^2 = 40.0$.

CONSTRUCTING THE MME

First, we need to set up $\mathbf{R}$, the matrix of residual variances, and its inverse. In the example data set, animals 4, 5 and 6 are parents; therefore the diagonal elements in $\mathbf{R}$ corresponding to these animals are equal to σ_e^2, that is, 40.0. Calves 7 and 8 are non-parents; therefore the diagonal elements for these animals in $\mathbf{R}$ are equal to $\sigma_e^2 + d_i\sigma_a^2$, assuming that the average inbreeding coefficients of the parents of these animals equal zero. For each calf, d_i equals half because both their parents are known; therefore $r_{77} = r_{88} = 40 + \frac{1}{2}(20) = 50$.

The matrix $\mathbf{R}$ for animals with records is:

$$\mathbf{R} = \text{diag}(40, 40, 40, 50, 50)$$

and:

$$\mathbf{R}^{-1} = \text{diag}(0.025, 0.025, 0.025, 0.020, 0.020)$$

The matrix $\mathbf{X}$ is the same as in Section 3.2 and relates records to sex effects.

Therefore:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{bmatrix} 0.065 & 0.000 \\ 0.000 & 0.050 \end{bmatrix}$$

For the matrix **W**, the rows for parents with records (animals 4, 5 and 6) consist of zeros except for the columns corresponding to these animals, which contain ones, indicating that they have records. However, the rows for non-parents with records (animals 7 and 8) contain halves in the columns that correspond to their parents and otherwise zeros. Thus:

$$\mathbf{W} = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.5 \end{bmatrix}$$

and:

$$\mathbf{W'R^{-1}W} = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.005 & 0.0 & 0.0 & 0.005 \\ 0.0 & 0.0 & 0.0 & 0.03 & 0.005 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.005 & 0.03 & 0.0 \\ 0.0 & 0.0 & 0.005 & 0.0 & 0.0 & 0.03 \end{bmatrix}$$

The transpose of the vector of observations, **y**, is as defined in Section 3.2. The remaining matrices, $\mathbf{X'R^{-1}W}$, $\mathbf{W'R^{-1}X}$, $\mathbf{X'R^{-1}y}$ and $\mathbf{Z'R^{-1}y}$, can easily be calculated through matrix multiplication since **X**, $\mathbf{R^{-1}}$, **W** and **y** have been set up. Therefore:

$$\mathbf{X'R^{-1}W} = \begin{bmatrix} 0.000 & 0.000 & 0.010 & 0.035 & 0.010 & 0.010 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.025 & 0.025 \end{bmatrix}$$

The matrix $\mathbf{W'R^{-1}X}$ is the transpose of $\mathbf{X'R^{-1}W}$.

$$\mathbf{X'R^{-1}y} = \begin{bmatrix} 0.282 \\ 0.170 \end{bmatrix} \quad \text{and} \quad \mathbf{W'R^{-1}y} = \begin{bmatrix} 0.000 \\ 0.000 \\ 0.050 \\ 0.148 \\ 0.107 \\ 0.148 \end{bmatrix}$$

The LSEs are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 0.065 & 0.000 & 0.000 & 0.000 & 0.010 & 0.035 & 0.010 & 0.010 \\ 0.000 & 0.050 & 0.000 & 0.000 & 0.000 & 0.000 & 0.025 & 0.025 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.010 & 0.000 & 0.000 & 0.000 & 0.005 & 0.000 & 0.000 & 0.005 \\ 0.035 & 0.000 & 0.000 & 0.000 & 0.000 & 0.030 & 0.005 & 0.000 \\ 0.010 & 0.025 & 0.000 & 0.000 & 0.000 & 0.005 & 0.030 & 0.000 \\ 0.010 & 0.025 & 0.000 & 0.000 & 0.005 & 0.000 & 0.000 & 0.030 \end{bmatrix}^{-1} \begin{bmatrix} 0.282 \\ 0.170 \\ 0.000 \\ 0.000 \\ 0.050 \\ 0.148 \\ 0.107 \\ 0.148 \end{bmatrix}$$

The relationship matrix is only for parents, that is, animals 1 to 6. Thus:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.833 & 0.500 & 0.000 & -0.667 & 0.000 & -1.000 \\ 0.500 & 2.000 & 0.500 & 0.000 & -1.000 & -1.000 \\ 0.000 & 0.500 & 1.500 & 0.000 & -1.000 & 0.000 \\ -0.667 & 0.000 & 0.000 & 1.333 & 0.000 & 0.000 \\ 0.000 & -1.000 & -1.000 & 0.000 & 2.000 & 0.000 \\ -1.000 & -1.000 & 0.000 & 0.000 & 0.000 & 2.000 \end{bmatrix}$$

Adding $\mathbf{A}^{-1}1/\sigma_a^2$ to the $\mathbf{W'R}^{-1}\mathbf{W}$ of the least-square equations gives the MME, which are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 0.065 & 0.000 & 0.000 & 0.000 & 0.010 & 0.035 & 0.010 & 0.010 \\ 0.000 & 0.050 & 0.000 & 0.000 & 0.000 & 0.000 & 0.025 & 0.025 \\ 0.000 & 0.000 & 0.092 & 0.025 & 0.000 & -0.033 & 0.000 & -0.050 \\ 0.000 & 0.000 & 0.025 & 0.100 & 0.025 & 0.000 & -0.050 & -0.050 \\ 0.010 & 0.000 & 0.000 & 0.025 & 0.080 & 0.000 & -0.050 & 0.005 \\ 0.035 & 0.000 & -0.033 & 0.000 & 0.000 & 0.097 & 0.005 & 0.000 \\ 0.010 & 0.025 & 0.000 & -0.050 & -0.050 & 0.005 & 0.130 & 0.000 \\ 0.010 & 0.025 & -0.050 & -0.050 & 0.005 & 0.000 & 0.000 & 0.130 \end{bmatrix}^{-1} \begin{bmatrix} 0.282 \\ 0.170 \\ 0.000 \\ 0.000 \\ 0.050 \\ 0.148 \\ 0.107 \\ 0.148 \end{bmatrix}$$

The solutions are:

Effects	Solutions
Sex	
1	4.358
2	3.404
Animal	
1	0.098
2	–0.019
3	–0.041
4	–0.009
5	–0.186
6	0.177

The solutions for sex effects and proofs for parents are exactly as obtained using the animal model in Example 3.1. However, the number of non-zero elements in the coefficient matrix is 38 compared with 46 for an animal model in Section 3.2 on the same data set. This difference will be more marked in large data sets or in data sets where the number of progeny far exceeds the number of parents. This is one of the main advantages of the reduced animal model, as the number of equations and therefore

non-zero elements to be stored are reduced. The solutions for non-parents can be obtained by back-solving, as discussed in the next section.

SOLUTIONS FOR NON-PARENTS

With the reduced animal model, solutions for non-parents are obtained by back-solving, using the solutions for the fixed effects and parents. Equation [3.9], derived earlier from the MME for an animal with its parents, can be used to back-solve for non-parents solutions. Since $\mathbf{R}^{-1}$ was not factored out of these MME, equation [3.9] can now be expressed as:

$$k = r^{11}/r^{11} + d_i^{-1}g^{-1} \qquad [3.26]$$

Solutions for non-parents in Example 3.3 can be solved using equation 3.9 but with k expressed as in [3.26]. However, because there is a fixed effect in the model, $m_i = k(y_c - b_j - 0.5a_s - 0.5a_d)$. In Example 3.3, both parents of non-parents (animals 7 and 8) are known; therefore:

$$k = 0.025/(0.025 + (2)0.05) = 0.20$$

Solutions for calves 7 and 8 are:

$$\begin{aligned}\hat{a}_7 &= 0.5(-0.009 + -0.186) + 0.20(3.5 - 4.358 - 0.5(-0.009 + -0.186)) \\ &= -0.249 \\ \hat{a}_8 &= 0.5(-0.041 + 0.177) + 0.20(5.0 - 4.358 - 0.5(-0.041 + 0.177)) \\ &= 0.183\end{aligned}$$

Again, these solutions are the same for these animals as under the animal model.

3.4.3 An alternative approach

Note that, if the example data had been analysed using equation [3.25], the design matrices would be of the following form:

$$\mathbf{X}'_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{X}'_n = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$\mathbf{Z}$ including ancestors is:

$$\mathbf{Z} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\mathbf{Z}_1$ is:

$$\mathbf{Z}_1 = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0.5 \end{bmatrix}$$

The remaining matrices can be calculated through matrix multiplication. The MME then are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = \begin{bmatrix} 2.600 & 0.000 & 0.000 & 0.000 & 0.400 & 1.400 & 0.400 & 0.400 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 \\ 0.400 & 0.000 & 0.000 & 1.000 & 3.200 & 0.000 & -2.000 & 0.200 \\ 1.400 & 0.000 & -1.333 & 0.000 & 0.000 & 3.867 & 0.200 & 0.000 \\ 0.400 & 1.000 & 0.000 & -2.000 & -2.000 & 0.200 & 5.200 & 0.000 \\ 0.400 & 1.000 & -2.000 & -2.000 & 0.200 & 0.000 & 0.000 & 5.200 \end{bmatrix}^{-1} \begin{bmatrix} 11.300 \\ 6.800 \\ 0.000 \\ 0.000 \\ 2.000 \\ 5.900 \\ 4.300 \\ 5.900 \end{bmatrix}$$

and these give the same solutions as those obtained from equation [3.24].

3.5 Animal Model with Groups

In Example 3.1, there were animals in the pedigree with unknown parents, usually termed base population animals. The use of the relationship matrix in animal model evaluation assumes that these animals were sampled from a single population with average breeding value of zero and common variance σ_a^2. The breeding values of animals in subsequent generations are usually expressed relative to those of the base animals. However, if it is known that base animals were actually from populations that differ in genetic means, for instance, sires from different countries, this must then be accounted for in the model. In the dairy cattle situation, due to differences in selection intensity, the genetic means for sires of bulls, sires of cows, dams of bulls and dams of cows may all be different from each other. These various subpopulation structures should be accounted for in the model to avoid bias in the prediction of breeding values. This can be achieved through a proper grouping of base animals using available information.

Westell and Van Vleck (1987) presented a procedure for grouping which has generally been adopted. For instance, if sires have been imported from several countries over a period of time and their ancestors are unknown, these sires could be assigned to groups on the basis of the expected year of birth of the ancestors and the country of origin. The sires born within a similar time period in a particular foreign country are assumed to come from ancestors of similar genetic merit. Thus each sire with one or both parents unknown is initially assigned phantom parents. Phantom parents are assumed to have had only one progeny each. Within each of the foreign countries, the phantom parents are grouped by the year of birth of their progeny and any other factor, such as sex of progeny. In addition for the dairy cattle situation, the four selection paths – sire of sires, sire of dams, dam of sires and dam of dams – are usually assumed to be of different genetic merit and this is accounted for in the grouping strategy.

With groups, the model (Thompson, 1979) is:

$$y_{ij} = h_j + a_i + \sum_{k=1}^{n} t_{ik} g_k + e_{ij} \qquad [3.27]$$

where h_j = effect of the jth herd, a_i = random effect of animal i, g_k = fixed group effect containing the kth ancestor, t_{ik} = the additive genetic relationship between the kth and ith animals and the summation is over all n ancestors of animal i, and e_{ij} = random environmental effect. From the model, it can be seen that the contribution of the group to the observation is weighted by the proportion of genes the ancestors in the group passed on to the animal with a record.

In matrix notation the model can be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{ZQg} + \mathbf{Za} + \mathbf{e} \qquad [3.28]$$

where:

$$\mathbf{Q} = \mathbf{TQ}^*$$

$\mathbf{Q}^*$ assigns unidentified ancestors to groups and $\mathbf{T}$, a lower triangular matrix, is obtained from $\mathbf{A} = \mathbf{TDT}'$ (see Chapter 2, Section 2.2). With this model the breeding value of an animal $k(\hat{a}_{k^*})$ is obtained as:

$$\hat{a}_{k^*} = Q\hat{g} + \hat{a}_k$$

The MMEs are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{ZQ} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha & \mathbf{Z}'\mathbf{ZQ} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{X} & \mathbf{Q}'\mathbf{Z}'\mathbf{Z} & \mathbf{Q}'\mathbf{Z}'\mathbf{ZQ} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{Q}'\mathbf{Z}'\mathbf{y} \end{bmatrix}$$

Solving the MME above will yield vectors of solutions for **a** and **g**, but the ranking criterion (breeding value) is $\hat{a}_{k^*} = \mathbf{Qg} + \hat{a}_k$ for animal k. Modification of the mixed model equations (Quaas and Pollak, 1981) and absorption of the group equations gave the following set of equations, which are usually solved to obtain $\hat{\mathbf{a}}*$ directly (Westell *et al.*, 1988):

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}_{nn}^{-1}\alpha & \mathbf{A}_{np}^{-1}\alpha \\ \mathbf{0} & \mathbf{A}_{pn}^{-1}\alpha & \mathbf{A}_{pp}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} + \hat{\mathbf{Q}}\mathbf{g} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{0} \end{bmatrix} \qquad [3.29]$$

where n is the number of animals and p the number of groups.

Let

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{nn}^{-1} & \mathbf{A}_{np}^{-1} \\ \mathbf{A}_{pn}^{-1} & \mathbf{A}_{pp}^{-1} \end{bmatrix}$$

The matrix $\mathbf{A}^{-1}$ is obtained by the usual rules for obtaining the inverse of the relationship matrix outlined in Chapter 2, Section 2.3.1. A list of pedigrees, consisting of only actual animals but with unknown ancestors assigned to groups, is set up. For the ith animal calculate the inverse (b_i) of the variance of Mendelian sampling as:

$b_i = 4/(2 + \text{number of parents of animal } i \text{ assigned to groups})$

Then add:

b_i to the (i,i) element of $\mathbf{A}^{-1}$

$-\frac{b_i}{2}$ to the (i,s), (i,d), (s,i) and (d,i) elements of $\mathbf{A}^{-1}$

$\frac{b_i}{4}$ to the (s,s), (s,d), (d,s) and (d,d) elements of $\mathbf{A}^{-1}$

Thus, for an animal i with both parents assigned to groups:

$b_i = 4/(2 + 2) = 1$

Then add:

1 to the (i,i) element of $\mathbf{A}^{-1}$

$-\frac{1}{2}$ to the (i,s), (i,d), (s,i) and (d,i) elements of $\mathbf{A}^{-1}$

$\frac{1}{4}$ to the (s,s), (s,d), (d,s) and (d,d) elements of $\mathbf{A}^{-1}$

3.5.1 An illustration

Example 3.4
An animal model evaluation with groups is illustrated below using the same data set and genetic parameters as in Example 3.1. The aim is to estimate sex effects and predict breeding values for animals and phantom parents (groups). The model [3.28] and the MME [3.29] are used for the analysis. The pedigree file for the data set is as follows:

Calf	Sire	Dam
1	Unknown	Unknown
2	Unknown	Unknown
3	Unknown	Unknown
4	1	Unknown
5	3	2
6	1	2
7	4	5
8	3	6

Assuming that males are of different genetic merit compared to females, the unknown sires can be assigned to one group (G1) and unknown dams

to another group (G2). The pedigree file now becomes:

Calf	Sire	Dam
1	G1	G2
2	G1	G2
3	G1	G2
4	1	G2
5	3	2
6	1	2
7	4	5
8	3	6

Recoding G1 as 9 and G2 as 10:

Calf	Sire	Dam
1	9	10
2	9	10
3	9	10
4	1	10
5	3	2
6	1	2
7	4	5
8	3	6

Setting up design matrices and MME

The design matrices **X** and the matrices **X′X**, **X′Z**, **Z′X**, **X′y** and **Z′y** in the MME are exactly as in Example 3.1. The MME without addition of the inverse of the relationship matrix for animals and groups are:

$$\begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \\ \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} = \begin{bmatrix} 13.0 \\ 6.8 \\ 0.0 \\ 0.0 \\ 0.0 \\ 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \\ 0.0 \\ 0.0 \end{bmatrix}$$

Using the procedure outlined above, $\mathbf{A}^{-1}$ for the example data is:

	1	2	3	4	5	6	7	8	9	10
1	1.83	0.50	0.00	–0.67	0.00	–1.00	0.00	0.00	–0.50	–0.17
2	0.50	2.00	0.50	0.00	–1.00	–1.00	0.00	0.00	–0.50	–0.50
3	0.00	0.50	2.00	0.00	–1.00	0.50	0.00	–1.00	–0.50	–0.50
4	–0.67	0.00	0.00	1.83	0.50	0.00	–1.00	0.00	0.00	–0.67
5	0.00	–1.00	–1.00	0.50	2.50	0.00	–1.00	0.00	0.00	0.00
6	–1.00	–1.00	0.50	0.00	0.00	2.50	0.00	–1.00	0.00	0.00
7	0.00	0.00	0.00	–1.00	–1.00	0.00	2.00	0.00	0.00	0.00
8	0.00	0.00	–1.00	0.00	0.00	–1.00	0.00	2.00	0.00	0.00
9	–0.50	–0.50	–0.50	0.00	0.00	0.00	0.00	0.00	0.75	0.75
10	–0.17	–0.50	–0.50	–0.67	0.00	0.00	0.00	0.00	0.75	1.08

and $\mathbf{A}^{-1}\alpha$ is easily obtained by multiplying every element of $\mathbf{A}^{-1}$ by 2, the value of α. The matrix $\mathbf{A}^{-1}\alpha$ is added to equations for animal and group to obtain the MME, which are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \\ \hat{g}_1 \\ \hat{g}_2 \end{bmatrix} = \begin{bmatrix} 3.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 3.667 & 1.000 & 0.000 & -1.333 & 0.000 & -2.000 & 0.000 & 0.000 & -1.000 & -0.333 \\ 0.000 & 0.000 & 1.000 & 4.000 & 1.000 & 0.000 & -2.000 & -2.000 & 0.000 & 0.000 & -1.000 & -1.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 4.000 & 0.000 & -2.000 & 1.000 & 0.000 & -2.000 & -1.000 & -1.000 \\ 1.000 & 0.000 & -1.333 & 0.000 & 0.000 & 4.667 & 1.000 & 0.000 & -2.000 & 0.000 & 0.000 & -1.333 \\ 0.000 & 1.000 & 0.000 & -2.000 & -2.000 & 1.000 & 6.000 & 0.000 & -2.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & -2.000 & -2.000 & 1.000 & 0.000 & 0.000 & 6.000 & 0.000 & -2.000 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & -2.000 & -2.000 & 0.000 & 5.000 & 0.000 & 0.000 & 0.000 \\ 1.000 & 0.000 & 0.000 & 0.000 & -2.000 & 0.000 & 0.000 & -2.000 & 0.000 & 5.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & -1.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.500 & 1.500 \\ 0.000 & 0.000 & -0.333 & -1.000 & -1.000 & -1.333 & 0.000 & 0.000 & 0.000 & 0.000 & 1.500 & 2.167 \end{bmatrix}^{-1} \begin{bmatrix} 13.0 \\ 6.8 \\ 0.00 \\ 0.00 \\ 0.00 \\ 4.50 \\ 2.90 \\ 3.90 \\ 3.50 \\ 5.00 \\ 0.00 \\ 0.00 \end{bmatrix}$$

There is dependency in the equations, that is, all effects cannot be estimated; therefore, the equation for the first group has been set to zero to obtain the following solutions:

Effects	Solutions
Sex	
1	5.474
2	4.327
Animal	
1	0.780
2	–0.936
3	–0.977
4	–1.287
5	–1.113
6	–0.741
7	–1.354
8	–0.782
Group	
9	0.000
10	–1.795

The animal proofs above are generally lower than those from Example 3.1, the model without groups. In addition, the ranking for animals is also different. However, the relationship between the two sets of solutions can be shown by recalculating the vector of solutions for animals using the group solutions ($\hat{\mathbf{g}}$) above and the estimated breeding values ($\hat{\mathbf{a}}$) from Example 3.1 as:

$$\hat{\mathbf{a}}_* = \hat{\mathbf{a}} + \mathbf{Qg}$$

where $\mathbf{Q} = \mathbf{TQ}^*$, as defined earlier.

Assigning phantom parents (M1 to M7) to animals with unknown ancestors, the pedigree for the example data can be written as:

Calf	Sire	Dam
1	M1	M2
2	M3	M4
3	M5	M6
4	1	M7
5	3	2
6	1	2
7	4	5
8	3	6

and the matrix **T** for the pedigree is:

	M1	M2	M3	M4	M5	M6	M7	1	2	3	4	5	6	7	8
M1	1.000	0.000	0.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M2	0.000	1.000	0.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M3	0.000	0.000	1.000	0.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M4	0.000	0.000	0.000	1.000	0.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M5	0.000	0.000	0.000	0.000	1.000	0.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M6	0.000	0.000	0.000	0.000	0.000	1.000	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
M7	0.000	0.000	0.000	0.000	0.000	0.000	1.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
1	0.500	0.500	0.000	0.000	0.000	0.000	0.00	1.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0
2	0.000	0.000	0.500	0.500	0.000	0.000	0.00	0.00	1.00	0.00	0.0	0.0	0.0	0.0	0.0
3	0.000	0.000	0.000	0.000	0.500	0.500	0.00	0.00	0.00	1.00	0.0	0.0	0.0	0.0	0.0
4	0.250	0.250	0.000	0.000	0.000	0.000	0.50	0.50	0.00	0.00	1.0	0.0	0.0	0.0	0.0
5	0.000	0.000	0.250	0.250	0.250	0.250	0.00	0.00	0.50	0.50	0.0	1.0	0.0	0.0	0.0
6	0.250	0.250	0.250	0.250	0.000	0.000	0.00	0.50	0.50	0.00	0.0	0.0	1.0	0.0	0.0
7	0.125	0.125	0.125	0.125	0.125	0.125	0.25	0.25	0.25	0.25	0.5	0.5	0.0	1.0	0.0
8	0.125	0.125	0.125	0.125	0.250	0.250	0.00	0.25	0.25	0.50	0.0	0.0	0.5	0.0	1.0

The matrices $\mathbf{Q}^*$, which assigns phantom parents to groups, and $\mathbf{Q}$ are:

$$\mathbf{Q}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{TQ}^* = \mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.25 & 0.75 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.375 & 0.625 \\ 0.5 & 0.5 \end{bmatrix}$$

Therefore the vector of solutions using the estimated breeding values from Example 3.1 is:

$$\hat{\mathbf{a}}_* = \hat{\mathbf{a}} + \mathbf{Q}\hat{\mathbf{g}} = \begin{bmatrix} 0.098 \\ -0.019 \\ -0.041 \\ -0.009 \\ -0.186 \\ 0.177 \\ -0.249 \\ 0.183 \end{bmatrix} + \begin{bmatrix} -0.898 \\ -0.898 \\ -0.898 \\ -1.346 \\ -0.898 \\ -0.898 \\ -1.122 \\ -0.898 \end{bmatrix} = \begin{bmatrix} -0.800 \\ -0.917 \\ -0.939 \\ -1.355 \\ -1.084 \\ -0.721 \\ -1.371 \\ -0.715 \end{bmatrix}$$

These solutions are similar to those obtained in the model with groups. The slight differences are due to differences in sex solutions in the two examples and this is explained later. This indicates that, when the solutions from the model without groups are expressed relative to the group solutions, similar solutions are obtained to those in the model with groups. Thus, the differences between the solutions in Examples 3.1 and 3.4 are due to the fact that the solutions in the former are expressed relative to base animals assumed to have an average breeding value of zero, while in the latter solutions are relative to the group solutions, one of which is lower than zero.

The inclusion of groups also resulted in a larger sex difference compared with Example 3.1. The solution for sex effect *i* can be calculated using

equation [3.5]. For instance, the solution for male calves in Example 3.4 is:

$$\hat{b}_1 = [(4.5 + 3.5 + 5.0) - (-1.287 + -1.354 + -0.782)]/3 = 5.474$$

Since $\sum_j y_{ij}$ in equation [3.5] is the same in both examples, differences in $\sum_j \hat{a}_{ij}$ between the sexes in both models would result in differences in the linear function of **b**. The difference between average breeding values of male and females calves is –0.02 and –0.214, respectively, in Examples 3.1 and 3.4. The larger difference in the latter accounted for the higher sex difference in Example 3.4. Males had a lower breeding value in Example 3.4 due to the higher proportionate contribution of group two to their solutions (see the matrix **Q** above).

The basic principles involved in the application of BLUP for genetic evaluations and the main assumptions have been covered in this chapter, and its application to more complex models involves an extension of these principles. Equation [3.1] is a very general model and **a** could include random animal effects for several traits (multivariate model), random environmental effects, such as common environmental effects affecting animals that are reared together, maternal effects (maternal model), non-additive genetic effects, such as dominance and epistasis (non-additive models) and repeated data on individuals (random regression model). The extension of the principles discussed in this chapter under these various models constitutes the main subject area of the subsequent chapters in the text.

4 Best Linear Unbiased Prediction of Breeding Value: Models with Random Environmental Effects

In some circumstances, environmental factors constitute an important component of the covariance between individuals such as members of a family reared together (common environmental effects) or between the records of an individual (permanent environmental effects). Such environmental effects are usually accounted for in the model to ensure accurate prediction of breeding. This chapter deals with models that take into account these two main types of environmental effects in genetic evaluations.

4.1 Repeatability Model

The repeatability model has been employed for the analysis of data when multiple measurements on the same trait are recorded on an individual, such as litter size in successive pregnancies or milk yield in successive lactations (Interbull, 2000). The details of the assumptions and the components of the phenotypic variance have been given in Section 1.2.2. Briefly, the phenotypic variance comprises the genetic (additive and non-additive) variance, permanent environmental variance and temporary environmental variance. For an animal, the repeatability model usually assumes a genetic correlation of unity between all pairs of records, equal variance for all records and equal environmental correlation between all pairs of records. In practice, some of these assumptions do not hold in the analysis of real data. A more appropriate way of handling repeated measurements over time is by fitting a random regression model or a covariance function, and this is discussed in Chapter 8. This section has therefore been included to help illustrate the evolution of the model for the analysis of repeated records over time. The phenotypic structure

for three observations of an individual under this model could be written (Quaas, 1984) as:

$$\text{var}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \sigma^2_{t1} + \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g \\ \sigma^2_{pe} + \sigma^2_g & \sigma^2_{t2} + \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g \\ \sigma^2_{pe} + \sigma^2_g & \sigma^2_{pe} + \sigma^2_g & \sigma^2_{t3} + \sigma^2_{pe} + \sigma^2_g \end{bmatrix}$$

with: σ^2_{ti} = temporary environmental variance specific to record *i*; σ^2_{pe} = covariance due to permanent environmental effects (variances and covariances are equal); and σ^2_g = genetic covariance (variances and covariances are equal). The correlation between records of an individual referred to as repeatability is $(\sigma^2_g + \sigma^2_{pe})/\sigma^2_y$. Genetic evaluation under this model is concerned not only with predicting breeding values but also permanent environmental effects.

4.1.1 Defining the model

The repeatability model is usually of the form

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e} \qquad [4.1]$$

where **y** = vector of observations, **b** = vector of fixed effects, **a** = vector of random animal effects, **pe** = vector of random permanent environmental effects and non-additive genetic effects, **e** = vector of random residual effect, and **X**, **Z** and **W** are incidence matrices relating records to fixed, animal and permanent environmental effects, respectively.

Note that the vector **a** only includes additive random animal effects; consequently, non-additive genetic effects are included in the **pe** term. It is assumed that the permanent environmental effects and residual effects are independently distributed with means of zero and variance σ^2_{pe} and σ^2_e, respectively; therefore:

$$\text{var}(\mathbf{pe}) = \mathbf{I}\sigma^2_{pe}$$
$$\text{var}(\mathbf{e}) = \mathbf{I}\sigma^2_e = \mathbf{R}$$
$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma^2_a$$

and:

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma^2_a + \mathbf{WI}\sigma^2_{pe}\mathbf{W}' + \mathbf{R}$$

The mixed model equations for the best linear unbiased estimator (BLUE) of estimable functions of **b** and for the best linear unbiased prediction (BLUP) of **a** and **pe** are:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1}1/\sigma^2_a & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{I}(1/\sigma^2_{pe}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

However, the mixed model equations (MME) with $\mathbf{R}^{-1}$ factored out from the above equations give the following equations, which are easier to set up:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \mathbf{I}\alpha_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix} \quad [4.2]$$

where:

$$\alpha_1 = \sigma_e^2/\sigma_a^2 \quad \text{and} \quad \alpha_2 = \sigma_e^2/\sigma_{pe}^2 .$$

4.1.2 An illustration

Example 4.1
For illustrative purposes, assume a single dairy herd with the following data structure for six cows:

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season.

It is assumed that $\sigma_a^2 = 20.0$, $\sigma_e^2 = 28.0$ and $\sigma_{pe}^2 = 12.0$, giving a phenotypic variance (σ_y^2) of 60. From the given parameters, $\alpha_1 = 1.40$, $\alpha_2 = 2.333$ and repeatability is $(\sigma_a^2 + \sigma_{pe}^2)/\sigma_y^2 = (20 + 12)/60 = 0.53$. The aim is to estimate the effects of lactation number and predict breeding values for all animals and permanent environmental effects for cows with records. The above genetic parameters are proportional to estimates reported by Visscher (1991) for fat yield for Holstein Friesians in the UK for the first two lactations using a repeatability model. Later, in Section 5.5, this data set is reanalysed using a multivariate model assuming an unequal design with different herd–year–season (HYS) effects defined for each lactation, using the corresponding multivariate genetic parameter estimates of Visscher (1991).

SETTING UP THE DESIGN MATRICES

The transpose of the matrix **X**, which relates records to HYS and parity, is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

The first four rows of **X′** relate records to HYS effects and the last two rows to parity effects.

Considering only animals with records, **Z′** and **W′** are equal and, for the example data set:

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Each row of **Z′** corresponds to each cow with records. The matrices **Z′Z** and **W′W** are both diagonal and equal and **Z′Z** is:

$$\mathbf{Z}'\mathbf{Z} = \text{diag}(2, 2, 2, 2, 2)$$

Note, however, that it is necessary to augment **Z′Z** by three columns and rows of zeros to account for animals 1 to 3, which are ancestors. The remaining matrices in the MME apart from $\mathbf{A}^{-1}$ can easily be calculated through matrix multiplication. The inverse of the relationship matrix ($\mathbf{A}^{-1}$) is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 2.50 & 0.50 & 0.00 & -1.00 & 0.50 & -1.00 & 0.50 & -1.00 \\ 0.50 & 2.00 & 0.50 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.50 & 2.00 & 0.50 & -1.00 & 0.00 & -1.00 & 0.00 \\ -1.00 & -1.00 & 0.50 & 2.50 & 0.00 & 0.00 & -1.00 & 0.00 \\ 0.50 & -1.00 & -1.00 & 0.00 & 2.50 & -1.00 & 0.00 & 0.00 \\ -1.00 & 0.00 & 0.00 & 0.00 & -1.00 & 2.00 & 0.00 & 0.00 \\ 0.50 & 0.00 & -1.00 & -1.00 & 0.00 & 0.00 & 2.50 & -1.00 \\ -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & 2.00 \end{bmatrix}$$

and $\mathbf{A}^{-1}\alpha_1$ is added to the **Z′Z** to obtain the mixed model equations.

The MME are too large to be shown. There is dependency in the MME because the sum of equations for HYS 1 and 2 equals that of parity 1 and

the sum of HYS 3 and 4 equals that for parity 2. The equations for HYS 1 and 3 were set to zero to obtain the following solutions by direct inversion of the coefficient matrix:

Effects	Solutions
HYS	
1	0.000
2	44.065
3	0.000
4	0.013
Parity	
1	175.472
2	241.893
Animal	
1	10.148
2	–3.084
3	–7.063
4	13.581
5	–18.207
6	–18.387
7	9.328
8	24.194
Permanent environment	
4	8.417
5	–7.146
6	–17.229
7	–1.390
8	17.347

The fixed-effect solutions for parity indicate that yield at second lactation is higher than that at first, which is consistent with the raw averages. From the MME, the solution for level i of the nth fixed effect can be calculated as:

$$\hat{b}_{in} = \sum_{f=1}^{\text{diag}_{in}} y_{inf} - \sum_{j} \hat{b}_{inj} - \sum_{k} \hat{a}_{ink} - \sum_{l} \hat{p}e_{inl} / \text{diag}_{in} \qquad [4.3]$$

where y_{inf} is the record for animal f in level i of the nth fixed effect, diag_{in} is the number of observations for level i of the nth fixed effect, b_{inj}, $\hat{a}_{ink}$ and pe_{inl} are solutions for levels j, k and l of any other fixed effect, random animal and permanent environmental effects, respectively, within level i of the nth fixed effect. Thus the solution for level two of HYS effect is:

$$\begin{aligned}\hat{b}_{21} &= [445 - (2\hat{b}_{12}) - (\hat{a}_6 + \hat{a}_8) - (\hat{p}e_6 + \hat{p}e_8)]/2 \\ &= [445 - 2(175.472) - 5.807 - (0.118)]/2 \\ &= 44.065\end{aligned}$$

Breeding values for animals with a repeatability model can also be calculated using equation [3.8], except that yield deviation (YD) is now yield corrected for the appropriate fixed effects, permanent environmental effect and averaged. Thus, for animal 4:

$$\hat{a}_4 = n_1[(\hat{a}_1 + \hat{a}_2)/2] + n_2[((y_{41} - \hat{b}_1 - \hat{b}_5 - pe_4) + (y_{42} - \hat{b}_3 - \hat{b}_5 - \hat{p}e_4))/2] + n_3(2\hat{a}_7 - \hat{a}_3)$$

where y_{ji} is yield for cow j in lactation i, $n_1 = 2.8/5.5$, $n_2 = 2/5.5$ and $n_3 = 0.7/5.5$, and 5.5 = the sum of the numerator of n_1, n_2 and n_3.

$$\hat{a}_4 = n_1(3.532) + n_2[((201 - 0.0 - 175.472 - 8.417) + (280 - 0.0 - 241.893))/2] + n_3(18.656 - (-7.063)) = 13.581$$

The higher breeding value for sire 1 compared with sire 3 is due to the fact that on average the daughters of sire 1 were of higher genetic merit after adjusting for the breeding values of mates. The very high breeding value for cow 8 results from the high parent average breeding value and she has the highest yield in the herd, resulting in a large YD.

The estimate of *pe* for animal *i* could be calculated as:

$$\hat{p}e_i = \left[\sum_{f}^{mi} y_{if} - \sum_{j} \hat{b}_{ij} - \sum_{k} \hat{a}_{ik}\right] \Big/ (m_i + \alpha_2) \qquad [4.4]$$

where m_i is the number of records for animal i and other terms are as defined in [4.3]. Thus, for animal 4:

$$\hat{p}e_4 = [(201 - 0.0 - 175.472 - 13.581) + (280 - 0.0 - 241.893 - 13.581)]/(2 + 2.333) = 8.417$$

The estimate of permanent environment effect for an animal represents environmental influences and non-additive genetic effect, which are peculiar to the animal and affect its performance for life. These environmental influences could either be favourable – for instance, animal 8 has the highest estimates of **pe** and this is reflected by her high average yield – or could reduce performance (for example, cow 6 has a very negative estimate of **pe** and low average yield). A practical example of such permanent environment effect could be the loss of a teat by a cow early in life due to infection. Thus differences in estimates of **pe** represent permanent environmental differences between animals and could assist the farmer in addition to the breeding value in selecting animals for future performance in the same herd. The sum of breeding value and permanent environment effect ($\hat{a}_i + \hat{p}e_i$) for animal i is termed the probable producing ability (PPA) and represents an estimate of the future

performance of the animal in the same herd. If the estimate of the management level (M) for animal i is known, its future record (y_i) can be predicted as:

$$y_i = M + \text{PPA}$$

This could be used as a culling guide.

4.1.3 Calculating daughter yield deviations

As indicated in Section 3.2.2, daughter yield deviation (DYD) is commonly calculated for sires in dairy cattle evaluations. The calculation of DYD for sire 1 in Example 4.1 is hereby illustrated.

First, the yield deviations for the daughters (cows 4, 6 and 8) of sire 1 are calculated. Thus, for cow i, $\text{YD}_i = (\mathbf{Z'Z})^{-1}\,\mathbf{Z'}(\mathbf{y}_i - \mathbf{X\hat{b}} - \mathbf{W\hat{p}e})$. Therefore:

$$\text{YD}_4 = \tfrac{1}{2}[(201 - 175.472 - 0 - 8.417) + (280 - 241.893 - 0 - 8417)] = 23.4005$$

$$\text{YD}_6 = \tfrac{1}{2}[(160 - 175.472 - 44.065 - (-17.229)) + (190 - 241.893 - 0 - (-17.229)] = -38.486$$

$$\text{YD}_8 = \tfrac{1}{2}[(285 - 175.472 - 44.065 - 17.347) + (300 - 241.893 - 0.013 - 17.347)] = 44.432$$

Both parents of these daughters are known; therefore, $n_{2prog} = 2/(2 + 2\alpha_1) = 0.4167$ and $u_{prog} = 1$ for each daughter. Using equation [3.12], DYD for sire 1 is:

$$\text{DYD}_1 = \left[(u_{(4)}n_{2(4)}(2\text{YD}_4 - \hat{a}_2) + u_{(5)}n_{2(6)}(2\text{YD}_6 - \hat{a}_5) + u_{(8)}n_{2(8)}(2\text{YD}_8 - \hat{a}_7)\right] \Big/ \left(\sum_3 (u_{prog} + n_{2prog})\right)$$

$$\text{DYD}_1 = [(1)(0.4167)((2(23.4005)) - (-3.084)) + (1)0.4167((2(-38.486)) - (-18.207)) + (1)0.4167((2(44.432)) - 9.328)]/(3(1)(0.4167)) = 23.552$$

Calculating the proof of sire 1 using equation [3.13] and a DYD of 23.552 gives a breeding value of 9.058. It is slightly lower than the breeding value of 10.148 from solving the MME, as the contribution of the granddaughter through cow 4 is not included.

4.2 Model with Common Environmental Effects

Apart from the resemblance between records of an individual due to permanent environmental conditions, discussed in Section 4.1, environmental

circumstances can also contribute to the resemblance between relatives. When members of a family are reared together, such as litters of pigs, they share a common environment and this contributes to the similarity between members of the family. Thus there is an additional covariance between members of a family due to the common environment they share and this increases the variance between different families. The environmental variance may be partitioned therefore into the between-family or between-group component (σ_c^2), usually termed the common environment, which causes resemblance between members of a family, and the within-family or within-group variance (σ_e^2). Sources of common environmental variance between families may be due to factors such as nutrition and/or climatic conditions. All sorts of relatives are subject to an environmental source of resemblance, but most analyses concerned with this type of variation in animal breeding tend to account for the common environment effects associated with full-sibs or maternal half-sibs, especially in pig and chicken studies.

4.2.1 Defining the model

Genetic evaluation under this model is concerned with prediction of breeding values and common environmental effects and the phenotypic variance may be partitioned into:

1. Additive genetic effects resulting from additive genes from parents.
2. Common environmental effects affecting full-sibs or all offspring of the same dam. In the case of full-sibs, it may be confounded with dominance effects peculiar to offspring of the same parents. Further explanation is given later on the components of the common environmental effect.
3. Random environmental effects.

In matrix notation, the model, which is similar to [4.1], is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wc} + \mathbf{e} \qquad [4.5]$$

where all terms are as given in [4.1] except **c**, which is the vector of common environmental effects, and **W** now relates records to common environmental effects.

It is assumed that common environmental and residual effects are independently distributed with means of zero and variance σ_c^2 and σ_e^2, respectively. Thus var(**c**) = $\mathbf{I}\sigma_c^2$, var(**e**) = $\mathbf{I}\sigma_e^2$ and var(**a**) = $\mathbf{A}\sigma_a^2$.

The mixed model equations for the BLUP of **a** and **c** and BLUE of estimable functions of **b** are exactly the same as [4.2] but with $\alpha_1 = \sigma_e^2/\sigma_a^2$ and $\alpha_2 = \sigma_e^2/\sigma_c^2$.

4.2.2 An illustration

Example 4.2
Consider the following data set on the weaning weight of piglets which are progeny of three sows mated to two boars:

Piglet	Sire	Dam	Sex*	Weaning weight (kg)
6	1	2	1	90
7	1	2	2	70
8	1	2	2	65
9	3	4	2	98
10	3	4	1	106
11	3	4	2	60
12	3	4	2	80
13	1	5	1	100
14	1	5	2	85
15	1	5	1	68

*1 = male, 2 = female (throughout chapter)

The objective is to predict breeding values for all animals and common environmental effects for full-sibs. Given that $\sigma_a^2 = 20$, $\sigma_c^2 = 15$ and $\sigma_e^2 = 65$, then $\sigma_y^2 = 100$, $\alpha_1 = 3.25$ and $\alpha_2 = 4.333$.

The model for the analysis is that presented in [4.5] and, as mentioned earlier, the mixed model equations for the BLUP of **a** and **c** and BLUE of estimable functions of **b** are as given in [4.2], using α_1 and α_2 as defined above.

Setting up the design matrices

The transpose of the matrix **X**, which relates records to sex effects in this example, is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

and $\mathbf{Z} = \mathbf{I}$, excluding parents. The transpose of matrix **W**, which relates records to full-sibs, is:

$$\mathbf{W}' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

The MME can be set up as discussed in Example 4.1. The solutions to the MMEs by direct inversion of the coefficient matrix are:

Effects	Solutions
Sex	
1	91.493
2	75.764
Animal	
1	–1.441
2	–1.175
3	1.441
4	1.441
5	–0.266
6	–1.098
7	–1.667
8	–2.334
9	3.925
10	2.895
11	–1.141
12	1.525
13	0.448
14	0.545
15	–3.819
Common environment	
2	–1.762
4	2.161
5	–0.399

The equations for the solution of the *i* level of fixed, animal and common environmental effects under this model are the same as those given for fixed [4.3], animal and permanent environmental effects [4.4], respectively, in Example 4.1. The inclusion of common environmental effects in the model allows for accurate prediction of breeding values of animals. Assuming each dam reared her progeny and full-sib families were kept under similar environmental conditions, the estimates of common environmental effects indicate that dam 4 provided the best environment for her progeny compared with dams 2 and 5. Also, dam 4 has the highest breeding value among the dams and would therefore be the dam of first choice, whether selection is for dams of the next generation on the basis of breeding value only or selection is for future performance of the dams in the same herd, which will be based on some combination of breeding value and estimate of common environmental effect.

The environmental covariance among full-sibs or maternal half-sibs might be due to influences from the dam (mothering ability or maternal effect); therefore, differences in mothering ability among dams would

cause environmental variance between families. For instance, resemblance among progeny of the same dam in body weight could be due to the fact that they share the same milk supply, and variation in milk yield among dams would result in differences between families in body weight. This variation in mothering ability of dams has a genetic basis and, to some degree, is due to genetic variation in some character of the dams. In Chapter 6, the genetic component of maternal effect is examined and the appropriate model that accounts for the genetic component in genetic evaluation is presented.

5 Best Linear Unbiased Prediction of Breeding Value: Multivariate Models

Selection of livestock is usually based on a combination of several traits of economic importance that may be phenotypically and genetically related. Such traits may be combined into an index on which animals are ranked. A multiple trait evaluation is the optimum methodology to evaluate animals on these traits since it accounts for the relationship between them. A multiple trait analysis involves the simultaneous evaluation of animals for two or more traits and makes use of the phenotypic and genetic correlations between the traits. The first application of best linear unbiased prediction (BLUP) for multiple trait evaluation was by Henderson and Quaas (1976).

One of the main advantages of multivariate BLUP (MBLUP) is that it increases the accuracy of evaluations. The gain in accuracy is dependent on the absolute difference between the genetic and residual correlations between the traits. The larger the differences in these correlations, the greater the gain in accuracy of evaluations (Schaeffer, 1984; Thompson and Meyer, 1986). When, for instance, the heritability, genetic and environmental correlations for two traits are equal, multivariate predictions are equivalent essentially to those from univariate analysis for each trait. Moreover, traits with lower heritabilities benefit more when analysed with traits with higher heritabilities in a multivariate analysis. Also, there is an additional increase in accuracy with multivariate analysis resulting from better connections in the data due to residual covariance between traits (Thompson and Meyer, 1986).

Additionally, in some cases one trait is used to decide whether animals should remain in the herd and be recorded for other traits. For instance, only calves with good weaning weight may be allowed the chance to be measured for yearling weight. A single trait analysis of yearling weight will be biased since it does not include information on the weaning weight on which the selection was based. This is often called culling bias.

However, a multi-trait analysis on weaning and yearling weight can eliminate this bias. Thus MBLUP accounts for culling selection bias.

One of the disadvantages of a multiple trait analysis is the high computing cost. The cost of multiple analysis of *n* traits is more than the cost of *n* single analyses. Secondly, a multiple trait analysis requires reliable estimates of genetic and phenotypic correlations among traits and these may not be readily available.

In this chapter MBLUP involving traits affected by the same effects (equal design matrices) and situations in which different traits are affected by different factors (non-identical design matrices) are discussed. Approximations of MBLUP when design matrices are equal with or without missing records are also examined.

5.1 Equal Design Matrices and No Missing Records

Equal design matrices for all traits imply that all effects in the model affect all traits in the multivariate analysis and there are no missing records for any trait.

5.1.1 Defining the model

The model for a multivariate analysis resembles a stack of the univariate models for each of the traits. For instance, consider a multivariate analysis for two traits, with the model for each trait of the form given in [3.1], that is for trait 1:

$$\mathbf{y}_1 = \mathbf{X}_1\mathbf{b}_1 + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{e}_1$$

and for trait 2:

$$\mathbf{y}_2 + \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}_2\mathbf{a}_2 + \mathbf{e}_2$$

If animals are ordered within traits, the model for the multivariate analysis for the two traits could be written as:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} \qquad [5.1]$$

where $\mathbf{y}_i$ = vector of observations for the *i*th trait, $\mathbf{b}_i$ = vector of fixed effects for the *i*th trait, $\mathbf{a}_i$ = vector of random animal effects for the *i*th trait, $\mathbf{e}_i$ = vector of random residual effects for the *i*th trait, and $\mathbf{X}_i$ and $\mathbf{Z}_i$ are incidence matrices relating records of the *i*th trait to fixed and random animal effects, respectively.

It is assumed that

$$\text{var}\begin{bmatrix} a_1 \\ a_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} & 0 & 0 \\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & 0 & 0 \\ 0 & 0 & r_{11} & r_{12} \\ 0 & 0 & r_{21} & r_{22} \end{bmatrix}$$

where **G** = additive genetic variance and covariance matrix for animal effect with each element defined as; g_{11} = additive genetic variance for direct effects for trait 1; $g_{12} = g_{21}$ = additive genetic covariance between both traits; g_{22} = additive genetic variance for direct effects for trait 2; **A** is the relationship matrix among animals; and, **R** = variance and covariance matrix for residual effects.

The mixed model equations (MME) are of the same form as in Section 3.1, and these are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}' \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad [5.2]$$

where:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix}, \quad \hat{\mathbf{b}} = \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \end{bmatrix}, \quad \hat{\mathbf{a}} = \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

Writing out the equations for each trait in the model separately, the MME becomes:

$$\begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{11}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{Z}_2 \\ \mathbf{X}_2'\mathbf{R}^{12}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{21}\mathbf{Z}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1 + \mathbf{A}^{-1}\mathbf{g}^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{A}^{-1}\mathbf{g}^{12} \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1 + \mathbf{A}^{-1}\mathbf{g}^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2 + \mathbf{A}^{-1}\mathbf{g}^{22} \end{bmatrix}^{-1}$$
$$\begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{22}\mathbf{y}_2 \\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{y}_2 \end{bmatrix} \qquad [5.3]$$

where g^{ij} are elements of $\mathbf{G}^{-1}$. It should be noted that if $\mathbf{R}^{12}$, $\mathbf{R}^{21}$, g^{12} and g^{21} were set to zero, the matrices in the equations above reduce to the usual ones computed when carrying out two single trait analyses since the two traits become uncorrelated and there is no flow of information from one trait to the other.

5.1.2 An illustration

Example 5.1

Assume that the data in Table 5.1 are the pre-weaning gain (WWG) and post-weaning gain (PWG) for five beef calves. The objective is to estimate sex effects for both traits and to predict breeding values for all animals

Table 5.1. Pre-weaning gain (kg) and post-weaning gain (kg) for five beef calves.

Calf	Sex	Sire	Dam	WWG	PWG
4	Male	1	–	4.5	6.8
5	Female	3	2	2.9	5.0
6	Female	1	2	3.9	6.8
7	Male	4	5	3.5	6.0
8	Male	3	6	5.0	7.5

WWG, pre-weaning gain; PWG, post-weaning gain.

using a MBLUP analysis. Assume that the additive genetic covariance (**G**) matrix is:

$$\begin{matrix} \text{WWG} \\ \text{PWG} \end{matrix} \begin{bmatrix} 20 & 18 \\ 18 & 40 \end{bmatrix}$$

and the residual covariance matrix (**R**) is:

$$\begin{matrix} \text{WWG} \\ \text{PWG} \end{matrix} \begin{bmatrix} 40 & 11 \\ 11 & 30 \end{bmatrix}$$

The inverses of **G** and **R** are:

$$\mathbf{G}^{-1} = \begin{bmatrix} 0.084 & -0.038 \\ -0.038 & 0.042 \end{bmatrix} \quad \text{and} \quad \mathbf{R}^{-1} = \begin{bmatrix} 0.028 & -0.010 \\ -0.010 & 0.037 \end{bmatrix}$$

Setting up the design matrices

The matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ relate records for WWG and PWG, respectively, to sex effects. Both matrices are exactly the same as **X** in Section 3.2.1. Considering only animals with records, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ relate records for WWG and PWG to animals, respectively. Both matrices are identity matrices since animals have only one record each for WWG and PWG. The matrix **y** is a vector of observations for WWG ($\mathbf{y}_1$) and PWG ($\mathbf{y}_2$). Thus its transpose is:

$$\mathbf{y}' = [\mathbf{y}_1' \quad \mathbf{y}_2'] = [4.5 \quad 2.9 \quad 3.9 \quad 3.5 \quad 5.0 \quad 6.8 \quad 5.0 \quad 6.8 \quad 6.0 \quad 7.5]$$

The other matrices in the MME can then easily be calculated from the design matrices and vector of observations through matrix multiplication. Examples of some blocks of equations are given below.

From equations [5.2] and [5.3], the fixed effects by fixed effects block of equations for both traits in the coefficient matrix of the MME is:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} 0.084 & 0.000 & -0.03 & 0.00 \\ 0.00 & 0.056 & 0.00 & -0.02 \\ -0.03 & 0.00 & 0.101 & 0.00 \\ 0.00 & -0.02 & 0.00 & 0.074 \end{bmatrix}$$

The right-hand side for the levels of sex effects for both traits is:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} = \begin{bmatrix} \mathbf{X}'_1\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{X}'_2\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{X}'_2\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{X}'_2\mathbf{R}^{22}\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} 0.364 + (-0.203) \\ 0.190 + (-0.118) \\ -0.130 + 0.751 \\ -0.068 + 0.437 \end{bmatrix}$$

The inverse of the relationship matrix for the example data is the same as that given in Example 3.1. The matrices $\mathbf{A}^{-1}g^{11}$, $\mathbf{A}^{-1}g^{12}$, $\mathbf{A}^{-1}g^{21}$ and $\mathbf{A}^{-1}g^{22}$ are added to $\mathbf{Z}'_1\mathbf{R}^{11}\mathbf{Z}_1$, $\mathbf{Z}'_1\mathbf{R}^{12}\mathbf{Z}_2$, $\mathbf{Z}'_2\mathbf{R}^{21}\mathbf{Z}_1$ and $\mathbf{Z}'_2\mathbf{R}^{22}\mathbf{Z}_2$, respectively, to obtain the MME. For example, the matrix $\mathbf{Z}'_1\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{12}$ is:

$$\mathbf{Z}'_1\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{12} = \begin{bmatrix} -0.069 & -0.019 & 0.000 & 0.025 & 0.000 & 0.038 & 0.000 & 0.000 \\ -0.019 & -0.076 & -0.019 & 0.000 & 0.038 & 0.038 & 0.000 & 0.000 \\ 0.000 & -0.019 & -0.076 & 0.000 & 0.038 & -0.019 & 0.000 & 0.038 \\ 0.025 & 0.000 & 0.000 & -0.080 & -0.019 & 0.000 & 0.038 & 0.000 \\ 0.000 & 0.038 & 0.038 & -0.019 & -0.105 & 0.000 & 0.038 & 0.000 \\ 0.038 & 0.038 & -0.019 & 0.000 & 0.000 & -0.105 & 0.000 & 0.038 \\ 0.000 & 0.000 & 0.000 & 0.038 & 0.038 & 0.000 & -0.086 & 0.000 \\ 0.000 & 0.000 & 0.038 & 0.000 & 0.000 & 0.038 & 0.000 & -0.086 \end{bmatrix}$$

The MME have not been presented because they are too large, but solving the MME by direct inversion of the coefficient matrix gives the solutions shown below. See also the solutions from a univariate analysis of each trait.

	Multivariate analysis traits		Univariate analysis traits	
Effects	WWG	PWG	WWG	PWG
*Sex**				
1	4.361	6.800	4.358	6.798
2	3.397	5.880	3.404	5.879
Animal				
1	0.151	0.280	0.098	0.277
2	–0.015	–0.008	–0.019	–0.005
3	–0.078	–0.170	–0.041	–0.171
4	–0.010	–0.013	–0.009	–0.013
5	–0.270	–0.478	–0.186	–0.471
6	0.276	0.517	0.177	0.514
7	–0.316	–0.479	–0.249	–0.464
8	0.244	0.392	0.183	0.384

*1 = male, 2 = female (throughout chapter)

The differences between the solutions for males and females for WWG and PWG in the multivariate analysis are more or less the same as those

obtained in the univariate analyses of both traits. The solutions for fixed effects in the multivariate analysis from the MME can be calculated as:

$$\begin{bmatrix} \hat{b}_{1j} \\ \hat{b}_{2j} \end{bmatrix} = \begin{bmatrix} n_j r^{11} & n_j r^{12} \\ n_j r^{21} & n_j r^{22} \end{bmatrix}^{-1} \left[\mathbf{R}^{-1} \begin{bmatrix} y_{1j.} - a_{1j.} - g^{12} a_{2j.} \\ y_{2j.} - g^{21} a_{1j.} - a_{2j.} \end{bmatrix} \right] \quad [5.4]$$

where $y_{ij.}$ and $\hat{a}_{ij.}$ are the sums of observations and estimated breeding values, respectively, for calves for trait i in sex subclass j, $\hat{b}_{ij}$ is the solution for trait i in sex subclass j and n_j is the number of observations for sex subclass j. Using equation [5.4], the solutions for sex effects for males for WWG and PWG are:

$$\begin{bmatrix} \hat{b}_{11} \\ \hat{b}_{21} \end{bmatrix} = \begin{bmatrix} 3r^{11} & 3r^{12} \\ 3r^{21} & 3r^{22} \end{bmatrix}^{-1} \left[\begin{bmatrix} r^{11} & r^{12} \\ r^{21} & r^{22} \end{bmatrix} \begin{bmatrix} 13.0 - (0.082) - g^{12}(-0.10) \\ 20.3 - g^{21}(-0.082) - (-0.10) \end{bmatrix} \right]$$

$$= \begin{bmatrix} 4.361 \\ 6.800 \end{bmatrix}$$

5.1.3 Partitioning animal evaluations from multivariate analysis

An equation similar to [3.8] for the partitioning of evaluations from multivariate models was presented by Mrode and Swanson (2004) in the context of a random regression model (see Chapter 7). Since the yield records of animals contribute to the breeding values through the vector of yield deviations (**YD**), equations for calculating **YD** are initially presented. From equation [5.2], the equations for the breeding values of animals are:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1})\hat{\mathbf{a}} = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

Therefore:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1})\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\mathbf{YD} \quad [5.5]$$

with:

$$\mathbf{YD} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})) \quad [5.6]$$

Just as in the univariate model, **YD** is a vector of the weighted average of a cow's yield records corrected for all fixed effects in the model.

Transferring the left non-diagonal terms of $\mathbf{A}^{-1}$ in equation [5.5] to the right side of the equation (VanRaden and Wiggans, 1991) gives:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{anim})\hat{\mathbf{a}}_{anim} = \mathbf{G}^{-1}\alpha_{par}(\hat{\mathbf{a}}_{sire} + \hat{\mathbf{a}}_{dam}) + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\mathbf{YD} + \mathbf{G}^{-1}\sum \alpha_{prog}(\mathbf{a}_{prog} - 0.5\hat{\mathbf{a}}_{mate})$$

where $\alpha_{par} = 1, \frac{2}{3}$ or $\frac{1}{2}$ if both, one or neither parents are known, respectively, and $\alpha_{\text{prog}} = 1$ if the animal's mate is known and $\frac{2}{3}$ if unknown. Note that $\alpha_{anim} = 2\alpha_{par} + 0.5\alpha_{prog}$.

The above equation can be expressed as:

$$(\mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{anim})\hat{\mathbf{a}}_{anim} = 2\mathbf{G}^{-1}\alpha_{par}(\mathbf{PA}) + (\mathbf{Z'R}^{-1}\mathbf{Z})\mathbf{YD} + 0.5\mathbf{G}^{-1}\sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - \hat{\mathbf{a}}_{mate}) \quad [5.7]$$

where **PA** = parent average.

Pre-multiplying both sides of the equation by $(\mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{anim})^{-1}$ gives:

$$\hat{\mathbf{a}}_{anim} = \mathbf{W}_1\mathbf{PA} + \mathbf{W}_2\mathbf{YD} + \mathbf{W}_3\mathbf{PC} \quad [5.8]$$

with:

$$\mathbf{PC} = \sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - \hat{\mathbf{a}}_{mate}) \Big/ \sum \alpha_{prog}$$

The weights $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3 = \mathbf{I}$, with $\mathbf{W}_1 = (\mathbf{DIAG})^{-1}2\mathbf{G}^{-1}\alpha_{par}$, $\mathbf{W}_2 = (\mathbf{DIAG})^{-1}(\mathbf{Z'R}^{-1}\mathbf{Z})$, and $\mathbf{W}_3 = (\mathbf{DIAG})^{-1}0.5\mathbf{G}^{-1}\Sigma\alpha_{prog}$, where $(\mathbf{DIAG}) = (\mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{anim})$. Equation [5.8] is similar to [3.8] but the weights are matrices of the order of traits in the multivariate analysis. Equation [5.8] is illustrated below using calf 8 in Example 5.1.

Since $\mathbf{Z} = \mathbf{I}$ for calf 8, then [5.6] becomes $\mathbf{YD} = \mathbf{RR}^{-1}(\mathbf{y} - \mathbf{Xb}) = (\mathbf{y} - \mathbf{Xb})$. Thus:

$$\begin{pmatrix} YD_{81} \\ YD_{82} \end{pmatrix} = \begin{pmatrix} y_{81} - \hat{b}_1 \\ y_{82} - \hat{b}_2 \end{pmatrix} = \begin{pmatrix} 5.0 - 4.361 \\ 7.5 - 6.800 \end{pmatrix} = \begin{pmatrix} 0.639 \\ 0.700 \end{pmatrix}$$

Both parents of calf 8 are known, therefore:

$$\mathbf{DIAG}_8 = \mathbf{R}^{-1} + 2\mathbf{G}^{-1} = \begin{pmatrix} 0.1958 & -0.0858 \\ -0.0858 & 0.1211 \end{pmatrix}$$

and:

$$\mathbf{W}_1 = (\mathbf{DIAG})^{-1}2\mathbf{G}^{-1} = \begin{pmatrix} 0.8476 & -0.1191 \\ -0.0237 & 0.6092 \end{pmatrix} \quad \text{and}$$

$$\mathbf{W}_2 = \mathbf{I} - \mathbf{W}_1 = \begin{pmatrix} 0.1524 & 0.1191 \\ 0.0237 & 0.3908 \end{pmatrix}$$

Then, from [5.8]:

$$\begin{pmatrix} \hat{a}_{81} \\ \hat{a}_{82} \end{pmatrix} = \mathbf{W}_1\begin{pmatrix} PA_{81} \\ PA_{82} \end{pmatrix} + \mathbf{W}_2\begin{pmatrix} YD_{81} \\ YD_{82} \end{pmatrix} = \mathbf{W}_1\begin{pmatrix} 0.099 \\ 0.1735 \end{pmatrix} + \mathbf{W}_2\begin{pmatrix} 0.639 \\ 0.700 \end{pmatrix}$$
$$= \begin{pmatrix} 0.06325 \\ 0.10335 \end{pmatrix} + \begin{pmatrix} 0.18075 \\ 0.28870 \end{pmatrix} = \begin{pmatrix} 0.244 \\ 0.392 \end{pmatrix}$$

In both traits the contributions from PA accounted for about 26% of the breeding value of the calf.

In general, the estimates of breeding value for PWG from the multivariate analysis above are similar to those from the univariate analysis.

The maximum difference between the multivariate and univariate breeding values is 0.008 kg (calf 8). The similarity of the evaluations for PWG from both models is due to the fact that the genetic regression of WWG on PWG (0.45) is almost equal to the phenotypic regression (0.41) (Thompson and Meyer, 1986). However, the breeding values for WWG from the multivariate analysis are higher than those from the univariate analysis, with a maximum difference of 0.10 kg (calf 8) in favour of the multivariate analysis. Thus, much of the gain from the multivariate analysis is in WWG and this is due to its lower heritability, as mentioned earlier. However, there was only a slight reranking of animals for both traits in the multivariate analysis.

5.1.4 Accuracy of multivariate evaluations

One of the main advantages of MBLUP is the increase in the accuracy of evaluations. Presented below are estimates of reliabilities for the proofs for WWG and PWG from the multivariate analysis and the univariate analysis of each trait:

	Multivariate analysis				Univariate analysis reliability	
	Diagonals[a]		Reliability			
Animal	WWG	PWG	WWG	PWG	WWG	PWG
1	18.606	35.904	0.070	0.102	0.058	0.102
2	19.596	38.768	0.020	0.031	0.016	0.031
3	17.893	33.799	0.105	0.155	0.088	0.155
4	16.506	29.727	0.175	0.257	0.144	0.256
5	16.541	29.865	0.173	0.253	0.144	0.253
6	17.152	31.504	0.142	0.212	0.116	0.212
7	17.115	31.364	0.144	0.216	0.116	0.216
8	16.285	29.160	0.186	0.271	0.156	0.270

[a] Diagonal elements of the inverse of the coefficient matrix from multivariate analysis.

The reliability for the proof of animal i and trait j (r^2_{ij}) in the multivariate analysis was calculated as $r^2_{ij} = (g_{jj} - PEV_{ij})/g_{jj}$, where PEV_{ij} is the diagonal element of the coefficient matrix pertaining to animal i and trait j. This formula is obtained by rearranging the equation given for reliability in Section 3.2.2. For instance, the reliabilities for the proofs for WWG and PWG for animal 1, respectively, are:

$$r^2_{11} = (20 - 18.606)/20 = 0.070$$

$$r^2_{21} = (40 - 35.904)/40 = 0.102$$

Similarly to the estimates of breeding values, the reliabilities for animals for PWG from the multivariate analysis were essentially the same from the univariate analysis as $G_{ij} = r_p G_{jj}$ (Thompson and Meyer, 1986), where the jth trait is PWG and r_p is the phenotypic correlation. However, there was an increase of about 20% in reliability for WWG for each animal under the multivariate analysis compared with the univariate analysis. Again much of the gain in accuracy from the multivariate analysis is WWG.

5.1.5 Calculating daughter yield deviations in multivariate models

The equations for calculating daughter yield deviations (DYD) with a multivariate model are similar to [3.12] for the univariate model except that the weights are matrices of order equal to the order of traits. The equations can briefly be derived (Mrode and Swanson, 2004) as follows.

Given the daughter (prog) of a bull, with no progeny of her own, equation [5.8] becomes:

$$\hat{\mathbf{a}}_{prog} = \mathbf{W}_{\mathbf{1}_{\text{prog}}}\mathbf{PA} + \mathbf{W}_{\mathbf{2}_{\text{prog}}}(\mathbf{YD}) \qquad [5.9]$$

Let **PC** be expressed as in equation [5.7]:

$$\mathbf{PC} = 0.5\mathbf{G}^{-1}\sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - \hat{\mathbf{a}}_{mate}) \qquad [5.10]$$

Substituting equation [5.9] into equation [5.10] gives:

$$\mathbf{P} = 0.5\mathbf{G}^{-1}\sum \alpha_{prog}(\mathbf{W}_{\mathbf{1}_{\text{prog}}}\hat{\mathbf{a}}_{anim} + \mathbf{W}_{\mathbf{1}_{\text{prog}}}\hat{\mathbf{a}}_{mate} + \mathbf{W}_{\mathbf{2}_{\text{prog}}}2\mathbf{YD} - \hat{\mathbf{a}}_{mate})$$

Since the daughter has no offspring of her own, $\mathbf{W}_3 = 0$; therefore $\mathbf{W}_{\mathbf{1}_{\text{prog}}} = \mathbf{I} - \mathbf{W}_{\mathbf{2}_{\text{prog}}}$. Then:

$$\mathbf{PC} = 0.5\mathbf{G}^{-1}\sum \alpha_{prog}((\mathbf{I} - \mathbf{W}_{\mathbf{2}_{\text{prog}}})\hat{\mathbf{a}}_{anim} + \mathbf{W}_{\mathbf{2}_{\text{prog}}}(2\mathbf{YD} - \hat{\mathbf{a}}_{mate})) \qquad [5.11]$$

Substituting equation [5.11] into equation [5.7] and moving all terms involving $\hat{\mathbf{a}}_{anim}$ to the left-hand side gives:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + 2\mathbf{G}^{-1}\alpha_{par} + 0.5\mathbf{G}^{-1}\sum \mathbf{W}_{\mathbf{2}_{\text{prog}}}\alpha_{prog})\hat{\mathbf{a}}_{anim}$$
$$= 2\mathbf{G}^{-1}\alpha_{par}\mathbf{PA} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\mathbf{YD} + 0.5\mathbf{G}^{-1}\sum \mathbf{W}_{\mathbf{2}_{\text{prog}}}\alpha_{prog}(2\mathbf{YD} - \hat{\mathbf{a}}_{mate})$$

Pre-multiplying both sides of the equation by the inverse of the coefficient matrix gives:

$$\hat{\mathbf{a}}_{anim} = \mathbf{M}_1(\mathbf{PA}) + \mathbf{M}_2(\mathbf{YD}) + \mathbf{M}_3(\mathbf{DYD}) \qquad [5.12]$$

where:

$$\mathbf{DYD} = \sum \mathbf{W}_{\mathbf{2}_{\text{prog}}}\alpha_{prog}(\mathbf{YD} - \hat{\mathbf{u}}_{mate}) \Big/ \sum \mathbf{W}_{\mathbf{2}_{\text{prog}}}\alpha_{prog} \qquad [5.13]$$

and $\mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 = \mathbf{I}$, with $\mathbf{M}_1 = (\mathbf{DIAG})^{-1}2\mathbf{G}^{-1}\alpha_{par}$, $\mathbf{M}_2 = (\mathbf{DIAG})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})$ and $\mathbf{M}_3 = (\mathbf{DIAG})^{-1}0.5\mathbf{G}^{-1}\Sigma\mathbf{W}_{\mathbf{2}_{\text{prog}}}\alpha_{prog}$, where $(\mathbf{DIAG}) = (\mathbf{ZR}^{-1}\mathbf{Z} + 2\mathbf{G}^{-1}\alpha_{par} + 0.5\mathbf{G}^{-1}\Sigma\mathbf{W}_{\mathbf{2}_{\text{prog}}}\alpha_{prog})$. The matrix $\mathbf{W}_{\mathbf{2}_{\text{prog}}}$ in equation [5.13] for DYD is not

symmetrical and is of the order of traits and the full matrix has to be stored. This could make the computation of DYD cumbersome, especially with a large multivariate analysis or when a random regression model is implemented (see Chapter 8). For instance, in the Canadian test day model, which involves analysing milk, fat and protein yields and somatic cell count in the first three lactations, it is a matrix of order 36 (Jamrozik *et al.*, 1997). Thus, for computational ease, pre-multiply $\mathbf{W_{2_{prog}}}$ with $\mathbf{G}^{-1}$, and the equation for **DYD** becomes:

$$\mathbf{DYD} = \sum \mathbf{G}^{-1}\mathbf{W_{2_{prog}}}\,\alpha_{prog}\,(\mathbf{YD} - \hat{\mathbf{u}}_{mate})\,\mathbf{G}^{-1}\mathbf{W_{2_{prog}}}\,\alpha_{prog}$$

The product of $\mathbf{G}^{-1}\mathbf{W_{2_{prog}}}$ is symmetric and only upper or lower triangular elements need to be stored. The computation of DYD for a multivariate model is illustrated in Section 5.4, using the example dairy data.

5.2 Canonical Transformation

In the example discussed in Section 5.1.2, both traits were affected by the same fixed effect and all animals were measured for both traits. Thus the design matrices **X** and **Z** were the same for both traits or, in other words, the traits are said to have equal design matrices. In addition, there was only one random effect (animal effect) for each trait apart from the residual effect. Under these circumstances, the multivariate analysis can be simplified into *n* (number of traits) single trait analyses through what is called a canonical transformation (Thompson, 1977b). Canonical transformation involves using special matrices to transform the observations on several correlated traits into new variables that are uncorrelated with each other. These new variables are analysed by the usual methods for single trait evaluation, but the results (predictions) are transformed back to the original scale of the observations. Ducrocq and Besbes (1993) have presented a methodology for applying canonical transformation when design matrices are equal for all traits but with some animals having missing traits; details of the methodology with an illustration are given in Appendix E.2.

Let **y** be vectors of observations for several traits:

$$\text{var}(\mathbf{y}) = \mathbf{G} + \mathbf{R} \qquad [5.14]$$

where **G** and **R** are variance and covariance matrices for the additive genetic and residual effects, respectively. Assuming **G** and **R** are positive definite matrices, then there exists a matrix **Q** such that:

$$\mathbf{QRQ}' = \mathbf{I} \quad \text{and} \quad \mathbf{QGQ}' = \mathbf{W}$$

where **I** is an identity matrix and **W** is a diagonal matrix (Anderson, 1958). This implies that pre- and post-multiplication of **R** by the transformation matrix (**Q**) reduces it to an identity matrix and **G** to a diagonal matrix. The multiplication of **y** by **Q** yields a new vector of observation, $\mathbf{y}^*$, with

uncorrelated observations:

$$\mathbf{y}^* = \mathbf{Q}\mathbf{y}$$

and:

$$\text{var}(\mathbf{y}^*) = \mathbf{W} + \mathbf{I}$$

which is a diagonal matrix. Since there are no covariances between the transformed traits, they can be independently evaluated. The procedure for calculating the transformation matrix **Q** is given in Appendix E.1.

5.2.1 The model

A single trait analysis is usually carried out on each of the transformed variables. The model for the *i*th transformed variable can be written as:

$$\mathbf{y}_i^* = \mathbf{X}\mathbf{b}_i^* + \mathbf{Z}\mathbf{a}_i^* + \mathbf{e}_i^* \qquad [5.15]$$

where $\mathbf{y}_i^*$ = vector of transformed variable for the *i*th transformed trait; $\mathbf{b}_i^*$ = vector of fixed effect for the *i*th transformed variable *i*; $\mathbf{a}_i^*$ = vector of random animal effect for transformed trait *i*; $\mathbf{e}_i^*$ = vector of random residual error fot the *i*th transformed trait; and **X** and **Z** are incidence matrices relating records to fixed and random effects, respectively.

The MME to be solved to obtain the best linear unbiased estimator (BLUE) of $\mathbf{b}_i^*$ and the BLUP of $\mathbf{a}_i^*$ are the same as those presented in Section 3.1 for the univariate model. These equations are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}\alpha_i \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}}_i^* \\ \hat{\mathbf{a}}_i^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y}_i^* \\ \mathbf{Z}'\mathbf{y}_i^* \end{bmatrix}$$

As explained earlier, it is assumed for the *i*th trait that:

$$\text{var}(\mathbf{a}_i^*) = \mathbf{A}w_{ii}; \quad \text{var}(\mathbf{e}_i^*) = \mathbf{I}$$

and:

$$\text{var}(\mathbf{y}_i^*) = \mathbf{Z}\mathbf{A}\mathbf{Z}'w_{ii} + \mathbf{I}$$

where w_{ii} refers to the *i*th element of the diagonal matrix **W**.

The MME are solved for $\mathbf{b}_i^*$ and $\mathbf{a}_i^*$ and the transformation back to the original scale is achieved as:

$$\mathbf{b}_i = \mathbf{Q}^{-1}\mathbf{b}_i^* \qquad [5.16]$$

$$\mathbf{a}_i = \mathbf{Q}^{-1}\mathbf{a}_i^* \qquad [5.17]$$

Thus the multivariate analysis is simplified to *i* single trait evaluations.

5.2.2 An illustration

Example 5.2

The multivariate analysis for WWG and PWG in Section 5.1.2 is repeated below, carrying out a canonical transformation, assuming the same genetic parameters. The calculation of the transformation **Q** and the diagonal matrix **W** are in Appendix E.1. Presented in Table 5.2 are the data for all calves in the original scale and as transformed variables (VAR1 and

Table 5.2. Pre-weaning gain and post-weaning gain for beef calves on the original and transformed scales.

				Original scale		Transformed scale	
Calf	Sex	Sire	Dam	WWG	PWG	VAR1	VAR2
4	Male	1	–	4.5	6.8	0.208	1.269
5	Female	3	2	2.9	5.0	0.085	0.926
6	Female	1	2	3.9	6.8	0.109	1.259
7	Male	4	5	3.5	6.0	0.106	1.112
8	Male	3	6	5.0	7.5	0.236	1.400

VAR1 and VAR2, transformed variables for WWG and PWG, respectively.

VAR2). The observations are transformed into new uncorrelated variables using the matrix **Q**. Thus, for animal 4, the record would be transformed as:

$$\mathbf{Q}y_4 = \begin{bmatrix} 0.1659 & -0.0792 \\ 0.0168 & 0.1755 \end{bmatrix} \begin{bmatrix} 4.5 \\ 6.8 \end{bmatrix} = \begin{bmatrix} 0.208 \\ 1.269 \end{bmatrix}$$

The residual variance for each of the transformed variables is 1; thus heritability for the ith transformed variable = $w_{ii}/(1 + w_{ii})$ and $\alpha_i = 1/w_{ii}$. Therefore $h_1^2 = 0.247$, $h_2^2 = 0.573$, $\alpha_1 = 1/0.3283 = 3.046$ and $\alpha_2 = 1/1.3436 = 0.744$. A single trait analysis is carried out on the transformed variates for WWG and PWG using the model and the MME in Section 5.2.1, and solutions are transformed back to the original scale.

Setting up design matrices

The matrix **X**, which relates records for either VAR1 or VAR2 to sex effects, is exactly the same as the matrix $\mathbf{X}_1$ in Section 5.1.2. Similarly, **Z** is the same as $\mathbf{Z}_1$ in Section 5.1.2. For animals with records, the vectors of observations $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ are equal to the column of transformed variates for WWG and PWG gains, respectively, in Table 5.2. The matrices in the MME are easily obtained through matrix multiplication and the addition to the animal equations of $\mathbf{A}^{-1}\alpha_1$ for VAR1 and $\mathbf{A}^{-1}\alpha_2$ for VAR2. $\mathbf{A}^{-1}$ has been given earlier, in Section 5.1.2. For instance, the MME for VAR1 only are:

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{bmatrix} = \begin{bmatrix} 3.0 & 0.0 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 1.000 & 1.000 \\ 0.0 & 2.0 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 \\ 0.0 & 0.0 & 5.584 & 1.523 & 0.000 & -2.031 & 0.000 & -3.046 & 0.000 & 0.000 \\ 0.0 & 0.0 & 1.523 & 6.092 & 1.523 & 0.000 & -3.046 & -3.046 & 0.000 & 0.000 \\ 0.0 & 0.0 & 0.000 & 1.523 & 6.092 & 0.000 & -3.046 & 1.523 & 0.000 & -3.046 \\ 1.0 & 0.0 & -2.031 & 0.000 & 0.000 & 6.584 & 1.523 & 0.000 & -3.046 & 0.000 \\ 0.0 & 1.0 & 0.000 & -3.046 & -3.046 & 1.523 & 8.615 & 0.000 & -3.046 & 0.000 \\ 0.0 & 1.0 & -3.046 & -3.046 & 1.523 & 0.000 & 0.000 & 8.615 & 0.000 & -3.046 \\ 1.0 & 0.0 & 0.000 & 0.000 & 0.000 & -3.046 & -3.046 & 0.000 & 7.092 & 0.000 \\ 1.0 & 0.0 & 0.000 & 0.000 & -3.046 & 0.000 & 0.000 & -3.046 & 0.000 & 7.092 \end{bmatrix}^{-1} \begin{bmatrix} 0.549 \\ 0.194 \\ 0.000 \\ 0.000 \\ 0.000 \\ 0.208 \\ 0.085 \\ 0.108 \\ 0.105 \\ 0.235 \end{bmatrix}$$

Solving the MME for each transformed trait by direct inversion of the coefficient matrix gives the following solutions on the canonical scales. Given also are solutions for WWG and PWG after transforming the solutions for the transformed variates to the original scale.

	Canonical scale		Original scale	
Effects	VAR1	VAR2	WWG	PWG
Sex				
1	0.185	1.266	4.361	6.800
2	0.098	1.089	3.397	5.880
Animal				
1	0.003	0.052	0.151	0.280
2	–0.002	–0.002	–0.015	–0.008
3	0.000	–0.031	–0.078	–0.170
4	–0.001	–0.002	–0.010	–0.013
5	–0.007	–0.088	–0.270	–0.478
6	0.005	0.095	0.276	0.517
7	–0.015	–0.089	–0.316	–0.479
8	0.009	0.073	0.244	0.392

The solutions are exactly the same as those obtained from the multivariate analysis in Section 5.1. The solutions are transformed to the original scale using equations [5.16] and [5.17]. For instance, the solutions for animal 1 for both traits on the original scale are:

$$\begin{bmatrix} \hat{a}_{11} \\ \hat{a}_{12} \end{bmatrix} = \begin{bmatrix} 5.7651 & 2.6006 \\ -0.5503 & 5.4495 \end{bmatrix} \begin{bmatrix} 0.0029 \\ 0.0516 \end{bmatrix} = \begin{bmatrix} 0.151 \\ 0.280 \end{bmatrix}$$

5.3 Equal Design Matrices with Missing Records

When all traits in a multivariate analysis are not observed in all animals, the same methodology as that described in Section 5.1 can also be employed to evaluate animals, except that different residual covariance matrices have to be set up corresponding to a different combination of traits present. If the loss of traits is sequential, that is, the presence of the *i*th record implies the presence of 1 to $(i - 1)$ records, then the number of residual covariance matrices is equal to the number of traits. In general, if there are n traits, there are $(2^n - 1)$ possible combinations of observed traits and therefore residual covariance matrices (Quaas, 1984).

5.3.1 An illustration

Example 5.3
For illustrative purposes, consider the data set below, obtained by modifying the data in Table 5.1:

Calf	Sex	Sire	Dam	WWG (kg)	PWG (kg)
4	Male	1	–	4.5	–
5	Female	3	2	2.9	5.0
6	Female	1	2	3.9	6.8
7	Male	4	5	3.5	6.0
8	Male	3	6	5.0	7.5
9	Female	7	8	4.0	–

The model for the analysis is the same as in Section 5.1.1 and the same genetic parameters applied in Example 5.1 are assumed. The loss of records is sequential; there are therefore two residual covariance matrices. For animals with missing records for PWG, the residual covariance matrix ($\mathbf{R}_m$) and its inverse ($\mathbf{R}_m^{-1}$) are $\mathbf{R}_m = r_{m11} = 40$ and $\mathbf{R}_m^{-1} = r_m^{11} = \frac{1}{40} = 0.025$. For animals with records for both WWG and PWG, the residual covariance matrix ($\mathbf{R}_o$) and its inverse ($\mathbf{R}_o^{-1}$) are:

$$\mathbf{R}_o = \begin{bmatrix} 40 & 11 \\ 11 & 30 \end{bmatrix} \quad \text{and} \quad \mathbf{R}_o^{-1} = \begin{bmatrix} 0.028 & -0.010 \\ -0.010 & 0.037 \end{bmatrix}$$

SETTING UP DESIGN MATRICES

The $\mathbf{X}_1'$ and $\mathbf{X}_2'$ matrices, which relate sex effects for WWG and PWG, respectively, are:

$$\mathbf{X}_1' = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2' = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{X}_1'\mathbf{X}_1 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2'\mathbf{X}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

In setting up $\mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1$, it is necessary to account for the fact that animals (one male and one female) have missing records for PWG. Thus:

$$\mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 = r_m^{11}\mathbf{W}'\mathbf{W} + r_o^{11}\mathbf{B}'\mathbf{B} = 0.025\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 0.028\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0.081 & 0.000 \\ 0.000 & 0.081 \end{bmatrix}$$

where the matrix $\mathbf{W}$ relates WWG records for animals 4 and 9 with missing records for PWG to sex effects and $\mathbf{B}$ relates WWG records for calves 5,

6, 7 and 8 to sex effects. The matrices $\mathbf{W}'$ and $\mathbf{B}'$ are:

$$\mathbf{W}' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B}' = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

However, all animals recorded for PWG also had records for WWG; therefore:

$$\mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 = r_o^{22}\mathbf{X}_2'\mathbf{X}_2 = 0.037\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0.074 & 0 \\ 0 & 0.074 \end{bmatrix}$$

and:

$$\mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 = r_o^{12}\mathbf{X}_1'\mathbf{X}_2 = -0.010\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -0.02 & 0.00 \\ 0.00 & -0.02 \end{bmatrix}$$

Excluding ancestors, the matrix $\mathbf{Z}_1$ is an identity matrix because every animal has a record for WWG. Therefore, $\mathbf{Z}_1'\mathbf{Z}_1 = \mathbf{I}$ and:

$$\mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1 = \text{diag}(0.025, 0.028, 0.028, 0.028, 0.028, 0.025)$$

However:

$$\mathbf{Z}_2 = \text{diag}(0, 1, 1, 1, 1, 0)$$

indicating that calves 4 and 9 have no records for PWG, and:

$$\mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2 = \text{diag}(0.0, 0.037, 0.037, 0.037, 0.037, 0.0)$$

To account for ancestors (animals 1 to 3), $\mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1$ and $\mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2$ given above are augmented with three rows and columns of zeros.

The other matrices in the MME can be calculated through matrix multiplication. The matrix $\mathbf{A}^{-1}$ can be set up and $\mathbf{A}^{-1}*\mathbf{G}^{-1}$ (where * means the Kronecker product) added to the appropriate matrices, as described in Section 5.1.2, to obtain the MME. The MME are too large to be presented but solutions from solving the equations are shown below, together with solutions from the univariate analyses of WWG and PWG.

	Multivariate analysis		Univariate analysis	
Effects	WWG	PWG	WWG	PWG
Sex				
1	4.327	6.793	4.325	6.784
2	3.598	5.966	3.599	5.873
Animal				
1	0.154	0.288	0.100	0.273
2	–0.059	–0.053	–0.059	0.000
3	–0.061	–0.163	–0.022	–0.165
4	0.027	0.036	0.022	–0.025
5	–0.307	–0.521	–0.218	–0.464
6	0.235	0.477	0.138	0.518
7	–0.280	–0.452	–0.211	–0.460
8	0.272	0.407	0.214	0.392
9	0.077	0.051	0.081	–0.034

The differences for sex solutions for WWG from the multivariate and univariate analyses are very similar to those in Section 5.1 since there are no missing records in WWG. However, sex differences in the two analyses are different for PWG due to the missing records. Again, most of the benefit in terms of breeding values from the multivariate analysis was observed in WWG, as explained in Section 5.1. However, for the calves with missing records for PWG, there was a substantial increase in their proofs compared with the estimates from the univariate analysis. The proofs for these calves for PWG are based on pedigree information only in the univariate analysis but include information from the records for WWG in the multivariate analysis due the genetic and residual correlations between the two traits. Thus the inclusion of a correlated trait in a multivariate analysis is of much benefit to animals with missing records for the other trait.

5.4 Cholesky Transformation

When all records are measured in all animals, MBLUP may be simplified by a canonical transformation, as described in Section 5.2. However, if animals have some records missing and the loss of records is sequential, then a Cholesky transformation can be applied (Quaas, 1984). Such situations can arise, for example, in dairy cattle due to sequential culling, and different lactations are regarded as different traits.

5.4.1 Calculating the transformation matrix and defining the model

Cholesky transformation involves forming transformed variables (traits) which are environmentally independent of each other; that is, there is no residual covariance among them and therefore the residual covariance matrix for the transformed traits is an identity matrix. The transformation matrix $\mathbf{T}^{-1}$ is obtained by carrying out a Cholesky decomposition of $\mathbf{R}$, the residual covariance matrix for the traits, such that:

$$\mathbf{R} = \mathbf{T}\,\mathbf{T}'$$

where $\mathbf{T}$ is a lower triangular matrix. The transformation matrix $\mathbf{T}^{-1}$ is the inverse of $\mathbf{T}$. The formula for calculating $\mathbf{T}$ is given in Appendix E.3.

The vector of observation $\mathbf{y}_{ki}$ for the ith animal is transformed as:

$$\mathbf{y}^{*}_{ki} = \mathbf{T}^{-1}\mathbf{y}^{*}_{ki}$$

where k is the number of traits recorded and $\mathbf{y}^{*}_{ki}$ is the transformed vector. If traits are missing in $\mathbf{y}_{ki}$, then the corresponding rows of $\mathbf{T}^{-1}$ are set to zero when transforming the vector of observation. Thus, if $\mathbf{y}_{ki}$ is a vector

of observation of n traits for the ith animal, the transformation of $\mathbf{y}$ can be illustrated as below:

$$\begin{aligned} y^*_{11} &= t^{11} y_{11} \\ y^*_{21} &= t^{21} y_{11} + t^{22} y_{21} \\ &\vdots \quad \vdots \quad \vdots \\ y^*_{n1} &= t^{n1} y_{11} + t^{n2} y_{21} + t^{nn} y_{n1} \end{aligned}$$

where t^{ij} above are the elements of $\mathbf{T}^{-1}$.

Given that the variance of $\mathbf{y}_{ki}$ is:

$$\text{var}(\mathbf{y}) = \mathbf{G} + \mathbf{R}$$

the variance of the transformed variables becomes:

$$\begin{aligned} \text{var}(\mathbf{y}^*) &= \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})' + \mathbf{I} \\ &= \mathbf{G}^* + \mathbf{I} \\ &= \mathbf{M} + \mathbf{I} \end{aligned} \qquad [5.18]$$

where $\mathbf{G}$ is the covariance matrix for additive genetic effect and $\mathbf{G}^*$ is the transformed additive genetic covariance matrix. Note that $\mathbf{G}^*$ is not diagonal. Vectors of solutions ($\mathbf{b}^*_i$ and $\mathbf{a}^*_i$) are tranformed back to the original scale ($\mathbf{b}_i$ and $\mathbf{a}_i$) as:

$$\mathbf{b}_i = \mathbf{T}\mathbf{b}^*_i \qquad [5.19]$$

$$\mathbf{a}_i = \mathbf{T}\mathbf{a}^*_i \qquad [5.20]$$

5.4.2 An illustration

Example 5.4

The methodology is illustrated using the growth data on beef calves in Section 5.3.1. The residual and additive genetic covariance matrices were:

$$\mathbf{R} = \begin{bmatrix} 40 & 11 \\ 11 & 30 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} 20 & 18 \\ 18 & 40 \end{bmatrix}$$

Now carry out a Cholesky decomposition of $\mathbf{R}$ such that $\mathbf{R} = \mathbf{TT}'$. For the $\mathbf{R}$ above:

$$\mathbf{T} = \begin{bmatrix} 6.324555 & 0.000 \\ 1.739253 & 5.193746 \end{bmatrix} \quad \text{with } \mathbf{T}^{-1} = \begin{bmatrix} 0.1581139 & 0.000 \\ -0.052948 & 0.1925393 \end{bmatrix}$$

The transformed additive genetic covariance matrix ($\mathbf{M}$) is:

$$\mathbf{M} = \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1})' = \begin{bmatrix} 0.5000 & 0.380539 \\ 0.380539 & 1.171972 \end{bmatrix} \quad \text{and}$$

$$\mathbf{M}^{-1} = \begin{bmatrix} 2.654723 & -0.862556 \\ -0.862556 & 1.133334 \end{bmatrix}$$

The transformed variables are calculated using the transforming matrix $\mathbf{T}^{-1}$. For the first two animals, the transformation is as follows.

Animal 1:

$$y^*_{11} = t^{11} y_{11} = 0.1581139(4.5) = 0.712$$

Animal 2:

$$y^*_{12} = t^{11} y_{12} = 0.1581139(2.9) = 0.459$$
$$y^*_{22} = t^{21} y_{12} + t^{22} y_{22} = -0.052948(2.9) + 0.1925393(5.0) = 0.809$$

where y_{ij} and y^*_{ij} are the original and transformed observations, respectively, for the ith trait and jth animal. The transformed variables for all calves are shown in the table below.

				Original traits		Transformed traits	
Calf	Sex	Sire	Dam	WWG	PWG	y^*_1	y^*_2
4	Male	1	–	4.5	–	0.712	–
5	Female	3	2	2.9	5.0	0.456	0.809
6	Female	1	2	3.9	6.8	0.617	1.103
7	Male	4	5	3.5	6.0	0.553	0.970
8	Male	3	6	5.0	7.5	0.791	1.179
9	Female	7	8	4.0	–	0.632	–

The model for analysis is the same as in Section 5.3.1 except that the variance of $\mathbf{y}^*$ is now:

$$\begin{aligned}\text{var}(\mathbf{y}^*) &= \mathbf{T}^{-1}\mathbf{G}(\mathbf{T}^{-1}) + \mathbf{I} \\ &= \mathbf{M} + \mathbf{I}\end{aligned}$$

The MME for the transformed variables are:

$$\begin{bmatrix} \hat{\mathbf{b}}^*_1 \\ \hat{\mathbf{b}}^*_2 \\ \hat{\mathbf{a}}^*_1 \\ \hat{\mathbf{a}}^*_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{0} & \mathbf{X}'_1\mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_2\mathbf{X}_2 & \mathbf{0} & \mathbf{X}'_2\mathbf{Z}_2 \\ \mathbf{Z}'_1\mathbf{X}_1 & \mathbf{0} & \mathbf{Z}'_1\mathbf{Z}_1 + \mathbf{A}^{-1}m^{11} & \mathbf{A}^{-1}m^{12} \\ \mathbf{0} & \mathbf{Z}'_2\mathbf{X}_2 & \mathbf{A}^{-1}m^{21} & \mathbf{Z}'_2\mathbf{Z}_2 + \mathbf{A}^{-1}m^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1\mathbf{y}^*_1 \\ \mathbf{X}'_2\mathbf{y}^*_2 \\ \mathbf{Z}'_1\mathbf{y}^*_1 \\ \mathbf{Z}'_2\mathbf{y}^*_2 \end{bmatrix}$$

The design matrices $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ and the inverse of the relationship matrix are exactly as in Section 5.3.1. The vector of observations $\mathbf{y}^*$ now contains the transformed variables shown in the above table. All other matrices in the MME above can be derived from the design matrices and vector of observations through matrix multiplication and the addition of the $\mathbf{A}^{-1}m^{11}$ and $\mathbf{A}^{-1}m^{22}$ to the animal equations for traits 1 and 2, respectively, and $\mathbf{A}^{-1}m^{12}$ to the block of animal equations for trait 1 by trait 2 and $\mathbf{A}^{-1}m^{21}$ to the block of equations for trait 2 by trait 1 that pertains to animals. The MME have not been shown because they are too large. However,

solving the MME gives the following solutions on the transformed scale. The solutions transformed to the original scale are also shown.

	Transformed scale		Original scale	
Effects	WWG	PWG	WWG	PWG
Sex				
1	0.684	1.079	4.327	6.793
2	0.569	0.958	3.598	5.966
Animal				
1	0.024	0.047	0.154	0.288
2	–0.009	–0.007	–0.059	–0.053
3	–0.010	–0.028	–0.061	–0.163
4	0.004	0.005	0.027	0.036
5	–0.048	–0.084	–0.307	–0.521
6	0.037	0.079	0.235	0.477
7	–0.044	–0.072	–0.280	–0.452
8	0.043	0.064	0.272	0.407
9	0.012	0.006	0.077	0.051

These are exactly the same solutions as those obtained in Section 5.3 without any transformation. The number of non-zero elements was 188 in the analysis on the transformed variables compared with 208 when no transformation is carried out. This difference could be substantial with large data sets and could reduce storage requirements when data are transformed. The solutions were transformed to the original scale using [5.19] and [5.20]. Thus the solutions for male calves on the original scale are:

$$\begin{bmatrix} \hat{b}_{11} \\ \hat{b}_{12} \end{bmatrix} = \begin{bmatrix} 6.324555 & 0.000 \\ 1.739253 & 5.193746 \end{bmatrix} \begin{bmatrix} 0.68420 \\ 1.07871 \end{bmatrix} = \begin{bmatrix} 4.327 \\ 6.793 \end{bmatrix}$$

5.5 Unequal Design Matrices

This situation arises when traits in the multivariate analysis are affected by different fixed or random effects – for instance, the multivariate analysis of yields in different lactations as different traits. Due to the fact that calving in different parities occurs in different years, herd–year–season (HYS) effects associated with each lactation are different, and an appropriate model should include different HYS effects for yield in each parity. An example where random effects might be different for different traits is the joint analysis for weaning weight and lean per cent in beef cattle. It might be considered that random maternal effect (see Chapter 6) is only important for weaning weight and the model for the analysis will include maternal effects only for weaning weight.

5.5.1 Numerical example

Example 5.5
Using the fat yield data in Chapter 4 analysed with a repeatability model, the principles of a multivariate analysis with unequal design are illustrated below, considering yield in each parity as different traits and fitting a different HYS effect for each trait. The data with each lactational yield treated as different traits and HYS recoded for each trait are as follows:

Cow	Sire	Dam	HYS1	HYS2	FAT1	FAT2
4	1	2	1	1	201	280
5	3	2	1	2	150	200
6	1	5	2	1	160	190
7	3	4	1	1	180	250
8	1	7	2	2	285	300

HYS1, HYS2, herd–year–season for parity 1 and 2, respectively; FAT1, FAT2, fat yield in parity 1 and 2.

The aim is to carry out a multivariate estimate of breeding values for fat yield in lactation 1 (FAT1) and 2 (FAT2) as different traits. Assume the genetic parameters are:

$$\mathbf{R} = \begin{bmatrix} 65 & 27 \\ 27 & 70 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} 35 & 28 \\ 28 & 30 \end{bmatrix}$$

and the inverses are:

$$\mathbf{R}^{-1} = \begin{bmatrix} 0.018 & -0.007 \\ -0.007 & 0.017 \end{bmatrix} \quad \text{and} \quad \mathbf{G}^{-1} = \begin{bmatrix} 0.113 & -0.105 \\ -0.105 & 0.132 \end{bmatrix}$$

The model for the analysis is the same as in Section 5.1 but the MME are different in Section 5.1 because HYS effects are peculiar to each trait. The MME with the equations written out separately for each trait are:

$$\begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{11}\mathbf{Z}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{Z}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{21}\mathbf{Z}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{Z}_2 \\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{X}_2 & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1 + \mathbf{A}^{-1}g^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{12} \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1 + \mathbf{A}^{-1}g^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2 + \mathbf{A}^{-1}g^{22} \end{bmatrix}^{-1}$$
$$\begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{22}\mathbf{y}_2 \\ \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{y}_2 \end{bmatrix}$$

Setting up the design matrices and MME

The matrix $\mathbf{X}_1$ now relates HYS effects to FAT1 while $\mathbf{X}_2$ relates HYS effects to FAT2. The transposes of these matrices are:

$$\mathbf{X}_1' = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2' = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are equal and they are identity matrices of order 5 by 5 considering only animals with records. The matrix $\mathbf{A}^{-1}$ has been presented in Section 4.1. The remaining matrices in MME can be obtained as described in previous sections. The MME have not been presented because they are too large. The solutions to the MME are as follows:

	Solutions			
	Multivariate analysis		Univariate analysis	
Effects	FAT1	FAT2	FAT1	FAT2
HYS				
1	175.7	243.2	175.8	237.1
2	219.6	240.6	220.4	250.0
Animal				
1	8.969	8.840	6.933	8.665
2	–2.999	–2.777	–2.590	–2.244
3	–5.970	–6.063	–4.341	–6.422
4	11.754	11.658	9.103	12.197
5	–16.253	–15.824	–12.992	–15.563
6	–17.314	–15.719	–15.197	–11.149
7	8.690	8.138	7.566	7.696
8	22.702	20.931	19.417	15.560

Similarly to the results in Section 5.1.2, the largest increase in breeding value under the multivariate analysis compared with the univariate was in FAT2. This may be due to the lower heritability of FAT2 compared with FAT1, as explained earlier.

Compared with the results from the repeatability model (Section 4.2) on the same data with corresponding estimates of genetic parameters, the mean breeding values for FAT1 and FAT2 for animals in the multivariate analysis are similar to the breeding value estimates from the former. The ranking of animals is the same under both models. Also the differences between solutions for corresponding levels of HYS are very similar. In general, the repeatability model on successive records of animals is very efficient compared with the multivariate model, especially when the genetic correlation among records is high. The genetic correlation used for the multivariate analysis was 0.86. Visscher (1991) reported a loss of

0 to 5% in efficiency in genetic gain with a repeatability model on first and second fat yield compared with the multivariate model using a selection index. Mrode and Swanson (1995) reported a rank correlation of 0.98 between breeding values estimates for milk yield in first and second lactations, from a repeatability model and multivariate analysis for bulls with 60 or more daughters. The benefit of the repeatability model compared with the multivariate is that it is less computationally demanding and fewer estimates of genetic parameters are required.

If there are missing records in addition to unequal design matrices for traits in a multivariate analysis, the analysis can be carried out using the principles outlined in Section 5.3, defining different residual covariance matrices for each pattern of missing traits.

5.5.2 Illustrating the computation of DYD from a multivariate model

The computation of DYD from a multivariate model is illustrated using sire 1 with three daughters (cows 4, 6 and 8) in Example 5.4. As shown in Section 5.1, since each daughter has one record per each trait, $\mathbf{YD}_{ij}$ for the daughter i and trait j equals $(y_{ij} - x_{ij}\hat{b})$. Thus:

$$\begin{pmatrix} YD_{41} \\ YD_{42} \end{pmatrix} = \begin{pmatrix} 201 - 175.5 \\ 280 - 243.2 \end{pmatrix} = \begin{pmatrix} 25.7 \\ 36.8 \end{pmatrix}; \quad \begin{pmatrix} YD_{61} \\ YD_{62} \end{pmatrix} = \begin{pmatrix} 160 - 219.6 \\ 190 - 243.2 \end{pmatrix} = \begin{pmatrix} -59.6 \\ -53.2 \end{pmatrix}$$

and

$$\begin{pmatrix} YD_{81} \\ YD_{82} \end{pmatrix} = \begin{pmatrix} 285 - 219.6 \\ 300 - 240.6 \end{pmatrix} = \begin{pmatrix} 65.4 \\ 59.4 \end{pmatrix}$$

For all three daughters, the dams are known; therefore $\mathbf{W}_{\mathbf{2}_{\text{prog}}}$ in [5.13] is the same for all daughters and is:

$$\begin{aligned} \mathbf{W}_{\mathbf{2}_{\text{prog}}} &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + 2\mathbf{G}^{-1})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}) \\ &= \begin{pmatrix} 0.2439 & -0.2176 \\ -0.2176 & 0.2802 \end{pmatrix}^{-1} \begin{pmatrix} 0.0183 & -0.0071 \\ -0.0071 & 0.0170 \end{pmatrix} = \begin{pmatrix} 0.1713 & 0.0821 \\ 0.1078 & 0.1244 \end{pmatrix} \end{aligned}$$

The correction of the daughters' YD for the breeding values of the mates of the sire is as follows:

$$\begin{pmatrix} 2YD_{41} - \hat{a}_{21} \\ 2YD_{42} - \hat{a}_{22} \end{pmatrix} = \begin{pmatrix} 51.4 - (-2.999) \\ 73.6 - (-2.777) \end{pmatrix} = \begin{pmatrix} 54.399 \\ 76.377 \end{pmatrix}$$

$$\begin{pmatrix} 2YD_{61} - \hat{a}_{51} \\ 2YD_{62} - \hat{a}_{52} \end{pmatrix} = \begin{pmatrix} -119.2 - (-16.253) \\ -106.4 - (-15.824) \end{pmatrix} = \begin{pmatrix} -102.947 \\ -90.576 \end{pmatrix}$$

$$\begin{pmatrix} 2YD_{81} - \hat{a}_{71} \\ 2YD_{82} - \hat{a}_{72} \end{pmatrix} = \begin{pmatrix} 130.8 - 8.690 \\ 118.8 - 8.138 \end{pmatrix} = \begin{pmatrix} 122.110 \\ 110.662 \end{pmatrix}$$

Since α_{prog} equals 1 for all daughters of the bull, **DYD** for sire 1, using [5.13], is:

$$\mathbf{DYD} = (3\mathbf{W_2}_{\text{prog}})^{-1}\left[\mathbf{W_2}_{\text{prog}}\begin{pmatrix}53.399\\76.377\end{pmatrix} + \mathbf{W_2}_{\text{prog}}\begin{pmatrix}-102.947\\-90.576\end{pmatrix} + \mathbf{W_2}_{\text{prog}}\begin{pmatrix}122.110\\110.662\end{pmatrix}\right]$$
$$= \begin{pmatrix}24.5207\\32.1543\end{pmatrix}$$

Using equation [5.12], the breeding value of sire 1 can be calculated as:

$$\begin{pmatrix}\hat{a}_{11}\\\hat{a}_{12}\end{pmatrix} = \mathbf{M}_2\begin{pmatrix}24.5207\\32.1543\end{pmatrix} = \begin{pmatrix}7.439\\7.387\end{pmatrix}$$

where:

$$\mathbf{M}_2 = \left(2G^{-1}0.5 + 0.5G^{-1}\sum_3 \mathbf{W_2}_{\text{prog}}\alpha_{prog}\right)^{-1}\left(0.5G^{-1}\sum_3 \mathbf{W_2}_{\text{prog}}\alpha_{prog}\right)$$

$$\mathbf{M}_2 = \begin{pmatrix}0.1247 & -0.1110\\-0.1110 & 0.1432\end{pmatrix}^{-1}\begin{pmatrix}0.0120 & -0.0058\\-0.0058 & 0.0116\end{pmatrix} = \begin{pmatrix}0.1937 & 0.0836\\0.1099 & 0.1459\end{pmatrix}$$

The vector of breeding value calculated for sire 1 using [5.12] is slightly lower than that shown earlier in the table of results as contributions from the grand-progeny of the sire are not included.

5.6 Multivariate Models with No Environmental Covariance

In some cases, a multivariate analysis may be necessary when individual animals have records for one trait (or subset of traits), but relatives have records on a different trait (or subset of traits). For instance, in beef cattle, if selection is for dual-purpose sires, male and female calves might be reared in different environments (different feedlots) and body weight recorded in male calves and milk yield in female calves. The evaluation of the sires will be based on multivariate analysis of these two traits. A special feature of such a multivariate analysis is that there is no environmental covariance between the traits as the two traits are not observed in the same individual. In Section 5.6.1, the details of such a model are discussed and its application to example data is illustrated.

Also, when the same trait is measured on relatives in different environments such that the genetic correlation between performances in the two environments is not one, a multivariate analysis might be the optimum means to evaluate sires. For example, milk yield may be recorded on the

daughters of a bull in two different environments, say, in a tropical environment and in a temperate environment. Such a multivariate analysis will treat milk yield in the various environments as different traits. However, as the number of environments increases, the data might be associated with a heterogeneous fixed effects structure that might be difficult to model correctly in multivariate analysis, such that it might be useful, for practical purposes of implementation, to analyse not the original data but summaries of the data. A very good illustration of such a multivariate analysis is the multi-trait sire model used by the international bull evaluation service Interbull, Uppsala, Sweden, for the across-country evaluation of dairy sires. This multi-trait sire model, commonly referred to as MACE (multi-trait across-country evaluations), analyses deregressed breeding values (DRB) of sires in different countries as different traits. The use of DRB could be regarded as utilizing a variable that summarizes the daughter performances of bulls in different countries. This avoids the need to model at the Interbull centre the heterogeneous fixed effects structure, such as different herd management systems and complex national climatic conditions associated with the daughters' milk performance records in the different countries. MACE plays a very important role in the international trade of dairy cattle and in Section 5.6.2 the model for MACE is discussed and illustrated.

5.6.1 Different traits recorded on relatives

Defining the model

In this situation, with different traits recorded on relatives in different environments, the different traits are not observed on the same individual; consequently, there is not environmental covariance between the traits. Therefore the residual covariance matrix **R** is diagonal. Thus, for n traits:

$$\mathbf{R} = \text{diag}(\sigma^2_{e1}, \sigma^2_{e2}, \ldots, \sigma^2_{en}) = \text{diag}(r_{11}, r_{22}, \ldots, r_{nn})$$

and:

$$\mathbf{R}^{-1} = \text{diag}(r^{11}, r^{22}, \ldots, r^{nn})$$

However, var(**a**), where $\mathbf{a}' = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$, the vector of breeding values, is:

$$\text{var}(\mathbf{a}) = \mathbf{A} * \mathbf{G}$$

where * refers to the direct product, **A** the relationship matrix and **G** the covariance matrix for additive genetic effects. Schaeffer *et al.* (1978) discussed this model in detail but from the standpoint of variance component estimation.

Assuming there are two traits, the model for analysis is as given in equation [5.1] but with **R** and **G** defined as above. The MME for the BLUP of **a** and estimable functions of **b** are similar to those presented in Section

5.4.2, when data are transformed using Cholesky decomposition, since there is no residual covariance between variables. The MME are:

$$\begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{bmatrix} \begin{bmatrix} r^{11}\mathbf{X}_1'\mathbf{X}_1 & 0 & r^{11}\mathbf{X}_1'\mathbf{Z}_1 & 0 \\ 0 & r^{22}\mathbf{X}_2'\mathbf{X}_2 & 0 & r^{22}\mathbf{X}_2'\mathbf{Z}_2 \\ r^{11}\mathbf{Z}_1'\mathbf{X}_1 & 0 & r^{11}\mathbf{Z}_1'\mathbf{Z}_1 + \mathbf{A}^{-1}g^{11} & \mathbf{A}^{-1}g^{12} \\ 0 & r^{22}\mathbf{Z}_2'\mathbf{X}_2 & \mathbf{A}^{-1}g^{21} & r^{22}\mathbf{Z}_2'\mathbf{Z}_2 + \mathbf{A}^{-1}g^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{y}_2 \\ \mathbf{Z}_1'\mathbf{y}_1 \\ \mathbf{Z}_2'\mathbf{y}_2 \end{bmatrix}$$

An illustration

Example 5.6

Consider the following data on the progeny of three sires born in the same herd; assuming that selection is for dual-purpose sires, such that the male and female calves are raised on different feeding regimes, with males recorded for yearling weight (kg) and females for fat yield (kg):

Calf	Sex	Sire	Dam	HYS	Yearling weight	Fat yield
4	Female	1	Unknown	–	–	–
9	Male	1	4	1	375.0	–
10	Male	2	5	2	250.0	–
11	Male	1	6	2	300.0	–
12	Male	3	Unknown	1	450.0	–
13	Female	1	7	1	–	200.0
14	Female	3	8	2	–	160.0
15	Female	2	Unknown	3	–	150.0
16	Female	2	13	2	–	250.0
17	Female	3	15	3	–	175.0

The aim is to estimate HYS effects for both traits and predict breeding values for yearling weight and fat yield for all animals, carrying out a multivariate analysis. Note that animal 4 is just an ancestor and has no yield record for either trait. Assume that the additive genetic covariance matrix (**G**) is:

$$\mathbf{G} = \begin{bmatrix} 43 & 18 \\ 18 & 30 \end{bmatrix}$$

and:

$$\mathbf{R} = \text{diag}(77, 70)$$

The inverse of **R** is:

$$\mathbf{R}^{-1} = \text{diag}(1/77,\ 1/70)$$

and:

$$\mathbf{G}^{-1} = \begin{bmatrix} 0.0311 & -0.0186 \\ -0.0186 & 0.0445 \end{bmatrix}$$

The MME given earlier can easily be set up using the principles discussed so far in this chapter. Solving the MME by the direct inverse of the coefficient matrix gave the following solutions:

	Solutions	
Effects	Yearling weight	Fat (kg)
HYS		
1	411.833	193.299
2	275.955	205.344
3	–	163.315
Animal		
1	–0.472	2.519
2	–3.350	0.381
3	0.856	–3.208
4	–5.142	–3.936
5	–4.778	–2.000
6	4.778	2.000
7	2.177	3.628
8	–4.940	–5.251
9	–10.234	–3.817
10	–8.842	–2.810
11	6.932	4.260
12	11.568	3.060
13	3.029	6.701
14	–6.395	–11.485
15	–2.797	–1.680
16	4.193	10.797
17	0.526	0.050

Selection of dual-purpose sires will be based on some combination of breeding value estimates for yearling weight and fat yield. If equal weights were given to yearling weight and fat yield, sire 1 would be the best of the three sires, followed by sire 3.

5.6.2 The multi-trait across-country evaluations (MACE)

The sire model for MACE was originally proposed by Schaeffer (1994) and involved the analysis of the DYD of bulls in different countries as different traits, with the number of daughters of a bull used as a weighting factor. The genetic correlations among DYDs of bulls in different countries were incorporated. The genetic correlations accounted for genotype by environment (G × E) interactions and differences in national models for genetic evaluations among the countries. The genetic correlation among several countries used by Interbull is usually of medium to high value.

However, due to the inability of some countries to compute DYDs for bulls, the deregressed proofs (DRP) of bulls became the variable of choice (Sigurdsson and Banos, 1995) and the weighting factor became the effective daughter contributions (EDC) of bulls (Fiske and Banos, 2001). The model in matrix notation is:

$$\mathbf{y}_i = 1\boldsymbol{\mu}_i + \mathbf{Z}_i\mathbf{Q}\mathbf{w}_i + \mathbf{Z}_i\mathbf{a}_i + \mathbf{e}_i \qquad [5.21]$$

where $\mathbf{y}_i$ is the vector of DRP from country i for one trait, such as milk yield, $\boldsymbol{\mu}_i$ is a mean effect for country i, which reflects the definition of the genetic base for that country, $\mathbf{w}_i$ is the vector of genetic group effects of phantom parents, $\mathbf{a}_i$ is the vector random sire proof for country i and $\mathbf{e}_i$ is the vector of random mean residuals. The matrix $\mathbf{Q}_i$ relates sires to phantom groups (see Section 3.5) and $\mathbf{Z}_i$ relates DRP to sires. Given two countries, the variance–covariance matrix for $\mathbf{w}$, $\mathbf{s}$ and $\mathbf{e}$ is:

$$\text{var}\begin{pmatrix}\mathbf{w}_1\\ \mathbf{w}_2\\ \mathbf{s}_1\\ \mathbf{s}_2\\ \mathbf{e}_1\\ \mathbf{e}_2\end{pmatrix} = \begin{pmatrix}\mathbf{A}_{pp}g_{11} & \mathbf{A}_{pp}g_{12} & \mathbf{A}_{pn}g_{11} & \mathbf{A}_{pn}g_{12} & 0 & 0\\ \mathbf{A}_{pp}g_{21} & \mathbf{A}_{pp}g_{22} & \mathbf{A}_{pn}g_{21} & \mathbf{A}_{pn}g_{22} & 0 & 0\\ \mathbf{A}_{np}g_{11} & \mathbf{A}_{np}g_{12} & \mathbf{A}_{nn}g_{11} & \mathbf{A}_{nn}g_{12} & 0 & 0\\ \mathbf{A}_{np}g_{21} & \mathbf{A}_{np}g_{22} & \mathbf{A}_{nn}g_{21} & \mathbf{A}_{nn}g_{22} & 0 & 0\\ 0 & 0 & 0 & 0 & \mathbf{D}_1\sigma^2_{e1} & 0\\ 0 & 0 & 0 & 0 & 0 & \mathbf{D}_2\sigma^2_{e2}\end{pmatrix}$$

where n and p are the number of bulls and groups, respectively, g_{ij} is the sire genetic (co)variance between countries i and j, and $\mathbf{A}$ is the additive genetic relationship for all bulls and phantom parent groups based on sire–maternal grandsire (MGS) relationships (see Section 3.5), σ^2_{ei} is the residual variance for country i, and $\mathbf{D}_i$ is a diagonal matrix containing the reciprocal of the effective daughter contribution of the bull in the ith country.

The variable DRP, analysed in [5.21], are obtained by deregressing the national breeding values of bulls such that they are independent of all country group effects and additive genetic relationships among bulls, their sires and paternal grandsires, which are included in the MACE analysis (Sigurdsson and Banos, 1995). DRP may therefore contain additive

genetic contributions from the maternal pedigree, which are included at the national level but not in MACE. The deregression procedure involves solving the MME associated with [5.21] for the right-hand-side vector. The details of the procedure are outlined in Appendix F. The computation of the EDC of bulls used as the weighting factor for the analysis of DRP in [5.21] is dealt with in a subsequent section.

The MME for the above model, which are modified such that sire solutions have group solutions incorporated (see Section 3.5), are:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & 0 \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{A}^{-1}\otimes\mathbf{G}^{-1} & -\mathbf{A}^{-1}\mathbf{Q}\otimes\mathbf{G}^{-1} \\ 0 & -\mathbf{Q}'\mathbf{A}^{-1}\otimes\mathbf{G}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}\otimes\mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{c}} \\ \mathbf{Q}\hat{\mathbf{w}}+\hat{\mathbf{a}} \\ \hat{\mathbf{w}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ 0 \end{pmatrix} \quad [5.22]$$

Genetic groups are defined for unknown sires and MGS on the basis of country of origin and year of birth of their progeny. Also maternal grandams (MGD) are always assumed unknown and assigned to phantom groups on the same basis.

Then $\mathbf{A}^{-1}$ can be obtained by the rules outlined in Chapter 2, Section 2.4, which can be briefly summarized in the table below, taking into account the contribution to the groups for MGD. Given a list of pedigrees with the *i*th line consisting of a bull, its sire or group, its MGS or group and a group for its MGD, then contributions to $\mathbf{A}^{-1}$ are as follows:

	Bull	Sire	MGS	MGD
Bull	*d*	–0.5*d*	–0.25*d*	–0.25*d*
Sire	–0.5*d*	0.25*d*	0.125*d*	0.125*d*
MGS	–0.25*d*	0.125d	0.0625*d*	0.0625*d*
MGD	–0.25*d*	0.125*d*	0.0626*d*	0.0625*d*

where $d = 16/(11 + m)$ and $m = 0$ if both sire and MGS are known, $m = 1$ if the sire is known but MGS is unknown, $m = 4$ if the sire is unknown and the MGS is known, and $m = 5$ if both sire and MGS are unknown.

Usually there are dependencies among group effect equations and 1 is added to the diagonals of the phantom group effects in the inverse of the relationship matrix to overcome these dependencies. Then the group solutions sum to zero; therefore the solutions for bulls are relative to the same genetic base within each country. The addition of 1 to the diagonals of the phantom groups implies that group effects are random, with expected values of zero. Since group effects represent differences in the effects of previous selection, which should not have expected values of zero, Schaeffer (1994) indicated that this approach could also be regarded as biased estimation of the fixed effects of phantom groups: that is, a small

amount of bias in the estimates of the phantom groups is accepted in exchange for the hope of getting estimates with smaller mean square errors.

Computing effective daughter contribution

The use of EDC instead of the number of daughters as a weighting factor was proposed by Fiske and Banos (2001) from a simulation study in which they demonstrated that using the numbers of daughters resulted in biased estimates of sire variances used in MACE and international reliabilities. The computation of EDC for a bull accounts for such factors as contemporary group (CG) structure for the bull's daughters, the correlation between observations on the same daughter and the reliability of the performance of the daughters' dams. Thus the EDC provides a measure of the precision of the daughter information used to compute the deregressed proof of the bull. The formula for the computation of EDC (Fiske and Banos, 2001), which included the performance of the dam of the daughter k of bull i, is:

$$\mathrm{EDC}_i = \sum_k \frac{\lambda rel_{k(o)}}{4 - rel_{k(o)}(1 + rel_{dam(o)})}$$

where the summation is over all the k daughters of the bull, $\lambda = (4 - h^2)/h^2$ with h^2 being the heritability, $rel_{dam(o)}$ is the reliability of the dam's own performance, $rel_{k(o)}$ is the reliability of the animal k's own performance, computed as:

$$rel_{k(o)} = \frac{n_k h^2}{1 + (n_k - 1)r}$$

with r being the reliability of the animal's records, n_k the number of lactations of the daughter k of the sire adjusted for the CG size, computed as:

$$n_k = \sum_l 1 - 1/n_{jkl}$$

where n_{jkl} is the size of the CG_j in which the daughter k of sire i made her lth lactation.

An example of MACE for two countries

Example 5.7

The data set below consists of bull breeding values (kg) and deregressed proofs for fat yield for six bulls from two countries. Two of the bulls have evaluations in both countries and in addition each country had two other bulls which were the only progeny tested in that country. A MACE is implemented using the data set. Assume residual variances of 206.5 kg^2

and 148.5 kg^2 for countries 1 and 2, respectively, with corresponding sire additive genetic variances of 20.5 kg^2 and 9.5 kg^2. The sire genetic covariance between fat yield in both countries was assumed to be 12.839 kg, giving a genetic correlation of 0.93. The computation of the DRP in the table below is shown in Appendix F.

	Country 1			Country 2		
Sire	EDC	BV	DRP	EDC	BV	DRP
1	58	9.0	9.7229	90	13.5	14.5088
2	150	10.1	9.9717	65	7.6	7.7594
3	20	15.8	19.2651	–	–	–
4	25	–4.7	–8.5711	–	–	–
5	–	–	–	30	19.6	23.9672
6	–	–	–	55	–5.3	–9.6226

EDC, effective daughter contribution; BV, breeding value; DRP, deregressed proof.

Assume that the sires in the data set have the following pedigree structure, with unknown sires, MGS and MGD assigned to group G_i, with $i = 1,\ldots, 5$.

Bull	Sire	MGS	MGD
1	7	G3	G5
2	8	9	G5
3	7	2	G5
4	1	G2	G5
5	8	G3	G4
6	1	9	G4
7	G1	G2	G4
8	G1	G2	G4
9	G1	G2	G4

Computing sire breeding values

The matrix $\mathbf{G}^{-1}$ for Example 5.7 is:

$$\mathbf{G}^{-1} = \begin{pmatrix} 0.31762 & -0.42925 \\ -0.42925 & 0.68539 \end{pmatrix}$$

The inverses of the matrix of residual variances for countries 1 and 2 are:

$$\mathbf{R}_1^{-1} = \text{diag}(0.2809, 0.7264, 0.0969, 0.1211, 0, 0)$$

and:

$$\mathbf{R}_2^{-1} = \text{diag}(0.6061, 0.4377, 0, 0, 0.2020, 0.3704)$$

The design matrix $\mathbf{X}$ in [5.7] is:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

and:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{pmatrix} 1.2252 & 0 \\ 0 & 1.6162 \end{pmatrix}$$

The matrix $\mathbf{Z}$ is an identity matrix of order 12, considering only bulls with evaluations. The matrix $\mathbf{A}^{-1}$ is set up using the rules outlined earlier. The remaining matrices in [5.22] could be obtained through matrix multiplication and addition. The MME are of the order of 30 by 30 and have not been shown. Solutions to the MME by direct inversion gave the following results:

	Solutions			
Effects	Country 1		Country 2	
Country effect				
	7.268		9.036	
Animal/group				
	A	B	A	B
1	2.604	9.871	2.661	11.697
2	2.176	9.444	0.403	9.439
3	8.059	15.327	5.001	14.037
4	–9.865	–2.597	–5.605	3.431
5	13.634	20.902	9.728	18.764
6	–18.086	–10.818	–13.203	–4.167
7	4.310	11.578	3.071	12.106
8	7.015	14.283	4.489	13.525
9	–6.299	0.969	–5.059	3.977
G1	0.174	7.442	–0.092	8.944
G2	–0.124	7.144	0.126	9.162
G3	–0.071	7.197	0.264	9.300
G4	0.087	7.355	–0.288	8.748
G5	–0.067	7.201	–0.010	9.026

A = solutions for animals and groups from the MME; B = solutions for animals and groups expressed in each country scale.

The solutions for animals and groups were expressed in each country scale by adding the solution for country effects for country *i* to the animal and group solutions of the *i*th country. As indicated earlier, the sum of the

group solutions is zero. In the next section, some of the bull solutions are partitioned to contributions from various sources to gain a better understanding of MACE.

Equations for partitioning bull evaluations from MACE

The equations for sire proofs from [5.22] are:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1})\hat{\mathbf{a}} = (\mathbf{A}^{-1}\mathbf{Q} \otimes \mathbf{G}^{-1})\hat{\mathbf{g}} + \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{c}}) \quad [5.23]$$

where:

$$\hat{\mathbf{a}} = \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{s}}$$

Thus equation [5.23] can be expressed as:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1})\,\hat{\mathbf{a}} = (\mathbf{A}^{-1}\mathbf{Q} \otimes \mathbf{G}^{-1})\hat{\mathbf{g}} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{CD}) \quad [5.24]$$

where:

$$\mathbf{CD} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{c}}))$$

CD (country deviation) is simply a vector of weighted average of corrected DRP in all countries where the bull has a daughter, the weighting factor being the reciprocal of EDC multiplied by the residual variance in each country. Since $\mathbf{R}^{-1}$ is diagonal, **CD** is equal to the vector $(\mathbf{y} - \mathbf{X}\hat{\mathbf{c}})$.

For a particular bull with a direct progeny (e.g. son), equation [5.24] can be written as:

$$\begin{aligned}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})\hat{\mathbf{a}}_{bull} &= \mathbf{G}^{-1}\alpha_{par}(\hat{\mathbf{a}}_{sire} + 0.5(\hat{\mathbf{a}}_{mgs} + \hat{\mathbf{g}})) + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{CD}) \\ &\quad + \mathbf{G}^{-1}\sum \alpha_{prog}(\hat{\mathbf{a}}_{prog} - 0.25\hat{\mathbf{a}}_{mate}) \quad [5.25]\end{aligned}$$

where $\alpha_{par} = \frac{8}{11}$, $\frac{8}{15}$, $\frac{2}{3}$ or $\frac{1}{2}$ if both sire and MGS (maternal grandsire), only mgs, only sire or no parents are known, respectively, and $\alpha_{prog} = \frac{8}{11}$ if the MGS of the bull's mate is known or $\frac{2}{3}$ if unknown. The above values for α_{par} and α_{prog} are based on the assumption that $\mathbf{A}^{-1}$ has been calculated without accounting for inbreeding. Note that in equation [5.25]:

$$\alpha_{bull} = 2\alpha_{par} + 0.5\alpha_{prog}$$

Equation [5.25] can be expressed as:

$$\begin{aligned}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})\hat{\mathbf{a}}_{bull} &= 2\mathbf{G}^{-1}\alpha_{par}(\mathbf{PA}) + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\mathbf{CD} \\ &\quad + 0.5\mathbf{G}^{-1}\sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - 0.5\hat{\mathbf{a}}_{mate})\end{aligned}$$

where:

$$\mathbf{PA} = 0.5\hat{\mathbf{a}}_{sire} + 0.25(\hat{\mathbf{a}}_{mgs} + \hat{\mathbf{g}})$$

Pre-multiplying both sides of the equation by $(\mathbf{Z}'\,\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})^{-1}$ gives:

$$\hat{\mathbf{a}}_{bull} = \mathbf{W}_1\mathbf{PA} + \mathbf{W}_2\mathbf{CD} + \mathbf{W}_3\mathbf{PC} \quad [5.26]$$

where:

$$\mathbf{PC} = \sum \alpha_{prog}(2\hat{\mathbf{a}}_{prog} - 0.5\hat{\mathbf{a}}_{mate}) \Big/ \sum \alpha_{prog} \quad \text{and} \quad \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3 = \mathbf{I}$$

The matrices $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ are the product of $(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})^{-1}$ and $2\mathbf{G}^{-1}\alpha_{par}$, $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ and $0.5\mathbf{G}^{-1}\Sigma\alpha_{prog}$, respectively. Using equation [5.26], the contributions from different sources of information from different countries to the MACE of a bull can be computed.

If the progeny in equation [5.25] is not a direct progeny of the bull but a maternal grandson of the bull, then α_{prog} equals $\frac{4}{11}$ if mate (sire) is known or $\frac{4}{15}$ if unknown. Then equation [5.25] becomes:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})\hat{\mathbf{a}}_{bull} = \mathbf{G}^{-1}\alpha_{par}(\hat{\mathbf{a}}_{sire} + 0.5(\hat{\mathbf{a}}_{mgs} + \hat{\mathbf{g}})) + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{CD}) + \mathbf{G}^{-1}\sum \alpha_{prog}(\hat{\mathbf{a}}_{prog} - 0.5\hat{\mathbf{u}}_{mate})$$

and α_{bull} now equals $2\alpha_{par} + 0.25\alpha_{prog}$ and $0.5\hat{\mathbf{a}}_{mate} = 0.5\hat{\mathbf{a}}_s$, the sire of the progeny. The above can be expressed as:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})\hat{\mathbf{a}}_{bull} = 2\mathbf{G}^{-1}_{\alpha_{par}}(\mathbf{PA}) + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z})\mathbf{CD} + 0.25\mathbf{G}^{-1}\sum \alpha_{prog}(4\hat{\mathbf{a}}_{prog} - 2\hat{\mathbf{a}}_{mate})$$

Pre-multiplying both sides by $(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})^{-1}$ gives the same equation as equation [5.26] but with:

$$\mathbf{PC} = \sum \alpha_{prog}(4\hat{\mathbf{a}}_{prog} - 2\hat{\mathbf{a}}_{mate}) \Big/ \sum \alpha_{prog} \qquad [5.27]$$

and now:

$$\mathbf{W}_3 = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull})^{-1}(0.25\mathbf{G}^{-1}\sum \alpha_{prog})$$

The use of [5.26] to partition proofs from MACE is illustrated for two bulls, one with no progeny and another with a maternal grandson. First, consider bull 3 in Example 5.7 that has deregressed proofs only in country 1 and has no progeny. Therefore $\mathbf{CD}_{3i}$ for bull 3 in country i is:

$$\mathbf{CD}_{31} = y_{31} - \mu_1 = 19.2651 - 7.261 = 12.0041 \quad \text{and} \quad \mathbf{CD}_{32} = 0$$

Parent average for bull 3 ($\mathbf{PA}_{3i}$) in country i is:

$$\mathbf{PA}_{31} = 0.5(\hat{\mathbf{a}}_{71}) + 0.25(\hat{\mathbf{a}}_{21} + \hat{\mathbf{g}}_{G51}) = 0.5(4.310) + 0.25(2.176 + (-0.067)) = 2.68225$$

and:

$$\mathbf{PA}_{32} = 0.5(\hat{\mathbf{a}}_{72}) + 0.25(\hat{\mathbf{a}}_{22} + \hat{\mathbf{g}}_{G52}) = 0.5(3.071) + 0.25(0.403 + (-0.010)) = 1.63375$$

where $\hat{\mathbf{a}}_{ji}$ is the breeding value of animal j in country i and $\hat{\mathbf{g}}_{Gji}$ is the solution for group j and in the ith country.

The residual variance for bull 3 in country 1 $(\mathbf{r}_{31}) = (\frac{1}{20})206.5 = 10.325$ and its inverse equals 0.09685. Both sire and MGS of bull 3 are known;

therefore $\alpha_{bull} = \frac{16}{11}$. Then:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull}) = \begin{pmatrix} 0.09685 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0.4620 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}$$

$$= \begin{pmatrix} 0.55884 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}$$

The matrices of weights ($\mathbf{W}_i$) using [5.26] are:

$$\mathbf{W}_1 = \begin{pmatrix} 0.55884 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}^{-1} \begin{pmatrix} 0.4620 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}$$

$$= \begin{pmatrix} 0.4229 & 0 \\ -0.3614 & 1.0000 \end{pmatrix}$$

and:

$$\mathbf{W}_2 = \begin{pmatrix} 0.55884 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}^{-1} \begin{pmatrix} 0.09685 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0.5771 & 0 \\ 0.3614 & 0 \end{pmatrix}$$

Therefore the vector proofs of bull 3 are:

$$\begin{pmatrix} \hat{\mathbf{a}}_{31} \\ \hat{\mathbf{a}}_{32} \end{pmatrix} = \mathbf{W}_1 \begin{pmatrix} 2.68225 \\ 1.63375 \end{pmatrix} + \mathbf{W}_2 \begin{pmatrix} 11.9971 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.1343 \\ 0.6644 \end{pmatrix} + \begin{pmatrix} 6.9235 \\ 4.3358 \end{pmatrix} = \begin{pmatrix} 8.058 \\ 5.000 \end{pmatrix}$$

The contribution from the DRP of bull 3 in country 1 accounts for over 85% of the MACE proof in both countries, although the bull has no DRP in country 2. Thus, with only 20 daughters, parental contribution was not very large, although, in general, parental contributions will be influenced by the heritability of the traits in both countries and the genetic correlation between them.

When a bull has a proof only in country *i* and not in *j*, its proof in country *j* can be obtained (Mrode and Swanson, 1999) as:

$$\hat{\mathbf{a}}_j = \mathbf{PA}_j + (g_{ij}/g_{ii})(\hat{\mathbf{a}}_i - \mathbf{PA}_i) \qquad [5.28]$$

where g_{ii} is the genetic variance in country *i* and g_{ij} the genetic covariance between countries *i* and *j*. Therefore, if interest was only in calculating the proof of bull 3 in country 2, it can be obtained from the above equation as:

$$\hat{\mathbf{a}}_{32} = 1.63375 + (12.839/20.5)(8.059 - 2.68225) = 5.001$$

Equation [5.28] can be derived from [5.25] as follows. The equation for $\hat{\mathbf{a}}_{32}$ from [5.25] is:

$$(g^{22}\alpha_{bull})\hat{\mathbf{a}}_{32} = g^{22}\alpha_{par}(\hat{\mathbf{a}}_{sire\,2} + 0.5(\hat{\mathbf{a}}_{mgs2} + \hat{\mathbf{g}}_{mgd2}) + g^{21}\alpha_{par}(\hat{\mathbf{a}}_{sire\,1} + 0.5(\hat{\mathbf{a}}_{mgs1} + \hat{\mathbf{g}}_{mgd1}) + (g^{21}\alpha_{bull})\hat{\mathbf{a}}_{31}$$

where $\hat{\mathbf{a}}_{sirej}$, $\hat{\mathbf{a}}_{mgsi}$ and $\hat{\mathbf{g}}_{mgdj}$ are the proofs for the sire, MGS and solution for the MGD in country *j*, respectively, and g^{ii} are the inverse elements

of $\mathbf{G}^{-1}$. Since $\alpha_{bull} = 2\alpha_{par}$ for bull 3, multiplying the above equation by $(2\alpha_{par})^{-1}$ gives:

$$g^{22}\hat{\mathbf{a}}_{32} = g^{22}(\mathbf{PA}_2) + g^{21}(\mathbf{PA}_1) - g^{21}\hat{\mathbf{a}}_{31}$$
$$g^{22}\hat{\mathbf{a}}_{32} = g^{22}(\mathbf{PA}_2) + g^{21}(\hat{\mathbf{a}}_{31} - \mathbf{PA}_1)$$
$$\hat{\mathbf{a}}_{32} = \mathbf{PA}_2 - g^{21}/g^{22}(\hat{\mathbf{a}}_{31} - \mathbf{PA}_1)$$
$$\hat{\mathbf{a}}_{32} = \mathbf{PA}_2 + g_{21}/g_{22}(\hat{\mathbf{a}}_{31} - \mathbf{PA}_1)$$

Thus the proof of a bull in country *j* is dependent on the parent average of the bull in country *j* and the Mendelian sampling of the bull in the *i*th country.

Partitioning the proof of bull 2 with records in both countries and a maternal grandson (bull 3) is as follows. The country deviations for bull 2 in both countries are:

$$\mathbf{CD}_{21} = y_{21} - \mu_1 = 9.9717 - 7.268 = 2.7037$$

and:

$$\mathbf{CD}_{22} = y_{22} - \mu_2 = 7.7594 - 9.036 = -1.2766$$

Parent average for sire 2 (PA_{2i}) for country *i* is:

$$\mathbf{PA}_{21} = 0.5(\hat{\mathbf{a}}_{81}) + 0.25(\hat{\mathbf{a}}_{91} + \hat{\mathbf{g}}_{G51}) = 0.5(7.015) + 0.25(-6.299 + (-0.067)) = 1.916$$

$$\mathbf{PA}_{22} = 0.5(\hat{\mathbf{a}}_{82}) + 0.25(\hat{\mathbf{a}}_{92} + \hat{\mathbf{g}}_{G52}) = 0.5(4.489) + 0.25(-5.059 + (-0.010)) = 0.97725$$

Progeny contributions (PC) from bull 3 to sire 2 (PC_{2i}) in country *i* are:

$$PC_{21} = 4(\hat{\mathbf{a}}_{31}) - 2(\hat{\mathbf{a}}_{71}) = 4(8.059) - 2(4.310) = 23.616$$

$$PC_{22} = 4(\hat{\mathbf{a}}_{32}) - 2(\hat{\mathbf{a}}_{72}) = 4(5.001) - 2(3.071) = 13.862$$

The residual variance for bull 2 in country 1, $(r_{21}) = (\frac{1}{150})206.5$ and in country 2, $(r_{22}) = (\frac{1}{65})148.5$. Corresponding inverses were 0.72639 and 0.43771, respectively. Since both sire and MGS of bull 2 are known and he has a maternal grandson, $\alpha_{bull} = 2\alpha_{par} + 0.25\alpha_{prog} = 2(\frac{8}{11}) + 0.25(\frac{4}{11}) = 1.54545$. Therefore:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\alpha_{bull}) = \begin{pmatrix} 0.72630 & 0 \\ 0 & 0.43771 \end{pmatrix} + \begin{pmatrix} 0.49087 & -0.66338 \\ -0.66338 & 1.05924 \end{pmatrix}$$
$$= \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}$$

From [5.26], the matrices of weights ($\mathbf{W}_i$) are:

$$\mathbf{W}_1 = \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}^{-1} \begin{pmatrix} 0.4620 & -0.62436 \\ -0.62436 & 0.99693 \end{pmatrix}$$
$$= \begin{pmatrix} 0.2007 & -0.1977 \\ -0.3281 & 0.5783 \end{pmatrix}$$

$$\mathbf{W}_2 = \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}^{-1} \begin{pmatrix} 0.72639 & 0 \\ 0 & 0.43771 \end{pmatrix} = \begin{pmatrix} 0.7867 & 0.2101 \\ 0.3487 & 0.3855 \end{pmatrix}$$

$$\mathbf{W}_3 = \begin{pmatrix} 1.21726 & -0.66338 \\ -0.66338 & 1.49695 \end{pmatrix}^{-1} \begin{pmatrix} 0.02887 & -0.03902 \\ -0.03902 & 0.06231 \end{pmatrix}$$

$$= \begin{pmatrix} 0.0125 & -0.0124 \\ -0.02051 & 0.0361 \end{pmatrix}$$

The vector of proof for bull 2 is:

$$\begin{pmatrix} \hat{\mathbf{a}}_{21} \\ \hat{\mathbf{a}}_{22} \end{pmatrix} = \mathbf{W}_1 \begin{pmatrix} 1.9160 \\ 0.9773 \end{pmatrix} + \mathbf{W}_2 \begin{pmatrix} 2.7037 \\ -1.2766 \end{pmatrix} + \mathbf{W}_3 \begin{pmatrix} 23.616 \\ 13.862 \end{pmatrix} = \begin{pmatrix} 2.176 \\ 0.403 \end{pmatrix}$$

Again, similarly to bull 3 above, the contributions from the DRP in both countries accounted for much of the MACE proofs of bull 2 in countries 1 and 2.

Some limitations of MACE

One of the limitations of MACE involves the use of deregressed breeding values; therefore different data editing procedures and genetic evaluation models among countries could lead to an artificial interaction between the breeding values of sires and therefore less than unity genetic correlations among countries, even though the expression of the underlying biological trait is identical. The definition of traits according to country borders rather than actual climatic differences has been regarded as another weakness of MACE. This implies that MACE is unable to account for G × E between herds due to differences in climatic conditions and management practices in large countries such as the USA or Australia. On the other hand, significant G × E may not exist between small neighbouring countries, such as The Netherlands and Belgium, but current MACE procedure assumes so by defining traits according to country boundaries.

Efforts to address these problems are currently only at the research stage. In an effort to overcome the first limitation, Weigel *et al.* (2001) analysed actual test day records of first lactation milk, fat and protein yields in Holstein-sired cows from 17 countries with a multiple-trait sire BLUP. The use of test day records avoided the problems of differences in editing procedures among countries. In general, the genetic correlations (r_g) from their study between countries with high genetic level, intensive management and adequate genetic ties (e.g. the USA, Canada, France, Germany and the UK) were similar to r_g calculated by Interbull using deregressed breeding values. The estimates of r_g tended to be lower than Interbull estimates between countries with weak genetic ties (Estonia, Ireland, Israel) but higher between Ireland and both Australia and New Zealand. The concept of borderless evaluation, which involves standardized data collection and evaluation across national borders, was introduced by Lohuis

and Dekkers (1998) to address the second limitation. They suggested that borderless evaluations could be based on similar herds grouped together across countries on the basis of similar production environments. Weigel and Rekaya (2000) examined a multiple-trait herd cluster model for such international borderless genetic evaluation. The estimates of r_g among clusters were generally in the same range (0.84 to 0.94) as those estimated among countries by Interbull.

6 Maternal Trait Models: Animal and Reduced Animal Models

The phenotypic expression of some traits in the progeny, such as weaning weight in beef cattle, is influenced by the ability of the dam to provide a suitable environment in the form of better nourishment. Thus the dam contributes to the performance of the progeny in two ways: first, through her direct genetic effects passed to the progeny and, secondly, through her ability to provide a suitable environment, for instance in producing milk. Traits such as birth and weaning weights in beef cattle fall into this category and are termed maternally influenced traits. The ability of the dam to provide a suitable environment for the expression of such traits in her progeny is partly genetic and partly environmental. Similarly to the genetic component of an individual, the maternal genetic component can be partitioned into additive, dominance and epistatic effects (Willham, 1963). The environmental part may be partitioned into permanent and temporary environmental components. It is the maternal additive genetic component of the dam that is passed on to all her offspring but it is expressed only when the female offspring have progeny of their own.

In the usual mixed linear model for maternally influenced traits [6.1], the phenotype is partitioned into the following:

1. Additive genetic effects from the sire and the dam, usually termed the direct genetic effect.
2. Additive genetic ability of the dam to provide a suitable environment, usually termed the indirect or maternal genetic effect.
3. Permanent environmental effects, which include permanent environmental influences on the dam's mothering ability and the maternal non-additive genetic effects of the dam.
4. Other random environmental effects, termed residual effects.

In this chapter, the mixed model methodology for genetic evaluation in models with maternal effects is discussed considering a univariate situation,

and the extension to multivariate analysis is also briefly presented. The application of best linear unbiased prediction (BLUP) to models with maternal effects was first presented by Quaas and Pollak (1980).

When repeated measurements for maternally influenced traits are available over a range of ages (for instance, body weight from birth to 630 days), a random regression model (see Chapter 7) might be more appropriate to analyse such a trait. A random regression model for maternally influenced traits is briefly defined in Section 7.2.5 of Chapter 7.

6.1 Animal Model for a Maternal Trait

The model for maternally influenced traits in matrix notation is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wm} + \mathbf{Spe} + \mathbf{e} \qquad [6.1]$$

where $\mathbf{y}$ = vector of observations, $\mathbf{b}$ = vector of fixed effects, $\mathbf{u}$ = vector of random animal effects, $\mathbf{m}$ = vector of random maternal (indirect) genetic effects, $\mathbf{pe}$ = vector of permanent environmental effects, as explained in **3** above, $\mathbf{e}$ = vector of random residual effects, and $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$ and $\mathbf{S}$ are incidence matrices relating records to fixed, animal, maternal genetic and permanent environmental effects, respectively.

It is assumed that:

$$\text{var}\begin{bmatrix}\mathbf{u}\\ \mathbf{m}\\ \mathbf{pe}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} & 0 & 0\\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & 0 & 0\\ 0 & 0 & \mathbf{I}\sigma^2_{pe} & 0\\ 0 & 0 & 0 & \mathbf{I}\sigma^2_{e}\end{bmatrix}$$

where g_{11} = additive genetic variance for direct effects, g_{22} = additive genetic variance for maternal effects, g_{12} = additive genetic covariance between direct and maternal effects, σ^2_{pe} = variance due to permanent environmental effects, and σ^2_e = residual error variance.

The variance of $\mathbf{y}$, using the same arguments as in Section 3.1, is:

$$\text{var}(\mathbf{y}) = [\mathbf{Z} \quad \mathbf{W}]\begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A}\\ g_{21}\mathbf{A} & g_{22}\mathbf{A}\end{bmatrix}\begin{bmatrix}\mathbf{Z}'\\ \mathbf{W}'\end{bmatrix} + \mathbf{SI}\,\sigma^2_{pe}\mathbf{S}' + \mathbf{I}\,\sigma^2_e$$

The best linear unbiased estimator (BLUE) of estimable functions of $\mathbf{b}$ and the BLUP of $\mathbf{u}$, $\mathbf{m}$ and $\mathbf{pe}$ in [6.1] are obtained by solving the following mixed model equations (MME):

$$\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{u}}\\ \hat{\mathbf{m}}\\ \hat{\mathbf{pe}}\end{bmatrix}\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} & \mathbf{X}'\mathbf{S}\\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\mathbf{A}^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W}+\mathbf{A}^{-1}\alpha_2 & \mathbf{Z}'\mathbf{S}\\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z}+\mathbf{A}^{-1}\alpha_2 & \mathbf{W}'\mathbf{W}+\mathbf{A}^{-1}\alpha_3 & \mathbf{W}'\mathbf{S}\\ \mathbf{S}'\mathbf{X} & \mathbf{S}'\mathbf{Z} & \mathbf{S}'\mathbf{W} & \mathbf{S}'\mathbf{S}+\mathbf{I}\,\alpha_4\end{bmatrix} = \begin{bmatrix}\mathbf{X}'\mathbf{y}\\ \mathbf{Z}'\mathbf{y}\\ \mathbf{W}'\mathbf{y}\\ \mathbf{S}'\mathbf{y}\end{bmatrix} \qquad [6.2]$$

with:

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}; \quad \mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} = \sigma_e^2 \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}$$

$$\alpha_4 = \sigma_e^2 / \sigma_{pe}^2$$

6.1.1 An illustration

Example 6.1

Assume the data in Table 6.1 to be the birth weight for a group of beef calves. The aim is to estimate solutions for herd and sex effects and predict solutions for direct and maternal effects for all animals and permanent environmental effects for dams of progeny with records. Suppose that the genetic parameters are $g_{11} = 150$, $g_{12} = -40$, $g_{22} = 90$, $\sigma_{pe}^2 = 40$ and $\sigma_e^2 = 350$. Then:

$$\mathbf{G}^{-1} = \begin{bmatrix} 0.00756 & 0.00336 \\ 0.00336 & 0.0126 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} = \begin{bmatrix} 2.647 & 1.176 \\ 1.176 & 4.412 \end{bmatrix}$$

and $\alpha_4 = 350/40 = 8.75$.

The model for the analysis is as presented in equation [6.1].

Setting up design matrices

Considering only animals with records, the first three rows of matrix **X** relate records to herd effects and the last two rows to sex effects. The transpose of **X** is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Table 6.1. Birth weight for a group of beef calves.

Calf	Sire	Dam	Herd	Sex[a]	Birth weight
5	1	2	1	1	35.0
6	3	2	1	2	20.0
7	4	6	1	2	25.0
8	3	5	1	1	40.0
9	1	6	2	1	42.0
10	3	2	2	2	22.0
11	3	7	2	2	35.0
12	8	7	3	2	34.0
13	9	2	3	1	20.0
14	3	6	3	2	40.0

[a] 1 = male and 2 = female (throughout chapter)

Excluding ancestors, each animal has one record; therefore **Z** is an identity matrix. However, **Z** is augmented with columns of zeros equal to the number of ancestors to take account of ancestors in the pedigree. The matrices **W** and **S** relate records through the dam to their effects – that is, maternal genetic effect and permanent environmental effect, respectively. However, since maternal effect is genetic and is passed from parent to offspring, estimates of maternal effect are for all animals in the analysis while estimates of permanent environmental effects are only for dams of progeny with records. Thus, in setting up **W**, all animals are considered, while only four dams with progeny having records are taken into account for **S**. For the example data set, **W** (with rows and columns numbered by the relevant animal they relate to) is:

$$\mathbf{W} = \begin{array}{c} \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \end{array} \begin{array}{c} \begin{array}{cccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

and:

$$\mathbf{S}' = \begin{array}{c} \\ 2 \\ 5 \\ 6 \\ 7 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

The matrix **S** above implies, for instance, that animals 5, 6, 10 and 13 have the same dam (animal 2), while animals 11 and 12 are from another dam (animal 7).

The transpose of the vector of observations is:

$$\mathbf{y}' = [35 \;\; 20 \;\; 25 \;\; 40 \;\; 42 \;\; 22 \;\; 35 \;\; 34 \;\; 20 \;\; 40]$$

The other matrices in the MME can be calculated through matrix multiplication. The inverse of the relationship matrix is calculated applying the rules in Section 2.3.1. The matrix $\mathbf{A}^{-1}\alpha_1$ is added to animal equations, $\mathbf{A}^{-1}\alpha_2$ to the equations for maternal genetic effects, $\mathbf{A}^{-1}\alpha_3$ to the animal by maternal genetic equations and maternal genetic by animal equations and α_4 to the diagonals of the equations for permanent environmental effects to obtain the MME. The MME are not presented because they are too large.

There is dependency between the equations for herds and sex; thus the row for the first herd was set to zero in solving the MME by direct inversion. Solutions to the MME are as follows:

Effects	Solutions	
Herd–year–season		
1	0.000	
2	3.386	
3	1.434	
Sex of calf		
1	34.540	
2	27.691	
Animal and maternal		
1	0.564[a]	0.262[b]
2	–1.244	–1.583
3	1.165	0.736
4	–0.484	0.586
5	0.630	–0.507
6	–0.859	0.841
7	–1.156	1.299
8	1.917	–0.158
9	–0.553	0.660
10	–1.055	–0.153
11	0.385	0.916
12	0.863	0.442
13	–2.980	0.093
14	1.751	0.362
Permanent environment		
2	–1.701	
5	0.415	
6	0.825	
7	0.461	

[a] Solutions for direct animal effects.
[b] Solutions for maternal genetic effects.

The solutions show little difference between the herds but the males are heavier than females by about 6.85 kg at birth. The solution for level i of the fixed effect n can be calculated using [4.3] except that the sum of yields for the level of fixed effect is corrected in addition for maternal effects. That is

$$\hat{b}_{in} = \frac{\sum_{f=1}^{\text{diag}_{in}} y_{inf} - \sum_{j} \hat{b}_{inj} - \sum_{k} \hat{a}_{ink} - \sum_{l} \hat{m}_{inl} - \sum_{t} \hat{p}e_{int}}{\text{diag}_{in}} \qquad [6.3]$$

where $\hat{m}_{inl}$ is the solution for level l of genetic maternal effects within level i of the nth fixed effect and all other terms are as defined in [4.3]. Thus the solution for level l of sex effect is:

$$\begin{aligned}\hat{b}_{11} &= [137 - (2\hat{h}d_1 + \hat{h}d_2 + \hat{h}d_3) - (\hat{a}_5 + \hat{a}_8 + \hat{a}_9 + \hat{a}_{13}) \\ &\quad - (2\hat{m}_2 + \hat{m}_5 + \hat{m}_6) - (2\hat{p}e_2 + \hat{p}e_5 + \hat{p}e_6)]/4 \\ &= [137 - 4.82 - (-0.986) - (-2.832) - (-2.162)]/4 \\ &= 34.540\end{aligned}$$

where $\hat{h}d_j$ is the solution for level j of herd effect.

From the MME, the solutions for direct and maternal effects for animal i with progeny o are:

$$\begin{bmatrix}\hat{u}_i \\ \hat{m}_i\end{bmatrix} = \begin{bmatrix}n_1 + (d + k_1)\alpha_1 & (d + k_1)\alpha_2 \\ (d + k_1)\alpha_2 & n_2 + (d + k_1)\alpha_3\end{bmatrix}^{-1} \mathbf{H}k_2 \begin{bmatrix}\hat{u}_s + \hat{u}_d \\ \hat{m}_s + \hat{m}_d\end{bmatrix} + \begin{bmatrix}y_i - \hat{b}_i - \hat{m}_{dam} - \hat{p}_{dam} \\ y_o - \hat{b}_o - \hat{u}_o - \hat{p}_i\end{bmatrix} + \mathbf{H}k_3 \begin{bmatrix}\hat{a}_o - 0.5(\hat{a}_{mate}) \\ \hat{m}_o - 0.5(\hat{m}_{mate})\end{bmatrix} \quad [6.4]$$

where n_1 is the number of records for animal i, n_2 is the number of progeny records with animal i as the dam, $d = 2$, $\frac{4}{3}$ or 1 when both, one or no parents of animal i are known, respectively, $k_2 = 1$ or $\frac{2}{3}$ when both or one parent of animal i are known, $k_1 = \frac{1}{2}$ and $k_3 = 1$ when the mate of animal i is known or $k_1 = \frac{1}{3}$ and $k_3 = \frac{2}{3}$ with the mate unknown and:

$$\mathbf{H} = \begin{bmatrix}\alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3\end{bmatrix}$$

For instance, the solutions for direct and genetic maternal effects for animal 5 are:

$$\begin{bmatrix}\hat{u}_5 \\ \hat{m}_5\end{bmatrix} = \begin{bmatrix}1 + (2 + 0.5)2.647 & (2 + 0.5)1.176 \\ (2 + 0.5)1.176 & 1 + (2 + 0.5)4.412\end{bmatrix}^{-1} \mathbf{H}k_2 \begin{bmatrix}u_1 + u_2 \\ m_2 + m_2\end{bmatrix} + \begin{bmatrix}y_5 - \hat{b}_1 - \hat{m}_2 - \hat{p}_2 \\ y_8 - \hat{b}_1 - \hat{u}_8 - \hat{p}_5\end{bmatrix} + \mathbf{H}k_3 \begin{bmatrix}\hat{a}_8 - 0.5(\hat{a}_3) \\ \hat{m}_8 - 0.5(\hat{m}_3)\end{bmatrix}$$

$$\begin{bmatrix}\hat{u}_5 \\ \hat{m}_5\end{bmatrix} = \begin{bmatrix}1 + (2 + 0.5)2.647 & (2 + 0.5)1.176 \\ (2 + 0.5)1.176 & 1 + (2 + 0.5)4.412\end{bmatrix}^{-1} \mathbf{H}(1) \begin{bmatrix}0.564 + (-1.244) \\ 0.262 + -(1.583)\end{bmatrix} + \begin{bmatrix}35 - 0 - 34.54 - (-1.583) - (-1.701) \\ 40 - 0 - 34.54 - 1.917 - 0.415\end{bmatrix} + \mathbf{H}(1) \begin{bmatrix}1.917 - 0.5(1.165) \\ -0.158 - 0.5(0.736)\end{bmatrix} = \begin{bmatrix}0.630 \\ -0.507\end{bmatrix}$$

The solution for the permanent environmental effect for dam j from the MME is:

$$\hat{p}e_j = (y_o - \hat{b}_o - \hat{u}_o - \hat{m}_j)/(n_2 + \alpha_3) \quad [6.5]$$

where all terms are as defined in [6.4]. For animal 5, the solution for the permanent environmental effect is:

$$\hat{pe}_5 = 40 - 0 - 34.54 - 1.917 - (-0.507)/(1 + 8.75) = 0.415$$

Additive genetic maternal effects represent good mothering ability, which is passed on from dams to progeny, while permanent environment effects refer to permanent environmental and maternal non-additive genetic influences on the mothering ability of the dam. Thus selection of dams for the next generation in a maternal line would place emphasis on good genetic maternal effects in addition to a good estimate of breeding value. If equal emphasis is placed on both effects, dams 7 and 5 would be the top two dams in the example while dam 2 ranks lowest. However, if the main interest is the performance of the future dams in the same herd, then selection of dams would be based on some combination of the solutions for direct, maternal genetic and permanent environmental effects for the dams. Again, in the example data, dam 2 ranks lowest while the best two dams are dams 6 and 7 if equal emphasis is placed on the three components.

In the case of males, the selection of sires for a maternal line, for instance, would be based on a combination of solutions for direct and maternal genetic effects. Obviously, sires 3 and 1 would be the top two bulls for such a purpose. However, if the emphasis is only on direct genetic effects, probably to breed a bull, then sire 8 in the example would be the bull of choice.

6.2 Reduced Animal Model with Maternal Effects

In Chapter 3, Section 3.4, the use of the reduced animal model (RAM), with only one random effect apart from residual error in the model, was considered. The records of non-parents in the MME were expressed as the average of parental breeding values plus Mendelian sampling. This has the advantage of reducing the number of random animal equations in the MME. The application of RAM with multiple random effects in the model is illustrated in this section using the example data used for the full animal model in Section 6.1. The model for the analysis is the same but design matrices and the variance of non-parental animals are different. From the arguments in Section 3.4, the model for the RAM can be expressed as:

$$\begin{bmatrix} \mathbf{y}_p \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_p \\ \mathbf{X}_n \end{bmatrix} \mathbf{b} + \begin{bmatrix} \mathbf{Z}_p \\ \mathbf{Z}_n \end{bmatrix} \mathbf{u}_p + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{pe} + \begin{bmatrix} \mathbf{e}_p \\ \mathbf{e}_n \end{bmatrix} \quad [6.6]$$

where $\mathbf{y}_p, \mathbf{y}_n$ = vector of observations for parent and non-parents, respectively, $\mathbf{b}$ = vector of fixed effects, $\mathbf{u}_p$ = vector of random animal

effect for parents, **m** = vector of maternal genetic effects for parents, **pe** = vector of permanent environmental effects, and $\mathbf{e}_p$, $\mathbf{e}_n$ = vector of residual error for parents and non-parents, respectively.

The incidence matrices $\mathbf{Z}_2$ and $\mathbf{Z}_3$ relate records to maternal genetic and permanent environmental effect, respectively. The matrices $\mathbf{Z}_p$ and $\mathbf{X}_p$ relate records of parents to animal and fixed effects, respectively, while $\mathbf{Z}_n$ and $\mathbf{X}_n$ relate records of non-parents to parents (animal effect) and fixed effects, respectively.

It is assumed that:

$$\text{var}\begin{bmatrix}\mathbf{u}_p\\ \mathbf{m}\\ \mathbf{p}e\\ \mathbf{e}_p\\ \mathbf{e}_n\end{bmatrix} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} & 0 & 0 & 0\\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & 0 & 0 & 0\\ 0 & 0 & \mathbf{I}\sigma^2_{pe} & 0 & 0\\ 0 & 0 & 0 & \mathbf{I}\sigma^2_{ep} & 0\\ 0 & 0 & 0 & 0 & \mathbf{I}\sigma^2_{en}\end{bmatrix}$$

where σ^2_{ep} is the residual variance for parents, which is equal to σ^2_e in Section 6.1, σ^2_{en} is the residual variance for non-parents and is equal to $\mathbf{I} + \mathbf{D}g_{11}$, with **D** being a diagonal matrix containing elements d_{jj}, which are equal to $\frac{3}{4}$ or $\frac{1}{2}$ depending on whether one or both parents are known. The matrix **G** and σ^2_{pe} are defined as in Section 6.1. Let:

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{bmatrix}, \quad \mathbf{Z}_1 = \begin{bmatrix}\mathbf{Z}_p\\ \mathbf{Z}_n\end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix}\mathbf{I}\sigma^2_{ep} & 0\\ 0 & \mathbf{I}\sigma^2_{en}\end{bmatrix} = \begin{bmatrix}\mathbf{R}_n & 0\\ 0 & \mathbf{R}_p\end{bmatrix} \text{ and}$$

$$\mathbf{R}^{-1} = \begin{bmatrix}\mathbf{R}_p^{-1} & 0\\ 0 & \mathbf{R}_n^{-1}\end{bmatrix}$$

Again, the MME provide the basis of the BLUE of estimable functions of **b** and BLUP of **a**, **m** and **pe** in [6.6]. The relevant MME are:

$$\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{u}}_p\\ \hat{\mathbf{m}}\\ \hat{\mathbf{p}}\mathbf{e}\end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}X & \mathbf{X}'\mathbf{R}^{-1}Z_1 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_2 & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_3\\ \mathbf{Z}_1'\mathbf{R}^{-1}X & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1' + \mathbf{A}_p^{-1}g^{11} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{A}_p^{-1}g^{12} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_3\\ \mathbf{Z}_2'\mathbf{R}^{-1}X & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_1' + \mathbf{A}_p^{-1}g^{21} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_2 + \mathbf{A}_p^{-1}g^{22} & \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{Z}_3\\ \mathbf{Z}_3'\mathbf{R}^{-1}X & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_1 & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_2 & \mathbf{Z}_3'\mathbf{R}^{-1}\mathbf{Z}_3' + \mathbf{I}1/\sigma^2_{pe}\end{bmatrix}^{-1} \begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}_2'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}_3'\mathbf{R}^{-1}y\end{bmatrix} \qquad [6.7]$$

where g^{ij} are the elements of the inverse of **G**.

As shown in Section 3.4, each block of equations in the MME above can be expressed as the sum of the contributions from parents' records and non-parents' records. Thus:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \mathbf{X}'_p\mathbf{R}_p^{-1}\mathbf{X}_p + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{X}_n$$

Expressing [6.5] as shown for the equations for the block of fixed effects above and multiplying by $\mathbf{R}_p$ gives:

$$\begin{bmatrix} \mathbf{X}'_p\mathbf{X}_p + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{X}'_p\mathbf{Z}_p + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{Z}_n \\ \mathbf{Z}'_p\mathbf{X}_p + \mathbf{Z}'_n\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}'_p\mathbf{Z}_p + \mathbf{Z}'_n\mathbf{R}_n^{-1}\mathbf{Z}_n + \mathbf{A}^{-1}\alpha_1 \\ \mathbf{Z}'_2\mathbf{X}_p + \mathbf{Z}'_2\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}'_2\mathbf{Z}_p + \mathbf{Z}'_2\mathbf{R}_n^{-1}\mathbf{Z}_n + \mathbf{A}^{-1}\alpha_2 \\ \mathbf{Z}'_3\mathbf{X}_p + \mathbf{Z}'_3\mathbf{R}_n^{-1}\mathbf{X}_n & \mathbf{Z}'_3\mathbf{Z}_p + \mathbf{Z}'_3\mathbf{R}_n^{-1}\mathbf{Z}_n \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'_p\mathbf{Z}_2 + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{Z}_2 & \mathbf{X}'_p\mathbf{Z}_3 + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{Z}_3 \\ \mathbf{Z}'_p\mathbf{Z}_2 + \mathbf{Z}'_n\mathbf{R}_n^{-1}\mathbf{Z}_2 + \mathbf{A}^{-1}\alpha_2 & \mathbf{Z}'_p\mathbf{Z}_3 + \mathbf{Z}'_n\mathbf{R}_n^{-1}\mathbf{Z}_3 \\ \mathbf{Z}'_2\mathbf{Z}_2 + \mathbf{Z}'_2\mathbf{R}_n^{-1}\mathbf{Z}_2 + \mathbf{A}^{-1}\alpha_3 & \mathbf{Z}'_2\mathbf{Z}_3 + \mathbf{Z}'_2\mathbf{R}_n^{-1}\mathbf{Z}_3 \\ \mathbf{Z}'_3\mathbf{Z}_3 + \mathbf{Z}'_3\mathbf{R}_n^{-1}\mathbf{Z}_3 & \mathbf{Z}'_3\mathbf{Z}_3 + \mathbf{Z}'_3\mathbf{R}_n^{-1}\mathbf{Z}_3 + \mathbf{I}\alpha_4 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{m}} \\ \hat{\mathbf{pe}} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{X}'_p\mathbf{y}_p + \mathbf{X}'_n\mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}'_p\mathbf{y}_p + \mathbf{Z}'_n\mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}'_2\mathbf{y}_p + \mathbf{Z}'_2\mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{Z}'_3\mathbf{y}_p + \mathbf{Z}'_3\mathbf{R}_n^{-1}\mathbf{y}_n \end{bmatrix}$$

The α terms are as defined in equation [6.2] and $\mathbf{R}_n^{-1}$ now equals $1/(1+\mathbf{D}\alpha^{-1})$. The MME for the solutions of **b**, **u**, **m** and **pe** can therefore be set up as shown above or as in equation [6.7].

6.2.1 An illustration

Example 6.2

The same data set and genetic parameters in Section 6.1 are used below to demonstrate the principles for setting up RAM with maternal effects in the model using [6.5]. Recollect that:

$$G = \begin{bmatrix} 40 & 10 \\ 10 & 20 \end{bmatrix} \quad \text{and} \quad G^{-1} = \begin{bmatrix} 0.029 & -0.014 \\ -0.014 & 0.057 \end{bmatrix}$$

The residual variance for parents, $\sigma^2_{ep} = 350$ and because both parents of non-parents in the data are known:

$$\sigma^2_{en} = \sigma^2_e + \tfrac{1}{2}(g_{11}) = 350 + \tfrac{1}{2}(150) = 425$$

with:

$$\mathbf{R} = \begin{bmatrix} I\sigma^2_{ep} & 0 \\ 0 & I\sigma^2_{en} \end{bmatrix}$$

Then:

$$\mathbf{R} = \text{diag}(350, 350, 350, 350, 350, 425, 425, 425, 425, 425)$$

and:

$$\mathbf{R}^{-1} = \text{diag}(0.00286, 0.00286, 0.00286, 0.00286, 0.00286, 0.00235, 0.00235, 0.00235, 0.00235, 0.00235)$$

The inverse of the variance of permanent environmental effect is:

$$1/\sigma^2_{pe} = \tfrac{1}{40} = 0.025$$

Setting design matrices

The matrix **X**, which relates records to fixed effects, is the same as in Section 6.1.1. The matrix $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$ in the MME can be calculated through matrix multiplication from **X** and $\mathbf{R}^{-1}$ already set up. For illustrative purposes, the matrix $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$, when expressed as the sum of the contributions from parents' and non-parents' records, is:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \mathbf{r}^{11}_p\mathbf{X}'_p\mathbf{X}_p + \mathbf{r}^{11}_n\mathbf{X}'_n\mathbf{X}_n$$

$$= \begin{bmatrix} 0.0114 & 0.0 & 0.0 & 0.0057 & 0.0057 \\ 0.0 & 0.0029 & 0.0 & 0.0029 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0057 & 0.0029 & 0.0 & 0.0086 & 0.0 \\ 0.0057 & 0.0 & 0.0 & 0.0 & 0.0057 \end{bmatrix} + \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0047 & 0.0 & 0.0 & 0.0047 \\ 0.0 & 0.0 & 0.0071 & 0.0024 & 0.0047 \\ 0.0 & 0.0 & 0.0024 & 0.0024 & 0.0 \\ 0.0 & 0.0047 & 0.0047 & 0.0 & 0.0094 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0114 & 0.0 & 0.0 & 0.0057 & 0.0057 \\ 0.0 & 0.0076 & 0.0 & 0.0029 & 0.0047 \\ 0.0 & 0.0 & 0.0071 & 0.0024 & 0.0047 \\ 0.0057 & 0.0029 & 0.0024 & 0.0109 & 0.0 \\ 0.0057 & 0.0047 & 0.0047 & 0.0 & 0.0151 \end{bmatrix}$$

where $\mathbf{X}_p$ and $\mathbf{X}_n$ are matrices relating parents and non-parents to fixed effects, respectively, and are:

$$\mathbf{X}'_p = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X}'_n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

The matrix $\mathbf{Z}_1$, which relates records to animal effect, is:

$$\mathbf{Z}_1 = \begin{array}{c} \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \end{array} \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \left[\begin{array}{ccccccccc} 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 \end{array}\right] \end{array}$$

The first five rows correspond to animals 5 to 9, which are parents and each has one record. The last five rows correspond to the records for animals 10 to 14 (non-parents), which are related to their parents. The matrices $\mathbf{Z}_2$ and $\mathbf{Z}_3$ are exactly the same as **W** and **S** in Section 6.1.1, respectively, and the vector of observation, **y**, is the same as in Section 6.1.1. Apart from the relationship matrix, all the matrices in the MME can easily be calculated through matrix multiplication from the design matrices and vector of observation set up above. The inverse of the relationship matrix is set up only for parents ($\mathbf{A}_p^{-1}$), that is, for animals 1 to 9, using the procedure outlined in Chapter 2. The matrix $\mathbf{A}_p^{-1}\mathbf{g}^{11}$ is added to animal equations, $\mathbf{A}_p^{-1}\mathbf{g}^{22}$ to the equations for maternal genetic effects, $\mathbf{A}_p^{-1}\mathbf{g}^{12}$ to the animal by maternal genetic equations, $\mathbf{A}_p^{-1}\mathbf{g}^{21}$ to the maternal genetic by animal equations and $1/\sigma_{pe}^2$ to the diagonals of the equations for permanent environmental effects to obtain the MME. The MME are not presented because they are too large. Solving the MME by direct inversion with the equation for the first herd set to zero gives the same solutions as from the animal model (Example 6.1). However, the number of non-zero elements in the coefficient matrix was 329 compared with 429 in the animal model, due to the reduced number of equations, indicating the advantages of the RAM.

Back-solving for non-parents

The solutions for direct animal and maternal effects for non-parents are back-solved after the MME have been solved.

BACK-SOLVING FOR DIRECT EFFECTS. Solutions for direct animal effect for the non-parents are obtained from parent average and an estimate of Mendelian sampling, using equation [3.9] but with k expressed as in [3.26]. Thus the solution for the non-parent i is:

$$\hat{u}_i = 0.5(u_s + u_d) + k_i(y_i - \hat{b}_j - \hat{m}_d - \hat{p}e_d - 0.5(\hat{u}_s + \hat{u}_d)) \qquad [6.8]$$

with:

$$k_i = r^{-1}/(r^{-1} + d^{-1}g^{-1}) = 1/(1 + d^{-1}\alpha); \quad \alpha = \sigma_e^2/\sigma_a^2$$

where d is either $\frac{1}{2}$ if both parents are known or $\frac{3}{4}$ if only one parent is known. For the example data, both parents of the non-parent individuals are known; therefore:

$$k_i = 1/(1 + 2(2.333)) = 0.17647$$

For animal 10, for instance, the breeding value is:

$$\begin{aligned}\hat{u}_{10} &= 0.5(\hat{u}_3 + \hat{u}_2) + k(y_{10} - \hat{b}_2 - \hat{b}_5 - \hat{m}_2 - \hat{p}e_2 - 0.5(\hat{u}_3 + \hat{u}_2)) \\ &= 0.5(1.165 + -1.244) + 0.17647(22 - 3.386 - 27.691 - (-1.583) \\ &\quad -(-1.701) - 0.5(1.165 + (-1.244))) \\ &= -1.055\end{aligned}$$

BACK-SOLVING FOR MATERNAL EFFECTS. The equation for obtaining genetic maternal effects for non-parents can be derived as follows. From the MME, the equation for direct and genetic maternal effects for non-parent i is:

$$\begin{bmatrix} r^{-1} + n^{-1}g^{11} & n^{-1}g^{12} \\ n^{-1}g^{21} & n^{-1}g^{22} \end{bmatrix} \begin{bmatrix} \hat{u}_i \\ \hat{m}_i \end{bmatrix} = \mathbf{G}^{-1}k_2 \begin{bmatrix} \hat{u}_s + \hat{u}_d \\ \hat{m}_s + \hat{m}_d \end{bmatrix} + r^{-1} \begin{bmatrix} y_i - \hat{b}_i - \hat{m}_{dam} - \hat{p}e_{dam} \\ 0 \end{bmatrix} \qquad [6.9]$$

where n is defined in equation [6.8] and other terms are as defined in [6.4].

From the above equations:

$$\begin{aligned}\hat{m}_i &= [g^{22}(\hat{m}_s + \hat{m}_d) + g^{21}(\hat{u}_s + \hat{u}_d) - n^{-1}g^{21}(\hat{u}_i)]/n^{-1}g^{22} \\ \hat{m}_i &= n(\hat{m}_s + \hat{m}_d) + [(g^{21}n(\hat{u}_s + \hat{u}_d) - g^{21}\hat{u}_i)/g^{22}] \\ \hat{m}_i &= n(\hat{m}_s + \hat{m}_d) + g^{21}/g^{22}(n(\hat{u}_s + \hat{u}_d) - \hat{u}_i)\end{aligned}$$

Note that:

$$\begin{aligned}g^{21}/g^{22} &= [-g_{12}/(g_{11}g_{22} - g_{12}g_{21})][(g_{11}g_{22} - g_{12}g_{21})/g_{11}] \\ &= -g_{12}/g_{11}\end{aligned}$$

Therefore:

$$\hat{m}_i = n(\hat{m}_s + \hat{m}_d) + g_{12}/g_{11}(\hat{u}_i - d(\hat{u}_s + \hat{u}_d)) \qquad [6.10]$$

When both parents are known:

$$\hat{m}_i = 0.5(\hat{m}_s + \hat{m}_d) + (g_{12}/g_{11})(\hat{u}_i - 0.5(\hat{u}_s + \hat{u}_d))$$

For instance, for animal 10:

$$\begin{aligned}\hat{m}_{10} &= 0.5(\hat{m}_3 + \hat{m}_2) + (g_{12}/g_{11})(\hat{u}_5 - 0.5(\hat{u}_3 + \hat{u}_2)) \\ &= 0.5(0.736 + (-1.583)) + (-40/150)(-1.055 - 0.5(1.165 + (-1.244))) \\ &= -0.153\end{aligned}$$

The solutions for direct and maternal effects of all non-parents in the example data (animals 10 to 14) applying equations [6.7] and [6.9] are exactly the same as those obtained for these animals in the animal model.

6.3 Multivariate Maternal Animal Model

In this section, the principles of a multivariate maternal animal model are briefly outlined considering two traits affected by both direct and maternal genetic effects. Due to the fact that it is a straightforward extension of the univariate model and because of the large sizes of the matrices involved, only the model, the assumptions and the MME have been presented.

In general, the model for such multivariate analysis for two traits is:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & 0\\ 0 & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & 0\\ 0 & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix} + \begin{bmatrix}\mathbf{W}_1 & 0\\ 0 & \mathbf{W}_2\end{bmatrix}\begin{bmatrix}\mathbf{m}_1\\ \mathbf{m}_2\end{bmatrix} + \begin{bmatrix}\mathbf{S}_1 & 0\\ 0 & \mathbf{S}_2\end{bmatrix}\begin{bmatrix}\mathbf{pe}_1\\ \mathbf{pe}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} \quad [6.11]$$

where $\mathbf{y}_i$ = vector of observations for the ith trait, $\mathbf{b}_i$ = vector of fixed effects for the ith trait, $\mathbf{u}_i$ = vector of random animal effects, $\mathbf{m}_i$ = vector of random maternal (indirect) genetic effects, $\mathbf{pe}_i$ = vector of permanent environmental effects, $\mathbf{e}_i$ = vector of random residual effects, and $\mathbf{X}_i$, $\mathbf{Z}_i$, $\mathbf{W}_i$ and $\mathbf{S}_i$ are incidence matrices relating records of the ith trait to fixed, animal, maternal genetic and permanent environmental effects, respectively.

It is assumed that:

$$\text{var}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \mathbf{m}_1\\ \mathbf{m}_2\\ \mathbf{pe}_1\\ \mathbf{pe}_2\\ \mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} & g_{13}\mathbf{A} & g_{14}\mathbf{A} & 0 & 0 & 0 & 0\\ g_{21}\mathbf{A} & g_{22}\mathbf{A} & g_{23}\mathbf{A} & g_{24}\mathbf{A} & 0 & 0 & 0 & 0\\ g_{31}\mathbf{A} & g_{32}\mathbf{A} & g_{33}\mathbf{A} & g_{34}\mathbf{A} & 0 & 0 & 0 & 0\\ g_{41}\mathbf{A} & g_{42}\mathbf{A} & g_{43}\mathbf{A} & g_{44}\mathbf{A} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & q_{11} & q_{12} & 0 & 0\\ 0 & 0 & 0 & 0 & q_{21} & q_{22} & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & r_{11} & r_{12}\\ 0 & 0 & 0 & 0 & 0 & 0 & r_{21} & r_{22}\end{bmatrix} \quad [6.12]$$

In [6.12], g_{ij} are the elements of $\mathbf{G}$ and g_{ij} is the additive genetic covariance between variables i and j, where $i = 1, 2$ refer to direct effects for traits 1, 2 and $i = 3, 4$ refer to maternal effects for traits 1, 2; q_{ij} are elements of $\mathbf{Q}$, the variance and covariance matrix for permanent environmental effects, and r_{ij} are elements of $\mathbf{R}$, the variance and covariance matrix for residual effects.

The MME to be solved to obtain BLUP of u, m and pe and BLUE of estimable functions of b in [6.11] are:

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{m}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{S} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+k_1 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W}+k_2 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{S} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z}+k_2 & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W}+k_3 & \mathbf{W}'\mathbf{R}^{-1}\mathbf{S} \\ \mathbf{S}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{S}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{S}'\mathbf{R}^{-1}\mathbf{W} & \mathbf{S}'\mathbf{R}^{-1}\mathbf{S}+\mathbf{I}*\mathbf{Q}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{S}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

where:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \quad b = \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \hat{\mathbf{u}}_1 \\ \hat{\mathbf{u}}_2 \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} \hat{\mathbf{m}}_1 \\ \hat{\mathbf{m}}_2 \end{bmatrix}, \quad \mathbf{pe} = \begin{bmatrix} \hat{\mathbf{pe}}_1 \\ \hat{\mathbf{pe}}_2 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & 0 \\ 0 & \mathbf{W}_2 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & 0 \\ 0 & \mathbf{S}_2 \end{bmatrix}$$

$$k_1 = \mathbf{G}_1 * \mathbf{A}^{-1}, \quad k_2 = \mathbf{G}_2 * \mathbf{A}^{-1} \quad \text{and} \quad k_3 = \mathbf{G}_3 * \mathbf{A}^{-1}$$

with:

$$\mathbf{G}_1 = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}, \quad \mathbf{G}_2 = \begin{bmatrix} g^{13} & g^{14} \\ g^{23} & g^{24} \end{bmatrix} \quad \text{and} \quad \mathbf{G}_3 = \begin{bmatrix} g^{33} & g^{34} \\ g^{43} & g^{44} \end{bmatrix}$$

In the above equations or expressions, * denotes the direct product of the matrices concerned. The MME above can easily be set up using the procedures so far discussed in the text.

7 Analysis of Longitudinal Data

In Chapter 4, the use of a repeatability model to analyse repeated measurements on individuals was discussed and illustrated. The basic assumption of the model was that repeated measurements were regarded as expression of the same trait over time. In other words a genetic correlation of unity was assumed between repeated measurements. The model has been employed mostly in the genetic evaluation of milk production traits of dairy cattle in most countries up to 1999 (Interbull, 2000). The main advantages of this model are its simplicity, fewer computation requirements and fewer parameters compared to a multivariate model (see Chapter 5). However, the model has some drawbacks. First, test day records within lactation are assumed to measure the same trait during the whole lactation length and are used to compute 305-day yields. These test day records are actually repeated observations measured along a trajectory (days in milk) and the mean and covariance between measurements change gradually along the trajectory. Several studies have reported that heritability of daily milk yields varied with days in milk. In addition, genetic correlations between repeated measurements usually tended to decrease as the time between them increased (Pander *et al.*, 1992). The extension of test records to compute 305-day yields is unable to account for these changes in the covariance structure. Secondly, the assumption that 305-day yields across parities measure the same trait suffers from the same limitations.

However, in beef cattle, repeated measurements of growth have been analysed somewhat differently, with the assumption that measurements are genetically different but correlated traits. Usually, a multivariate model has been employed in the genetic evaluation of these traits. While the multivariate model is an improvement on the repeatability model by accounting for the genetic correlations among different records, it would be highly over-parameterized if records were available at many ages or

time periods. For instance, not only will a multivariate model for daily body weight up to yearly weight in beef cattle as different traits be over-parameterized but it will be difficult to obtain accurate estimates of the necessary genetic parameters.

An appropriate model for the analysis of repeated measurements over time or age (also termed longitudinal data) should account for the mean and covariance structure that changes with time or age and should be feasible in terms of estimating the required genetic parameters. In 1994, Schaeffer and Dekkers introduced the concept of the random regression (RR) model for the analysis of test day records in dairy cattle as a means of accounting for the covariance structure of repeated records over time or age. Almost at the same time, Kirkpatrick *et al.* (1990, 1994) introduced covariance functions (CF) to handle the analysis of longitudinal data, illustrating their methodology with growth data. The application of RR models in animal breeding for the analysis of various types of data has been comprehensively reviewed by Schaeffer (2004). Prior to the development of the RR model for genetic evaluation, milk yield test day records were analysed by Ptak and Schaeffer (1993) using a fixed regression model. Initially, the details of this model are discussed and illustrated in the next section, followed by its extension to an RR model. This is then followed by a brief presentation of CF, and the equivalence of the RR model and CF is demonstrated.

7.1 Fixed Regression Model

The theoretical framework for the fixed regression model and its application for the analysis of longitudinal data such as test day milk production traits were presented by Ptak and Schaeffer in 1993. At a national scale, a fixed regression model was implemented for the genetic evaluation of test day records of milk production traits and somatic cell counts in Germany from the 1995 until 2002. The model involved the use of individual test day records, thereby avoiding the problem of explicitly extending test day yields into 305-day yield, and accounted for the effects peculiar to all cows on the same test day within herds (herd–test–day (HTD) effect). Therefore corrections for temporary environmental effects on the day of test are more precise compared to evaluations based on 305-day yields. The model also accounted for the general shape of the lactation curve for groups of cows of similar age, calving in the same season and region. The latter is accomplished by regressing lactation curve parameters on days in milk (hence the name of the model) within the groupings for cows. The inclusion of the curve therefore allows for the correction of the means of test day yields at different stages of lactation. Fitting residual variances relevant to the appropriate stage of lactation could also account for the variation of test day yields with days in milk. The only major disadvantage is that the volume of data to be analysed is much larger,

especially in the dairy situation, as ten or more test day observations are stored relative to a single 305-day yield.

Similarly to the repeatability model, at the genetic level the fixed regression model assumes that test day records within a lactation are repeated measurements of the same trait, that is, a genetic correlation of unity among test day observations. Usually the permanent environmental effect is included in the model to account for environmental factors with permanent effects on all test day yields within lactation.

The fixed regression model is of the form:

$$y_{tij} = htd_i + \sum_{k=0}^{nf} \phi_{tjk}\beta_k + u_j + pe_j + e_{tij}$$

where y_{tij} is the test day record of cow j made on day t within herd–test–date (*htd*) subclass i; β_k are fixed regressions coefficients; u_j and pe_j refer to animal additive genetic and permanent environmental effects, respectively, for animal j; ϕ_{tjk} is of the kth Legendre polynomials or any other curve parameter, for the test day (TD) record of cow j made on day t; nf is the order of fit for Legendre polynomials used to model the fixed regressions (fixed lactation curves); and e_{tij} is the random residual. In matrix notation, the model may be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Qu} + \mathbf{Zpe} + \mathbf{e} \qquad [7.1]$$

where **y** is the vector of TD yields, **b** is a vector of solutions for HTD and fixed regressions, **u** and **pe** are vectors of animal additive genetic and permanent environmental effects, respectively. The variances of **u** and **pe** are as defined in [4.1]. The matrices **X**, **Q** and **Z** are incidence matrices and are described in detail in the next section, which illustrates the application of the model. It is assumed that var(**u**) = $\mathbf{A}\sigma_u^2$, var(**pe**) = $\mathbf{I}\sigma_p^2$, and var(**e**) = $\mathbf{I}\sigma_e^2 = \mathbf{R}$. The mixed model equations (MME) for equation [7.1] are:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Q} & \mathbf{X'Z} \\ \mathbf{Q'X} & \mathbf{Q'Q} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Q'Z} \\ \mathbf{Z'X} & \mathbf{Z'Q} & \mathbf{Z'Z} + \alpha_2 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \mathbf{p\hat{e}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Q'y} \\ \mathbf{Z'y} \end{pmatrix}$$

with:

$$\alpha_1 = \sigma_e^2/\sigma_u^2 \quad \text{and} \quad \alpha_2 = \sigma_e^2/\sigma_p^2$$

7.1.1 Illustration

Example 7.1

Given in Table 7.1 are the test day fat yields of five cows in a herd with details of HTD and days in milk (DIM). The aim is to estimate solutions for

Table 7.1. Test day fat yields (TDY) for some cows in a herd.

	Animals									
	4		5		6		7		8	
DIM	HTD	TDY	HTD	TDY	HTD	TDY	HTD	TDY	HTD	TDY
4	1	17.0	1	23.0	6	10.4	4	22.8	1	22.2
38	2	18.6	2	21.0	7	12.3	5	22.4	2	20.0
72	3	24.0	3	18.0	8	13.2	6	21.4	3	21.0
106	4	20.0	4	17.0	9	11.6	7	18.8	4	23.0
140	5	20.0	5	16.2	10	8.4	8	18.3	5	16.8
174	6	15.6	6	14.0			9	16.2	6	11.0
208	7	16.0	7	14.2			10	15.0	7	13.0
242	8	13.0	8	13.4					8	17.0
276	9	8.2	9	11.8					9	13.0
310	10	8.0	10	11.4					10	12.6

DIM, days in milk; HTD, herd–test–day.

HTD effects, regression coefficients for a fixed lactation curve fitting Legendre polynomials of order 4, solutions for permanent environmental effects and breeding values for animal effects using equation [7.1]. Assume that the pedigree structure is the same as in Example 4.1 and that the estimated variances for additive genetic effects, permanent environmental effects and residual variances were 5.521 kg^2, 8.470 kg^2 and 3.710 kg^2, respectively. Then:

$$\alpha_1 = \sigma_e^2/\sigma_u^2 = 3.710/5.521 = 0.672$$

and:

$$\alpha_2 = \sigma_e^2/\sigma_p^2 = 3.710/8.470 = 0.438$$

The modelling of the fixed lactation curve by means of Legendre polynomials implies the need to compute $\mathbf{\Phi}$, which is the matrix of Legendre polynomials evaluated at the different DIM. The matrix $\mathbf{\Phi}$ is of order t (the number of DIM) by k (where k is the order of fit) with element $\phi_{ij} = \phi_j(a_t)$, which is the jth Legendre polynomial evaluated at the standardized DIM t (a_t). Therefore $\mathbf{\Phi} = \mathbf{M\Lambda}$, where $\mathbf{M}$ is the matrix containing the polynomials of the standardized DIM values and $\mathbf{\Lambda}$ is a matrix of order k containing the coefficients of Legendre polynomials. The calculation of $\mathbf{\Phi}$ is outlined in Appendix G, and matrix $\mathbf{\Phi}$ for Example 7.1 is shown in equation [g.1] in the appendix.

Setting up the incidence matrices for the MME

In equation [7.1], let $\mathbf{Xb} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2$; then, in Example 7.1, the matrix $\mathbf{X}_1$, which relates records to HTD effects, is of order n_{td} (number of TD records)

and is too large to be presented. However, $\mathbf{X}_1'\mathbf{X}_1$ is diagonal and is:

$$\mathbf{X}_1'\mathbf{X}_1 = \text{diagonal}[3, 3, 3, 4, 4, 5, 5, 5, 5, 5]$$

The matrix $\mathbf{X}_2$ of order n_{td} by nf contains Legendre polynomials (covariables) corresponding to the DIM of the ith TD yield. Thus the ith row of $\mathbf{X}_2$ contains elements of the row of $\mathbf{\Phi}$ corresponding to the DIM for the ith record. The matrix $\mathbf{X}_2$, with rows for the first three TD records of cow 4 and the last three TD records of cow 8, is:

$$\begin{bmatrix} 0.7071 & -1.2247 & 1.5811 & -1.8704 & 2.1213 \\ 0.7071 & -0.9525 & 0.6441 & -0.0176 & -0.6205 \\ 0.7071 & -0.6804 & -0.0586 & 0.7573 & -0.7757 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.7071 & 0.6804 & -0.0586 & -0.7573 & -0.7757 \\ 0.7071 & 0.9525 & 0.6441 & 0.0176 & -0.6205 \\ 0.7071 & 1.2247 & 1.5811 & 1.8704 & 2.1213 \end{bmatrix}$$

and $\mathbf{X}_2'\mathbf{X}_2$ is:

$$\mathbf{X}_2'\mathbf{X}_2 = \begin{bmatrix} 20.9996 & -4.4261 & 4.0568 & -0.8441 & 8.7149 \\ -4.4261 & 24.6271 & -4.7012 & 11.1628 & -3.0641 \\ 4.0568 & -4.7012 & 31.0621 & -6.6603 & 19.0867 \\ -0.8441 & 11.1628 & -6.6603 & 38.6470 & -8.8550 \\ 8.7149 & -3.0641 & 19.0867 & -8.8550 & 48.2930 \end{bmatrix}$$

Considering only animals with records, $\mathbf{Q} = \mathbf{Z}$ and is a matrix of order 5 (number of animals) by n_{td}. The matrix $\mathbf{Q}'$ could be represented as:

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{q}_4' & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{q}_5' & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{q}_6' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_7' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{q}_8' \end{bmatrix}$$

where $\mathbf{q}_i'$ is a vector of ones with size equal to the number of TD records for the ith cow. The matrices $\mathbf{Q}'\mathbf{Q}$ and $\mathbf{Z}'\mathbf{Z}$ are both diagonal and equal. Thus:

$$\mathbf{Q}'\mathbf{Q} = \mathbf{Z}'\mathbf{Z} = \text{diag}[10, 10, 5, 7, 10]$$

The matrix $\mathbf{A}^{-1}$ has been given in Example 4.1. The remaining matrices in MME could be obtained as outlined in earlier chapters. Solving the MME, with the solution for the tenth level of HTD effects constrained to zero,

gives the following results:

Effects	Solutions	
Herd–test–day		
1	10.9783	
2	7.9951	
3	8.7031	
4	8.2806	
5	6.3813	
6	3.1893	
7	3.3099	
8	3.3897	
9	0.6751	
10	0.0000	
Fixed regression coefficients		
1	16.3082	
2	–0.5227	
3	–0.1245	
4	0.5355	
5	–0.4195	
Animal effect		
	EBV for daily yield	EBV for 305-day yield
1	–0.3300	–100.6476
2	–0.1604	–48.9242
3	0.4904	149.5718
4	0.0043	1.3203
5	–0.2449	–74.7065
6	–0.8367	–255.2063
7	1.1477	350.0481
8	0.3786	115.4757
Permanent environmental effects		
Cow	Solutions for daily yield	Solutions for daily yield
4	–0.6156	–187.7634
5	–0.4151	–126.6150
6	–1.6853	–514.0274
7	2.8089	856.7092
8	–0.0928	–28.3035

EBV, estimated breeding value.

The solutions for the fixed regressions are regression coefficients from which plots of lactation curves can be obtained. In practice, the fixed regressions are usually fitted within a group of cows calving in the same season in the same parity and of similar age. Thus, the curves obtained for various groups of cows are useful for examining the influence of different

environmental factors on lactation curves. In Example 7.1, one fixed lactation curve was fitted for all cows and a vector (**v**) of actual daily fat yield (kg) from days 4 to 310 can be obtained as:

$$\mathbf{v} = \mathbf{\Phi}\hat{\mathbf{b}} = \sum_{i=4}^{310}\sum_{j=1}^{nf}\phi_{ij}\hat{b}_{2j}$$

where $\mathbf{\Phi}$ is a matrix of Legendre polynomials evaluated from 4 to 310 DIM, as described in Appendix G. From the above equation, v_{38}, for instance, is:

$$v_{38} = [0.7071 \quad -0.9525 \quad 0.6441 \quad -0.0176 \quad -0.6205]\hat{\mathbf{b}}_2 = 12.2001$$

For the DIM in the example data set, **v** is:

(DIM)	4	38	72	106	140	174	208	242	276	310
v =	[10.0835	12.2001	12.6254	12.2077	11.5679	11.0407	10.9156	11.1111	11.2500	10.8297]

A graph of the fixed lactation curve can be obtained by plotting the elements of **v** against DIM.

The estimated breeding value (BV) for animals and solutions for permanent environmental effect obtained by solving the MME are those for daily fat yield. To obtain estimated BV or solutions for pe effects on the *n*th DIM, these solutions are multiplied by *n*. This is implicit from the assumptions stated earlier of genetic correlations of unity among TD records. Thus estimated BV for 305 days, shown in the table of results above, was obtained by multiplying the solutions for daily fat yield by 305.

Partitioning breeding values and solutions for permanent environmental effects

Similarly to the repeatability model, estimated breeding values of animals can be partitioned in terms of contributions from various sources, using equation [3.8]. The yield deviation (YD) for an animal is now calculated as the average of corrected TD records. The correction is for effects of HTD, fixed regressions and pe. Thus for cow 6 with five TD records, YD_6 is:

$$YD_6 = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'(\mathbf{y}_6 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \hat{\mathbf{p}}\mathbf{e})$$

with:

$$\mathbf{y}_6 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \mathbf{p}\hat{\mathbf{e}} = \mathbf{y}_c = \begin{bmatrix}10.4\\12.3\\13.2\\11.6\\8.4\end{bmatrix} - \begin{bmatrix}3.1893\\3.3099\\3.3897\\0.6751\\0.0000\end{bmatrix} - \begin{bmatrix}10.0835\\12.2001\\12.6254\\12.2077\\11.5679\end{bmatrix} - \begin{bmatrix}-1.6853\\-1.6853\\-1.6853\\-1.6853\\-1.6853\end{bmatrix}$$

$$= \begin{bmatrix}-1.1875\\-1.5247\\-1.1298\\0.4025\\-1.4826\end{bmatrix}$$

and:

$$YD_6 = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'(\mathbf{y}_c) = \frac{1}{5}\left(\mathbf{q}_6' \begin{bmatrix} -1.1875 \\ -1.5247 \\ -1.1298 \\ 0.4025 \\ -1.4826 \end{bmatrix}\right) = -4.9221/5 = -0.9844$$

Then the solution for additive genetic effect for animal 6 using equation [3.8] is:

$$\begin{aligned} \hat{u}_6 &= w_1((\hat{u}_1 + \hat{u}_5)/2) + w_2(YD_6) \\ &= w_1((-0.3300 + -0.2449)/2) + w_2(-0.9844) = -0.8367 \end{aligned}$$

with $w_1 = 2(0.672)/6.344$, $w_2 = 5/6.344$ and 6.344 = the sum of the numerators of w_1 and w_2.

For animal 8 with ten TD records, the solution for additive genetic effect is:

$$\begin{aligned} \hat{u}_8 &= w_1((\hat{u}_1 + \hat{u}_7)/2) + w_2(YD_8) \\ &= w_1((-0.3300 + 1.1477)/2) + w_2(0.3746) = 0.3786 \end{aligned}$$

with $w_1 = 2(0.672)/11.344$, $w_2 = 10/11.344$ and 11.344 = the sum of the numerators of w_1 and w_2. The weights on yield deviations were 0.7882 and 0.8815 for animals 6 and 8, respectively. This illustrates the fact that, as the number of TD increases, more emphasis is placed on the performance records of the animal. Considering animal 4 with ten TD records and a progeny, her breeding value can be calculated as:

$$\begin{aligned} \hat{u}_4 &= w_1((\hat{u}_1 + \hat{u}_2)/2) + w_2(YD_4) + w_3(\hat{u}_7 - 0.5\hat{u}_3) \\ &= w_1((-0.3300 + -0.1604)/2) + w_2(-0.0226) + w_3(2(1.1477) - 0.4934) \\ &= 0.0043 \end{aligned}$$

where $w_1 = 2(0.672)/11.68$, $w_2 = 10/11.68$ and $w_3 = 0.5(0.672)/11.68$ and 11.68 is the sum of the numerators of w_1, w_2 and w_3. There was a slight reduction to the weight given to parent average from 0.1185 (animal 8) to 0.1151 (animal 4) due to the additional information from progeny.

The solution for pe of an animal can be calculated as in Section 4.1.2, using equation [4.4]. Here the correction of the TD records is for the estimates for HTD effects and fixed regressions and animal effect. Thus, for cow 6, pe_6 can be calculated as:

$$\hat{pe}_6 = \left(\mathbf{t}' \begin{bmatrix} 10.4 \\ 12.3 \\ 13.2 \\ 11.6 \\ 8.4 \end{bmatrix} - \begin{bmatrix} 3.1893 \\ 3.3099 \\ 3.3897 \\ 0.6751 \\ 0.0000 \end{bmatrix} - \begin{bmatrix} 10.0835 \\ 12.2001 \\ 12.6254 \\ 12.2077 \\ 11.5679 \end{bmatrix} - \begin{bmatrix} -0.8367 \\ -0.8367 \\ -0.8367 \\ -0.8367 \\ -0.8367 \end{bmatrix}\right) \Big/ 5.4380$$

$$= -9.1650/5.4380 = -1.6853$$

where **t** is a column vector of order 5 (number of TD records for the animal), with all elements equal to one. However, in contrast to pe estimates in Example 4.1, these pe estimates represent permanent environmental factors affecting TD records within lactation.

7.2 Random Regression Model

In Section 7.1, the advantage of including fixed regressions on days in milk in the model was to account for the shape of the lactation curve for different groups of cows. However, the breeding values estimated represented genetic differences between animals at the height of the curves. Although different residual variances associated with different stages of lactation could be fitted with the fixed regression model, the model did not account for the covariance structure at the genetic level. Schaeffer and Dekkers (1994) extended the fixed regression model for genetic evaluation by considering the regression coefficients on the same covariables as random, thus allowing for between-animal variation in the shape of the curve. Thus the genetic differences among animals could be modelled as deviations from the fixed lactation curves by means of random parametric curves (see Guo and Swalve, 1997) or orthogonal polynomials, such as Legendre polynomials (Brotherstone *et al.*, 2000), or even non-parametric curves, such as natural cubic splines (White *et al.*, 1999). Most studies have used Legendre polynomials as they make no assumption about the shape of the curve and are easy to apply. The RR model has also been employed for the analysis of growth data in pigs (Andersen and Pedersen, 1996) and beef cattle (Meyer, 1999). An additional benefit of the RR model in dairy cattle is that it provides the possibility of genetic evaluation for persistence of the lactation. A typical random regression model (RRM) especially for the analysis of dairy cattle test day records is of the form:

$$y_{tijk} = htd_i + \sum_{k=0}^{nf} \phi_{jtk}\beta_k + \sum_{k=0}^{nr} \phi_{jtk}\mathrm{u}_{jk} + \sum_{k=0}^{nr} \phi_{jtk}\mathrm{pe}_{jk} + e_{tijk}$$

where y_{tijk} is the test day record of cow j made on day t within htd subclass i; β_k are fixed regression coefficients; u_{jk} and pe_{jk} are the kth random regression for animal and permanent environmental effects, respectively, for animal j; ϕ_{jtk} is the kth Legendre polynomial for the test day record of cow j made on day t; nf is the order of polynomials fitted as fixed regressions; nr is the order of polynomials for animal and pe effects; and e_{tijk} is the random residual. The model in matrix notation is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Qu} + \mathbf{Zpe} + \mathbf{e}$$

The vectors **y**, **b** and the matrix **X** are as described in Example 7.1. However, **u** and **pe** are now vectors of random regressions for animal additive genetic and pe effects. The matrices **Q** and **Z** are covariable

matrices and, if only animals with records are considered, the *i*th row of these matrices contains the orthogonal polynomials (covariables) corresponding to the DIM of the *i*th TD yield. If the order of fit is the same for animal and pe effects, $\mathbf{Q} = \mathbf{Z}$, considering only animals with records. This would not be the case if the order of fit is different for animal and pe effects. In general, considering animals with records, the order of either **Q** or **Z** is n_{td} (number of TD records) by *nk*, where *nk* equals *nr* times the number of animals with records. It is assumed that $\text{var}(\mathbf{u}) = \mathbf{A} \otimes \mathbf{G}$, $\text{var}(\mathbf{pe}) = \mathbf{I} \otimes \mathbf{P}$ and $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}$, where **A** is the numerator relationship matrix, ⊗ is the Kronecker product and **G** and **P** are of the order of polynomial fitted for animal and pe effects. The MME are:

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Q} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Q}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} + \mathbf{A}^{-1} \otimes \mathbf{G} & \mathbf{Q}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Q} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{P} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \\ \mathbf{p}\hat{\mathbf{e}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Q}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

7.2.1 Numerical application

Example 7.2
Analysis of the data in Table 7.1 is undertaken fitting an RR model with Legendre polynomials of order 4 fitted for the fixed lactation curve and Legendre polynomials of order 3 fitted for both random animal and pe effects. The covariance matrices for the random regression coefficients for animal effect and pe effects are:

$$\mathbf{G} = \begin{bmatrix} 3.297 & 0.594 & -1.381 \\ 0.594 & 0.921 & -0.289 \\ -1.381 & -0.289 & 1.005 \end{bmatrix}; \quad \mathbf{P} = \begin{bmatrix} 6.872 & -0.254 & -1.101 \\ -0.254 & 3.171 & 0.167 \\ -1.101 & 0.167 & 2.457 \end{bmatrix}$$

and the residual variance equals 3.710 for all stages of lactation.

As indicated earlier, the above **G** or **P** matrix models the genetic or permanent environment covariance structure of fat yields over the whole lactation length. Thus the genetic covariance between DIM *i* and *j* along the trajectory can be calculated from **G**. For instance, the genetic variance for DIM *i* (v_{ii}) can be calculated as:

$$v_{ii} = \mathbf{t}_i \mathbf{G} \mathbf{t}_i'$$

where $\mathbf{t}_i = \phi_{ik}$, the *i*th row vector of **Φ**, for day *i*, and *k* is the order of fit. The genetic covariance between DIM *i* and *j* (v_{ij}) therefore is:

$$v_{ii} = \mathbf{t}_i \mathbf{G} \mathbf{t}_j'$$

Using the **G** matrix in Example 7.2, the genetic variance for DIM 106 equals 2.6433 kg^2, with $\mathbf{t}_{106} = [0.7071 \ -0.4082 \ -0.5271]$, and the genetic

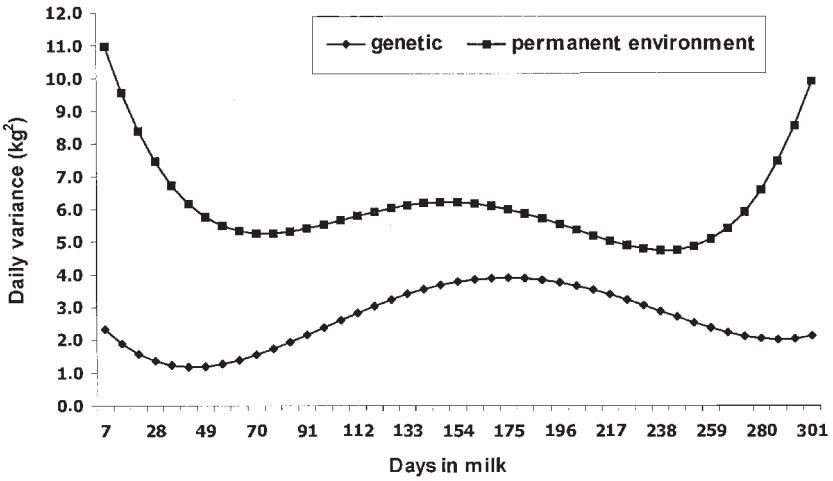


Fig. 7.1. The estimates of daily genetic and permanent environmental variances by days in milk.

covariance between DIM 106 and 140 equals 3.0219 kg, with $\mathbf{t}_{140}$ = [0.7071 −0.1361 −0.7613]. Similar calculations can be carried out using **p**. The plots of daily genetic and permanent environmental variances against DIM are shown in Fig. 7.1, indicating how these variances change through the lactation length.

Setting up the matrices for the MME

The setting up of the matrix **X** has been described in Example 7.1. The matrix $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$ can easily be obtained by matrix multiplication. Considering only animals with records, $\mathbf{Q}'$ can be represented as:

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{Q}_4' & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_5' & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_6' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_7' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_8' \end{bmatrix}$$

where $\mathbf{Q}_i'$ is the matrix of order nr by k (number of TD records for animal i). Thus for animal 6, $\mathbf{Q}_6'$ is:

$$\mathbf{Q}_6' = \begin{bmatrix} 0.7071 & 0.7071 & 0.7071 & 0.7071 & 0.7071 \\ -1.2247 & -0.9525 & -0.6804 & -0.4082 & -0.1361 \\ 1.5811 & 0.6441 & -0.0586 & -0.5271 & -0.7613 \end{bmatrix}$$

For all animals with records, $\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ and are block diagonals. For instance, $\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q}$ for the first three cows (cows 4, 5 and 6) with

records is:

$$\begin{bmatrix} 1.348 & 0.000 & 0.335 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.000 & 1.647 & 0.000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.335 & 0.000 & 2.035 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.348 & 0.000 & 0.335 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.000 & 1.647 & 0.000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.335 & 0.000 & 2.035 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.674 & -0.648 & 0.167 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.648 & 0.824 & -0.591 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.167 & -0.591 & 1.018 \end{bmatrix}$$

When all animals are considered, $\mathbf{Q'R^{-1}Q}$ is augmented by *nr* columns and rows per ancestor without records (that is, animals 1–3). The matrix $\mathbf{A^{-1}}\otimes\mathbf{G^{-1}}$ is then added to $\mathbf{Q'R^{-1}Q}$ and $\mathbf{P^{-1}}$ added to $\mathbf{Z'R^{-1}Z}$ to obtain the MME. Solving the MME by direct inversion with the solution for level 10 of HTD effects constrained to zero gave the following results:

Effects	Solutions			
Herd–test–day				
1	10.0862			
2	7.5908			
3	8.5601			
4	8.2430			
5	6.3161			
6	3.0101			
7	3.1085			
8	3.1718			
9	0.5044			
10	0.0000			
Fixed regression				
1	16.6384			
2	–0.6253			
3	–0.1346			
4	0.3479			
5	–0.4218			
Animal	Regression coeffients			305-day breeding value
1	–0.0583	0.0552	0.0442	–12.3731
2	–0.0728	–0.0305	–0.0244	–15.7347
3	0.1311	–0.0247	0.0686	28.1078
4	0.3445	0.0063	–0.3164	74.8132
5	–0.4537	–0.0520	0.2798	–98.4153

(*Continued*)

Effects	Solutions			
Animal	Regression coeffients			305-day breeding value
6	–0.5485	0.0730	0.1946	–118.4265
7	0.8518	–0.0095	–0.3131	184.1701
8	0.2209	0.0127	–0.0174	47.6907
Permanent environmental effects				
Cow	Regression coefficients			305-day solutions
4	–0.6487	–0.3601	–1.4718	–138.4887
5	–0.7761	0.1370	0.9688	–168.5531
6	–1.9927	0.9851	–0.0693	–427.2378
7	3.5188	–1.0510	–0.4048	756.9415
8	–0.1013	0.2889	0.9771	–22.6619

The solutions for HTD and fixed regression for the RRM are similar to those from the fixed regression model. Lactation curves can be constructed from the fixed regression, as described in Section 7.1.1, and influences of different environmental factors on the curves can be evaluated. Each animal has *nr* regression coefficients as solutions for animal and permanent environmental effects. These are not useful for ranking animals and need to be converted to breeding values for any particular day of interest. Usually, in dairy cattle, values are calculated for 305 days' yield and these have been shown above in the table of results. The estimated breeding value from days 6 to *m* for animal *k* (EBV_{km}) is calculated as:

$$EBV_{km} = \mathbf{t}\hat{\mathbf{u}}_k; \quad \text{with } \mathbf{t} \text{ containing elements } t_j = \sum_{i=6}^{m}\sum_{j=0}^{nr}\phi_{ij} \qquad [7.2]$$

where $\mathbf{t}$ is a row vector of order *nr*, with the *j*th elements equal to the sum of the *j*th orthogonal polynomial from days 6 to *m*, and $\hat{\mathbf{u}}_k$ is a vector for the regression coefficient of animal *k*. For Example 7.2, the matrix $\mathbf{\Phi}$ for days 4 to 310 has not been shown because of the size but can be generated as described in Appendix G. Assuming 305-day breeding values are computed from days 6 to 310, then the vector $\mathbf{t}$ for Example 7.2, calculated from days 6 to 310, is:

$$\mathbf{t} = [215.6655 \quad 2.4414 \quad -1.5561]$$

The breeding value for the 305-day yield for animal 4, for instance, can be calculated as:

$$\mathbf{t}\hat{\mathbf{u}}_4 = [215.6655 \quad 2.4414 \quad -1.5561]\begin{pmatrix} 0.3445 \\ 0.0063 \\ -0.3164 \end{pmatrix} \approx 74.81$$

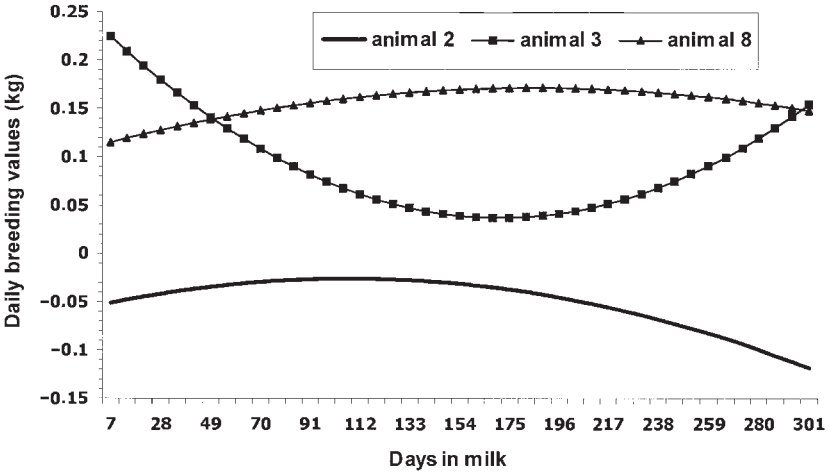


Fig. 7.2. The estimates of daily breeding values for some animals by days in milk.

Over the lactation length, daily breeding values can be computed for each animal from the random regression coefficients. Genetic lactation curves can be obtained for each animal by plotting these daily breeding values against DIM and differences between curves for different animals can then be studied. Let **v** be a vector containing daily breeding values for days 6 to 310, then **v** can be calculated as:

$$\mathbf{v} = \mathbf{T}\hat{\mathbf{u}}_k; \quad \text{with } \mathbf{T} = t_{ij} = \sum_{i=6}^{310}\sum_{j=0}^{nr}\phi_{ij}$$

The plots of the daily breeding values for animals 2, 3 and 8 are shown in Fig. 7.2. The plots indicate that the animal with the highest 305-day breeding value for fat yield also had the highest daily breeding values along the lactation length.

If the trait being analysed is milk yield, persistence breeding values can be calculated from the daily breeding values. For instance, persistence predicted transmitting ability (PS_{PTA}) for milk yield can be calculated (Schaeffer *et al.*, 2000) as:

$$PS_{PTA} = \frac{PTA_{280} - PTA_{60} + \bar{y}_{280}}{\bar{y}_{60}}(100)$$

where PTA_{60} and PTA_{280} are predicted transmitting abilities for day milk yield for an animal at days 60 and 280, respectively, and $\bar{y}_{60}$ and $\bar{y}_{280}$ are the average milk yields of cows in the genetic base at days 60 and 280, respectively.

7.2.2 Partitioning animal solutions from the random regression model

Equations for calculating the contribution of information from various sources to the solutions (random regression coefficients) of an animal from an RRM were presented by Mrode and Swanson (2004). These

equations are the same as those presented in Section 5.1.3 for the multivariate model. Test day records of cows contribute to random regressions for the animal effect through the yield deviations. The calculation of the vector of yield deviations (**YD**) is first examined. Using the same argument for deriving [5.7], the equation for **YD** for an RRM is:

$$\mathbf{YD} = (\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q})^{-1}(\mathbf{Q}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Zp}\hat{\mathbf{e}})) \qquad [7.3]$$

While this equation is similar to equation [5.6] for yield deviation under a multivariate model, here **YD** is a vector of weighted regressions of the animal's TD yields adjusted for all effects other than additive genetic effect. Since **YD** is a vector of regressions, it can be used to generate actual yield deviations for any DIM using equation [7.2]. Thus actual yield deviation (yd*) for day *m*, for instance, equals $\mathbf{v}'\mathbf{YD}$, where **v** is a vector of order *nr* with $v_m = \phi_{mj}$ and $j = 1,nr$. The actual yield deviation for 305-day yield can be calculated using equation [7.2] but with $\hat{\mathbf{u}}$ replaced with **YD**.

The calculation of **YD** for cow 6 in Example 7.2 is illustrated below. First, the vector of TD records for cow 6 corrected for all effects ($\mathbf{y}_c$) other than the additive genetic effect is:

$$\mathbf{y}_c = \mathbf{y}_6 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \hat{\mathbf{p}}\mathbf{e}$$

$$\mathbf{y}_c = \begin{bmatrix} 10.4 \\ 12.3 \\ 13.2 \\ 11.6 \\ 8.4 \end{bmatrix} - \begin{bmatrix} 3.0101 \\ 3.1085 \\ 3.1718 \\ 0.5044 \\ 0.0000 \end{bmatrix} - \begin{bmatrix} 10.7725 \\ 12.5295 \\ 12.7890 \\ 12.3454 \\ 11.7641 \end{bmatrix} - \begin{bmatrix} -2.7251 \\ -2.3920 \\ -2.0752 \\ -1.7746 \\ -1.4904 \end{bmatrix} = \begin{bmatrix} -0.6576 \\ -0.9460 \\ -0.6856 \\ 0.5249 \\ -1.8738 \end{bmatrix}$$

where $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$ are vectors of solutions for HTD and fixed regression coefficients. The matrices $\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q}$ and $\mathbf{Q}'\mathbf{R}^{-1}\mathbf{y}_c$ are:

$$\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} = \begin{bmatrix} 0.6738 & -0.6484 & 0.1674 \\ -0.6484 & 0.8235 & -0.5906 \\ 0.1674 & -0.5906 & 1.0177 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}'\mathbf{R}^{-1}\mathbf{y}_c = \begin{bmatrix} -0.6934 \\ 0.5967 \\ -0.1237 \end{bmatrix}$$

Using equation [7.3], yield deviation for cow 6 ($\mathbf{YD}_6$) is:

$$\mathbf{YD}_6 = (\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{R}^{-1}\mathbf{y}_c = \begin{bmatrix} -5.0004 \\ -4.6419 \\ -1.9931 \end{bmatrix}$$

The actual yield deviation at 305 DIM for cow 6 using [7.2] with $\hat{\mathbf{u}}$ replaced with $\mathbf{YD}_6$ is –1086.6450.

The equation for the partitioning of random regression coefficients for animals to contributions for parent average, yield deviations and progeny is:

$$\hat{\mathbf{u}}_{anim} = \mathbf{W}_1\mathbf{PA} + \mathbf{W}_2(\mathbf{YD}) + \mathbf{W}_3\mathbf{PC} \qquad [7.4]$$

with:

$$\mathbf{PC} = \sum \alpha_{prog}(2\hat{\mathbf{u}}_{prog} - \hat{\mathbf{u}}_{mate})/\sum \alpha_{prog} \quad \text{and} \quad \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3 = \mathbf{I}$$

This is the same equation as [5.8], which partitioned breeding values under the multivariate model. The weights $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ are as defined in [5.8] but here $\mathbf{W}_i$ is of the order of orthogonal polynomials for animal effects. Illustrating with cow 6, the weights on parent average ($\mathbf{W}_1$) and yield deviation ($\mathbf{W}_2$) can be calculated as:

$$\mathbf{W}_1 = \begin{bmatrix} 2.1520 & -0.9957 & 2.0986 \\ -0.9957 & 3.2921 & -0.3580 \\ 2.0986 & -0.3580 & 5.7284 \end{bmatrix}^{-1} \begin{bmatrix} 1.4781 & -0.3473 & 1.9313 \\ -0.3473 & 2.4685 & 0.2326 \\ 1.9313 & 0.2326 & 4.7107 \end{bmatrix}$$

$$(\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} + \mathbf{G}^{-1}\alpha_6)^{-1} \qquad (2\mathbf{G}^{-1}\alpha_{par})$$

$$= \begin{bmatrix} 0.6156 & 0.1940 & 0.2987 \\ 0.0935 & 0.8107 & 0.2402 \\ 0.1175 & 0.0202 & 0.7279 \end{bmatrix}$$

$$(\mathbf{W}_1)$$

and:

$$\mathbf{W}_2 = \begin{bmatrix} 2.1520 & -0.9957 & 2.0986 \\ -0.9957 & 3.2921 & -0.3580 \\ 2.0986 & -0.3580 & 5.7284 \end{bmatrix}^{-1} \begin{bmatrix} 0.6738 & -0.6484 & 0.1674 \\ -0.6484 & 0.8235 & -0.5906 \\ 0.1674 & -0.5906 & 1.0177 \end{bmatrix}$$

$$(\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q} + \mathbf{G}^{-1}\alpha_6)^{-1} \qquad (\mathbf{Q}'\mathbf{R}^{-1}\mathbf{Q})$$

$$= \begin{bmatrix} 0.3844 & -0.1940 & -0.2987 \\ -0.0935 & 0.1893 & -0.2402 \\ -0.1175 & -0.0202 & 0.2721 \end{bmatrix}$$

$$(\mathbf{W}_2)$$

The contributions from **PA** (parent average) and **YD** to the random regression coefficients for cow 6 are:

$$\begin{bmatrix} \hat{u}_0 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \mathbf{W}_1 \begin{bmatrix} -0.2560 \\ 0.0016 \\ 0.1178 \end{bmatrix} + \mathbf{W}_2 \begin{bmatrix} -5.0004 \\ -4.6419 \\ -1.9931 \end{bmatrix} = \begin{bmatrix} -0.1221 \\ 0.0057 \\ 0.0557 \end{bmatrix} + \begin{bmatrix} -0.4265 \\ 0.0674 \\ 0.1389 \end{bmatrix} = \begin{bmatrix} -0.5485 \\ 0.0730 \\ 0.1946 \end{bmatrix}$$

For cow 8 with ten TD records and no progeny, equation [7.4] is:

$$\begin{bmatrix} \hat{u}_0 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} 0.3893 & -0.0844 & 0.1763 \\ -0.0604 & 0.5903 & 0.0353 \\ 0.1576 & 0.0425 & 0.6379 \end{bmatrix} \begin{bmatrix} 0.3967 \\ 0.0228 \\ -0.1787 \end{bmatrix} + \begin{bmatrix} 0.6107 & 0.0844 & -0.1763 \\ 0.0604 & 0.4097 & -0.0353 \\ -0.1576 & -0.0425 & 0.3621 \end{bmatrix} \begin{bmatrix} 0.2102 \\ 0.0574 \\ 0.1893 \end{bmatrix}$$

$$(\mathbf{W}_1) \qquad (\mathbf{PA}) \qquad (\mathbf{W}_2) \qquad (\mathbf{YD}_8)$$

and:

$$\begin{bmatrix}\hat{u}_0\\ \hat{u}_1\\ \hat{u}_2\end{bmatrix} = \begin{bmatrix}0.1210\\ -0.0168\\ -0.0505\end{bmatrix} + \begin{bmatrix}0.0998\\ 0.0295\\ 0.0330\end{bmatrix} = \begin{bmatrix}0.2208\\ 0.0127\\ -0.0175\end{bmatrix}$$

Considering cow 4, with ten TD records and a progeny:

$$\begin{bmatrix}\hat{u}_0\\ \hat{u}_1\\ \hat{u}_2\end{bmatrix} = \underset{(\mathbf{W}_1)}{\begin{bmatrix}0.3488 & -0.0684 & 0.1393\\ -0.0490 & 0.5132 & 0.0284\\ 0.1245 & 0.0343 & 0.5451\end{bmatrix}} \underset{(\mathbf{PA})}{\begin{bmatrix}-0.0655\\ 0.0123\\ -0.0343\end{bmatrix}} + \underset{(\mathbf{W}_2)}{\begin{bmatrix}0.5640 & 0.0856 & -0.1741\\ 0.0613 & 0.3585 & -0.0355\\ -0.1557 & -0.0428 & 0.3186\end{bmatrix}} \underset{(\mathbf{YD})}{\begin{bmatrix}0.2711\\ -0.0508\\ -0.6412\end{bmatrix}}$$

$$+ \underset{(\mathbf{W}_3)}{\begin{bmatrix}0.0872 & -0.0171 & 0.0348\\ -0.0123 & 0.1283 & 0.0071\\ 0.0311 & 0.0086 & 0.1363\end{bmatrix}} \underset{(\mathbf{PC})}{\begin{bmatrix}1.5725\\ 0.0057\\ -0.6948\end{bmatrix}}$$

$$\begin{bmatrix}\hat{u}_0\\ \hat{u}_1\\ \hat{u}_2\end{bmatrix} = \begin{bmatrix}-0.0285\\ 0.0086\\ -0.0264\end{bmatrix} + \begin{bmatrix}0.2602\\ 0.0212\\ -0.2443\end{bmatrix} + \begin{bmatrix}0.1128\\ -0.0235\\ -0.0457\end{bmatrix} = \begin{bmatrix}0.3445\\ 0.0063\\ -0.3164\end{bmatrix}$$

Equation [7.4] is useful in explaining the evaluations for animals in terms of contributions from different sources of information, and how these contributions vary with different DIM could also be examined. However, equation [7.4] relates to random regression coefficients. Usually the estimated breeding value at a particular stage of the longitudinal scale, such as 305 days for milk yield or body weight at 1 year of age, is published. Therefore the interest might be in calculating the contributions from the various sources of information to the published EBV. Using milk yield as an example, the contribution to 305-day estimated BV from various sources of information can be calculated as:

$$\begin{aligned}\hat{\mathbf{u}}_{(305)anim} &= \mathbf{V}_1\mathbf{PA} + \mathbf{V}_2\mathbf{YD} + \mathbf{V}_3\mathbf{PC}\\ &= \mathbf{PA}^* + \mathbf{YD}^* + \mathbf{PC}^*\end{aligned} \qquad [7.5]$$

where $\mathbf{V}_i = \mathbf{DW}_i$, with $\mathbf{D}$ being a diagonal matrix such that $d_{ii} = t_i$, with t_i being the element of the row vector $\mathbf{t}$ in [7.2], $\mathbf{PA}^* = \mathbf{V}_1\mathbf{PA}$, $\mathbf{YD}^* = \mathbf{V}_2\mathbf{YD}$ and $\mathbf{PC}^* = \mathbf{V}_3\mathbf{PC}$, and $\mathbf{W}_i$ as defined in [7.4]. However, $\mathbf{V}_1 + \mathbf{V}_2 + \mathbf{V}_3 \neq \mathbf{I}$. Thus the estimated BV at 305 days ($BV_{(305)anim}$) from [7.5] is:

$$BV_{(305)anim} = \sum_{i=1}^{nr} \hat{u}_{(305)anim} = \sum_{i=1}^{nr} PA_i^* + \sum_{i=1}^{nr} YD_i^* + \sum_{i=1}^{nr} PC_i^*$$

where the contributions to the estimated BV at 305 days from PA, YD and PC are:

$\sum_{i=1}^{nr} PA_i^*$, $\sum_{i=1}^{nr} YD_i^*$ and $\sum_{i=1}^{nr} PC_i^*$, respectively

Using [7.5], the contributions from various sources of information can be calculated for EBV at days or ages *j* to *n* along the longitudinal scale and this could be plotted to examine how the contributions vary with days or age.

Using cow 6 in Example 7.2, the matrix **D** used in calculating the **V** terms in [7.5] is:

$$\mathbf{D} = \text{diag}(215.6655,\ 2.4414,\ -1.5561)$$

Using the $\mathbf{W}_1$ and $\mathbf{W}_2$ calculated earlier for cow 6:

$$\mathbf{V}_1 = \mathbf{D}\mathbf{W}_1 = \begin{pmatrix} 132.7637 & 41.8391 & 64.4193 \\ 0.2283 & 1.9792 & 0.5864 \\ -0.1828 & -0.0314 & -1.1327 \end{pmatrix}$$

$$\mathbf{V}_2 = \mathbf{D}\mathbf{W}_2 = \begin{pmatrix} 82.9018 & -41.8391 & -64.4193 \\ -0.2283 & 0.4622 & -0.5864 \\ 0.1828 & 0.0314 & -0.4234 \end{pmatrix}$$

$$\hat{\mathbf{u}}_{(305)6} = \mathbf{V}_1\mathbf{P}\mathbf{A} + \mathbf{V}_2\mathbf{Y}\mathbf{D} = \begin{pmatrix} -26.3320 \\ 0.0138 \\ 0.1178 \end{pmatrix} + \begin{pmatrix} -91.9351 \\ 0.1649 \\ -0.2160 \end{pmatrix}$$

Therefore contributions from **PA** and **YD** are −26.4049 and −91.9862, respectively. Then:

$$BV_{(305)6} = -26.4049 + -91.9862 = -118.4$$

Thus contribution from parent average is about 22% of the EBV at 305 days.

7.2.3 Calculating daughter yield deviations

The equation for calculating daughter yield deviation under an RRM is the same as [5.12] presented for the multivariate models. However, with the RRM, **DYD** in [5.12] is a vector of random regression coefficients and the weights $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{M}_3$ are of the order *nr*. Actual daughter yield deviation (DYD) for any DIM can be generated using equation [7.2].

As indicated in Section 5.1, for ease of computation, $\mathbf{W}_{2prog}$ in [5.12] is pre-multiplied with $\mathbf{G}^{-1}$, such that the equation for **DYD** becomes:

$$\mathbf{DYD} = \sum \mathbf{G}^{-1}\mathbf{W}_{2prog}\alpha_{prog}(2\mathbf{YD} - \hat{\mathbf{u}}_{mate})/\mathbf{G}^{-1}\mathbf{W}_{2prog}\alpha_{prog}$$

7.2.4 Reliability of breeding values

The reliability of an estimated BV depends on its prediction error variance (PEV) relative to the genetic variance. It can therefore be regarded as a statistic summarizing the value of information available in calculating the estimated BV. The published estimated BV from an RR model is

usually a linear function of the random regression coefficients obtained by solving the MME. The principles for calculating PEV and reliability under this situation are presented, using the diagonal elements of the inverse of the coefficient matrix of the MME for Example 7.2.

Let $\mathbf{k}'\mathbf{u}_i$ define the estimated BV for the trait of interests for animal i from the RR model. The vector $\mathbf{k} = \mathbf{w}_i\mathbf{t}$, where $\mathbf{w}_i$ might be the weighting factor for the ith age or lactation if the study was on body weight at several ages or fat yield in different lactations analysed as different traits. For instance, if fat yields in lactations 1 and 2 were analysed as different traits, $\mathbf{w}'_i$ might be [0.7 0.3], indicating a weight of 0.7 and 0.3, respectively, for first and second lactation estimated BV. The vector **t** defines how within lactation estimated BV was calculated and is the same as in [7.2]. For Example 7.2, $\mathbf{w}_i$ is a scalar with a value of 1. Given that **G** is the additive genetic covariance matrix for random regression effect for animal effects and **P** is the covariance matrix for pe effects, then the additive genetic variance of $\mathbf{k}'\mathbf{u} = g = \mathbf{k}'\mathbf{G}\mathbf{k}$ and the variance for the pe effect for the trait of interest $\mathbf{k}'\mathbf{pe} = p = \mathbf{k}'\mathbf{P}\mathbf{k}$. The heritability of $\mathbf{k}'\mathbf{u}$ can therefore be calculated as g/(g + p + e) and $\alpha = (4 - h^2)/h^2$.

Let $\mathbf{C}^{ii}$ be the subset of the inverse of MME corresponding to the genetic effect for the ith animal. Then, for animal i, prediction error variance $(PEV_i) = \mathbf{k}'\mathbf{C}^{ii}\mathbf{k}$. The reliability of $\mathbf{k}'\mathbf{u}$ can therefore be calculated as $(1 - \text{PEV}_i)/g$. As an illustration, in Example 7.2, $\mathbf{k}' = \mathbf{wt}$ = [215.6655 2.4414 −1.5561], g = $\mathbf{k}'\mathbf{G}\mathbf{k}$ = 154896.766 kg^2, p = $\mathbf{k}'\mathbf{P}\mathbf{k}$ = 323462.969 kg^2 and h^2 = 0.32. For animal 1, the matrix $\mathbf{C}^{11}$ is:

$$\mathbf{C}^{11} = \begin{bmatrix} 2.9911 & 0.5159 & -1.2295 \\ 0.5159 & 0.8683 & -0.2480 \\ -0.2295 & -0.2480 & 0.9183 \end{bmatrix}$$

and:

$$\text{PEV}_1 = \mathbf{k}'\mathbf{C}^{11}\mathbf{k} = 140499.97$$

Therefore reliability for animal 1 equals 1 − 140499.97/154896.766 = 0.09. The reliabilities for the animals in Example 7.2 are as follows:

Animal	Reliability
1	0.09
2	0.04
3	0.07
4	0.12
5	0.15
6	0.06
7	0.10
8	0.05

In practice, calculating the inverse of the MME is not feasible for large populations and PEV has to be approximated. As indicated earlier, estimated BV from RR models are linear functions of the random regressions; therefore methods to approximate reliabilities should simultaneously approximate PEV and the prediction error covariance (PEC) among the individual random regressions (Liu *et al.*, 2002; Meyer and Tier, 2003). Such an approximation method, presented by Meyer and Tier (2003), is outlined in Appendix D.2.

7.2.5 Random regression model for maternal traits

Maternal genetic effects are important in growth traits in beef cattle, and models that account for these effects have been discussed in Chapter 6. However, the RR model could also be augmented to include random regressions for maternal genetic and maternal permanent environmental effects. Albuquerque and Meyer (2001) examined different orders of fit for the random regressions for both effects. One of the favoured models was the one in which the order of Legendre polynomials for direct genetic, maternal genetic, animal pe and maternal pe effects were 5, 5, 5 and 3, respectively. Such a model, excluding all fixed effects, could be written as:

$$y_{ijktd} = \sum_{i=0}^{k_1-1} \phi_{jti} u_{ji} + \sum_{i=0}^{k_2-1} \phi_{jti} m_{ji} + \sum_{i=0}^{k_3-1} \phi_{jti} pe_{ji} + \sum_{i=0}^{k_4-1} \phi_{dti} pp_{di} + e_{ijktd}$$

where y_{ijktd} is the body weight of cow j taken at age t that has a dam d; u_{ji}, m_{ji} and pe_{ji} are random regressions for direct, maternal genetic and animal pe effects for animal j, respectively; pp_{di} is the random regression for dam pe effects and e_{ijktd} is random error, ϕ_{jti} and ϕ_{dti} are the ith Legendre polynomial for body weight at age t for cow j and dam d, respectively. Albuquerque and Meyer (2001) assumed a zero covariance between direct and maternal genetic effects to simplify the computation. The variance for direct effects increased from birth to 365 days, while maternal genetic variance increased from birth to about 115 days and decreased thereafter.

7.3 Covariance Functions

Kirkpatrick *et al.* (1990, 1994) introduced the concept of analysing repeated records taken along a trajectory such as time or age by means of covariance functions. In view of the fact that such a trait can take on a value at each of an infinite number of ages and its value at each age can be regarded as a distinct trait, the trajectory for such a trait could be regarded as an infinite-dimensional trait. Thus the growth trajectory or milk yield trajectory of an individual could be represented by a continuous function. Covariance function describes the covariance structure of an infinite-dimension character as a function of time. Therefore the covariance

function is the infinite-dimension equivalent of a covariance matrix for a given number of records taken over time at different ages. The value of the phenotypic covariance function, $P(t_i, t_j)$, gives the phenotypic covariance between the value of the trait at ages t_i and t_j. Similarly, the value of the additive genetic covariance function, $f(t_i, t_j)$ gives the additive genetic covariance between the value of the trait at ages t_i and t_j. In mathematical terms, given t ages, the covariance between breeding values u_l and u_m on an animal at ages a_l and a_m could be written as:

$$\text{cov}(u_l, u_m) = f(a_i, a_m) = \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \phi_i(a_l)\phi_j(a_m)C_{ij} \quad [7.6]$$

$$= \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \tau_{ij} a_l^i a_m^j \quad [7.7]$$

where f with factors τ_{ij} is the covariance function (CF), $\mathbf{C}$ is the coefficient matrix associated with the CF with elements C_{ij}, a_l is the lth age standardized to the intervals for which the polynomials are defined and k is the order of fit. Kirkpatrick *et al.* (1990, 1994) used Legendre polynomials, which span the interval –1 to +1. The ages can be standardized as described in Appendix G.

Given that $\mathbf{G}$ is the observed genetic covariance matrix of order t, and assuming a full-order polynomial fit, ($t = k$), [7.6] can be written in matrix notation as:

$$\hat{\mathbf{G}} = \mathbf{\Phi}\,\hat{\mathbf{C}}\,\mathbf{\Phi}' \quad [7.8]$$

and $\hat{\mathbf{C}}$ can be estimated as:

$$\hat{\mathbf{C}} = \mathbf{\Phi}^{-1}\hat{\mathbf{G}}(\mathbf{\Phi}^{-1}) \quad [7.9]$$

where $\mathbf{\Phi}$ is the matrix of Legendre polynomials of order t by k with element $\phi_{ij} = \phi_{j(a_i)}$ = the jth polynomial evaluated at standardized age t.

As an illustration, assume body weight measurements in beef cattle have been taken at three different ages – 90, 160 and 240 months old – and that the genetic covariance matrix ($\hat{\mathbf{G}}$) estimated was:

$$\hat{\mathbf{G}} = \begin{bmatrix} 132.3 & 127.0 & 136.6 \\ 127.0 & 172.8 & 200.8 \\ 136.6 & 200.8 & 288.0 \end{bmatrix}$$

Using the method described in Appendix G, the vector of standardized ages is:

$$\mathbf{a}' = [-1.0 \;\; -0.0667 \;\; 1.000]$$

and $\mathbf{M}$ becomes:

$$\mathbf{M} = \begin{bmatrix} 1.0000 & -1.0000 & 1.0000 \\ 1.0000 & -0.0667 & 0.0044 \\ 1.0000 & 1.0000 & 1.0000 \end{bmatrix}$$

Thus for $t = 3$, $\mathbf{\Lambda}$ (see Appendix G) is:

$$\mathbf{\Lambda} = \begin{bmatrix} 0.7071 & 0.0000 & -0.7906 \\ 0.0000 & 1.2247 & 0.0000 \\ 0.0000 & 0.0000 & 2.3717 \end{bmatrix}$$

and $\mathbf{\Phi}$ is:

$$\mathbf{\Phi} = \begin{bmatrix} 0.7071 & -1.2247 & 1.5811 \\ 0.7071 & -0.0816 & -0.7801 \\ 0.7071 & 1.2247 & 1.5811 \end{bmatrix}$$

and from [7.9], the coefficient matrix $\hat{\mathbf{C}}$ is:

$$\hat{\mathbf{C}} = \begin{bmatrix} 344.7117 & 45.2787 & -3.2062 \\ 45.2787 & 24.5185 & -0.1475 \\ -3.2062 & -0.1475 & 3.2768 \end{bmatrix}$$

The covariance between two different ages can be calculated using [7.8]. For instance, the variances at days 90 and 200 of body weight and the covariance between body weight on both days are $\mathbf{\Phi}_{90}\hat{\mathbf{C}}\mathbf{\Phi}'_{90} = 132.30$, $\mathbf{\Phi}_{200}\hat{\mathbf{C}}\mathbf{\Phi}'_{200} = 218.50$, $\mathbf{\Phi}_{90}\hat{\mathbf{C}}\mathbf{\Phi}'_{200} = 129.71$, respectively, with:

$$\mathbf{\Phi}_{90} = \mathbf{m}_{90}\mathbf{\Lambda} = [0.7071 \quad -1.2247 \quad 1.5811]$$

and:

$$\mathbf{\Phi}_{200} = \mathbf{m}_{200}\mathbf{\Lambda} = [0.7071 \quad 0.5716 \quad -0.2740]$$

where $\mathbf{m}_i$ are the appropriate row vectors of the matrix $\mathbf{M}$.

Also from [7.8] and Appendix G, $\hat{\mathbf{G}}$ can be written as:

$$\hat{\mathbf{G}} = \mathbf{M}\mathbf{\Lambda}\hat{\mathbf{C}}\mathbf{\Lambda}'\mathbf{M}'$$

Therefore $\hat{\mathbf{G}} = \mathbf{MTM}'$ with $\mathbf{T} = \mathbf{\Lambda}\hat{\mathbf{C}}\mathbf{\Lambda}$. Then $\mathbf{T}$ can be calculated as $\mathbf{T} = \mathbf{M}^{-1}\hat{\mathbf{G}}(\mathbf{M}^{-1})'$, where $\mathbf{T}$ is the matrix with elements τ_{ij} in [7.7]. Substituting $\mathbf{T}$ in [7.7] the full estimate of the CF, $f(a_l, a_m)$, can be obtained. Using the example data:

$$\mathbf{T} = \begin{pmatrix} 177.99 & 39.35 & -11.52 \\ 39.35 & 36.78 & -0.43 \\ -11.52 & -0.43 & 18.43 \end{pmatrix}$$

Therefore the full estimate of the covariance function, $f(a_l, a_m)$, is:

$$f(a_l, a_m) = 177.99 + 39.35\,(a_l + a_m) + 36.78\,a_l a_m - 11.52\,(a_l^2 + a_m^2) - 0.43\,(a_l^2 a_m + a_l a_m^2) + 18.43\,a_l^2 a_m^2$$

The application of CF in genetic evaluation involves defining an equivalent model using [7.8]. For instance, using the example of the body

weight of beef cattle, assume that the multivariate model for observations measured on one animal is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{a} + \mathbf{e}$$

where **y**, **X**, **b**, **a** and **e** are matrices and vectors defined as in equation [5.1], with $t = i$, var $(\mathbf{a}) = \breve{\mathbf{G}}$ and var$(\mathbf{e}) = \mathbf{R}$. Assuming a CF has also been fitted for the covariance matrix for environmental effects with a term included to account for measurement error, then:

$$\mathbf{R} = \mathbf{\Phi}\,\mathbf{C}_p\mathbf{\Phi}' + \mathbf{I}\sigma^2\varepsilon$$

where $\mathbf{C}_p$ contains the coefficient matrix associated with the CF for pe and variance ε is $\mathbf{I}\sigma^2\varepsilon$. Using this equation and [7.8], an equivalent model to the multivariate model can be written as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{\Phi u} + \mathbf{\Phi pe} + \boldsymbol{\varepsilon}$$

where $\mathbf{a} = \mathbf{\Phi u}$; and **u** and **pe** are now vectors of random regression coefficients for random animal and pe effects. Then var$(\mathbf{u}) = \mathbf{\Phi}\,\mathbf{C}\,\mathbf{\Phi}'$ (see [7.8]) and var$(\mathbf{pe}) = \mathbf{\Phi}\,\mathbf{C}'_p\,\mathbf{\Phi}'$. The application of the above model in genetic evaluation is as illustrated in Example 7.2. Thus the breeding value a_n for any time n can be calculated (see [7.2]) as:

$$a_n = \sum_{i=0}^{k-1} \phi_i(t_n)\,\mathbf{u}_i$$

where $\phi(t_n)$ is the vector of Legendre polynomial coefficients evaluated at age t_n.

Thus, with a full-order fit, the covariance function model is exactly equivalent to the multivariate model. However, in practice, the order of fit is chosen such that estimated covariance matrix can be appropriately fitted with as few parameters as possible. In the next section, the fitting of a reduced-order CF is discussed.

7.3.1 Fitting a reduced-order covariance function

Equation [7.8] and the illustration given in the above section assumed a full-order polynomial fit of **G** ($k = t$). Therefore it was possible to get an inverse of **Φ** and hence estimate **C**. However, for a reduced-order ($k < t$) fit, **Φ** has only k columns and a direct inverse may not be possible. With the reduced fit, the number of coefficients to be estimated are reduced to $k(k + 1)/2$. This is particularly important for large **Λ**, such as test day milk yield within a lactation with t equal to 10 or 305 assuming monthly or daily sampling and requiring $t(t + 1)/2$ coefficients to be estimated. Thus a reduced-order fit with k substantially lower than t could be very beneficial.

Kirkpatrick *et al.* (1990) proposed weighted least squares as an efficient method to obtain an estimate of the reduced coefficient matrix ($\breve{\mathbf{C}}$) from the linear function of the elements of $\breve{\mathbf{G}}$. They outlined the following steps for the weighted least-square procedure. The procedure is

illustrated using the example $\breve{\mathbf{G}}$ for the body weight in beef cattle given earlier, fitting polynomials of order one; that is, only the first two Legendre polynomials are fitted, thus $k = 2$. Initially, a vector $\breve{\mathbf{g}}$ of order t^2 is formed by stacking the successive columns of $\breve{\mathbf{G}}$. Thus:

$$\breve{\mathbf{g}}' = [\breve{G}_{11}, \ldots, \breve{G}_{n1},\ \breve{G}_{12}, \ldots, \breve{G}_{n2},\ \breve{G}_{1n}, \ldots, \breve{G}_{nn}]$$

Thus, for the example $\breve{\mathbf{G}}$:

$$\breve{\mathbf{g}}' = [132.3 \quad 127.0 \quad 136.6 \quad 127.0 \quad 172.8 \quad 200.8 \quad 136.6 \quad 200.8 \quad 288.0]$$

Define $\mathbf{\Phi}_r$ of order t by k, obtained by deleting $(t - k)$ columns of $\mathbf{\Phi}$ corresponding to those ϕ_j not in the reduced-order fit. The relationship between the observed covariance matrix, $\breve{\mathbf{g}}$, and the coefficient matrix of the reduced fit to be estimated is given by the following regression equation:

$$\breve{\mathbf{g}} = \mathbf{X}_s \breve{\mathbf{c}} + \mathbf{e} \qquad [7.10]$$

where $\mathbf{e}$ is the vector of difference between observed covariances and those predicted by the covariance function, $\hat{\mathbf{c}}$ is a vector of dimension k^2, containing the elements of the coefficient matrix of the reduced fit ($\breve{\mathbf{C}}$). The order of elements of $\breve{\mathbf{C}}$ in $\breve{\mathbf{c}}$ is the same as in $\breve{\mathbf{g}}$; that is, $\breve{\mathbf{c}}' = [\breve{C}_{00}, \ldots, \breve{C}_{k0}, \ldots, \breve{C}_{1k}, \ldots, \breve{C}_{kk}]$. $\mathbf{X}_s$ is the Kronecker product of $\mathbf{\Phi}_r$ with itself ($\mathbf{X}_s = \mathbf{\Phi}_r \otimes \mathbf{\Phi}_r$) and is of the order t^2 by k^2. Since only the first two polynomials are fitted, the matrix $\mathbf{\Phi}_r$ can be derived by deleting from $\mathbf{\Phi}$ the third column, corresponding to the missing second-degree polynomial. Thus for the beef cattle example:

$$\mathbf{\Phi}_r = \begin{bmatrix} 0.7071 & -1.2247 \\ 0.7071 & -0.0816 \\ 0.7071 & 1.2247 \end{bmatrix}$$

and $\mathbf{X}_s$ is:

$$\mathbf{X}_s = \begin{bmatrix} 0.5000 & -0.8660 & -0.8660 & 1.4999 \\ 0.5000 & -0.0577 & -0.8660 & 0.0999 \\ 0.5000 & 0.8660 & -0.8660 & -1.4999 \\ 0.5000 & -0.8660 & -0.0577 & 0.0999 \\ 0.5000 & -0.0577 & -0.0577 & 0.0067 \\ 0.5000 & 0.8660 & -0.0577 & -0.0999 \\ 0.5000 & -0.8660 & 0.8660 & -1.4999 \\ 0.5000 & -0.0577 & 0.8660 & -0.0999 \\ 0.5000 & 0.8660 & 0.8660 & 1.4999 \end{bmatrix}$$

The application of weighted least squares to obtain solutions for $\breve{\mathbf{c}}$ in equation [7.10] requires the covariance matrix ($\mathbf{V}$) of sampling errors of $\breve{\mathbf{g}}$. Kirkpatrick *et al.* (1990) presented several methods for estimating $\mathbf{V}$,

examining three different experimental designs. However, in animal breeding, most estimates of $\breve{\mathbf{G}}$ are from field data and may not fit strictly to the designs they described, but estimates of sampling variances from restricted maximum likelihood (REML) analysis could be used. For the example $\breve{\mathbf{G}}$ for the beef cattle data, **V** has been estimated using the formula given by Kirkpatrick *et al.* (1990) for a half-sib design, assuming that 60 sires were each mated to 20 dams. The mean cross-product for the residual effect ($\hat{W}_e$) was estimated as $\hat{W}_{e,ij} = P_{ij} - 0.25\,\breve{G}_{ij}$ and that among sires ($\hat{W}_a$) as $\hat{W}_{a,ij} = (n - 1/4)\breve{G}_{ij} + P_{ij}$, where P_{ij} is the phenotypic variance and n is the number of dams. Sampling variance for $\breve{\mathbf{g}}$ was then calculated as: $\mathbf{V} = (16/\mathrm{n}^2)\ [\mathrm{cov}(\hat{W}_{a,ij}, \hat{W}_{a,kl}) + \mathrm{cov}(\hat{W}_{e,ij}, \hat{W}_{e,kl})]$, where $\mathrm{cov}(\hat{W}_{ij}, \hat{W}_{kl}) = (\hat{W}_{ik}\hat{W}_{jl} + \hat{W}_{il}\,\hat{W}_{jk})/\mathrm{df}$, with df = number of degrees of freedom plus 2. In estimating $\mathrm{cov}(\hat{W}_{a,ij}, \hat{W}_{a,kl})$ and $\mathrm{cov}(\hat{W}_{e,ij}, \hat{W}_{e,kl})$, df = $(s - 1) + 2$ and $s(n - 1) + 2$, respectively. The estimated **V** therefore is:

$$\hat{\mathbf{V}} = \begin{bmatrix}
3450.0 & 2256.4 & 2184.6 & 2256.4 & 1480.3 & 1434.7 & 2184.6 & 1434.7 & 1390.9 \\
2256.4 & 2959.6 & 2430.9 & 2959.6 & 2903.5 & 2530.2 & 2430.9 & 2530.2 & 2180.1 \\
2184.6 & 2430.9 & 3889.7 & 2430.9 & 2249.1 & 3181.6 & 3889.7 & 3181.6 & 4051.5 \\
2256.4 & 2959.6 & 2430.9 & 2959.6 & 2903.5 & 2530.2 & 2430.9 & 2530.2 & 2180.1 \\
1480.3 & 2903.5 & 2249.1 & 2903.5 & 5711.4 & 4410.0 & 2249.1 & 4410.0 & 3417.5 \\
1434.7 & 2530.2 & 3181.6 & 2530.2 & 4410.0 & 5818.8 & 3181.6 & 5818.8 & 6354.3 \\
2184.6 & 2430.9 & 3889.7 & 2430.9 & 2249.1 & 3181.6 & 3889.7 & 3181.6 & 4051.5 \\
1434.7 & 2530.2 & 3181.6 & 2530.2 & 4410.0 & 5818.8 & 3181.6 & 5818.8 & 6354.3 \\
1390.9 & 2180.1 & 4051.5 & 2180.1 & 3417.5 & 6354.3 & 4051.5 & 6354.3 & 11835.0
\end{bmatrix}$$

However, the symmetry of $\breve{\mathbf{G}}$ resulted in redundancies in the vector $\breve{\mathbf{g}}$ such that **V** is singular. The vector $\breve{\mathbf{g}}$ can be redefined to be of the order s by 1, which contains only the elements in the lower half of $\breve{\mathbf{G}}$, where $s = t(t + 1)/2$. Therefore delete from $\breve{\mathbf{g}}$ the elements $\breve{\mathbf{G}}_{ij}$ for which $i < j$. Thus, for the example $\breve{\mathbf{G}}$, the vector $\breve{\mathbf{g}}$ becomes:

$$\breve{\mathbf{g}} = [132.3 \quad 127.0 \quad 136.6 \quad 172.8 \quad 200.8 \quad 288.0]$$

Then delete from **V** those columns and rows corresponding to elements $\breve{\mathbf{G}}_{ij}$ with $i < j$. This involves deleting rows and columns 4, 7 and 8 from the matrix **V** given above. The **V** of reduced order (s by s) is:

$$\hat{\mathbf{V}} = \begin{bmatrix}
3450.0 & 2256.4 & 2184.6 & 1480.3 & 1434.7 & 1390.9 \\
2256.4 & 2959.6 & 2430.9 & 2903.5 & 2530.2 & 2180.1 \\
2184.6 & 2430.9 & 3889.7 & 2249.1 & 3181.6 & 4051.5 \\
1480.3 & 2903.5 & 2249.1 & 5711.4 & 4410.0 & 3417.5 \\
1434.7 & 2530.2 & 3181.6 & 4410.0 & 5818.8 & 6354.3 \\
1390.9 & 2180.1 & 4051.5 & 3417.5 & 6354.3 & 11835.0
\end{bmatrix}$$

Similarly, the rows corresponding to those elements of $\breve{\mathbf{g}}$ for which $\breve{\mathbf{G}}_{ij}$ has $i < j$ are deleted from $\mathbf{X}_s$. In the example $\mathbf{X}_s$, rows 4, 7 and 8 are deleted. Thus $\mathbf{X}_s$ becomes:

$$\mathbf{X}_s = \begin{bmatrix} 0.5000 & -0.8660 & -0.8660 & 1.4999 \\ 0.5000 & -0.0577 & -0.8660 & 0.0999 \\ 0.5000 & 0.8660 & -0.8660 & -1.4999 \\ 0.5000 & -0.0577 & -0.0577 & 0.0067 \\ 0.5000 & 0.8660 & -0.0577 & -0.0999 \\ 0.5000 & 0.8660 & 0.8660 & 1.4999 \end{bmatrix}$$

Also for each element of $\hat{\mathbf{c}}$ for which $\breve{\mathbf{C}}_{ij}$ has $i < j$, add the corresponding column of $\mathbf{X}_s$ to the column corresponding to $\breve{\mathbf{C}}_{ji}$ and then delete the former column. For the beef cattle example, the vector of coefficients, $\breve{\mathbf{c}}' = [\breve{C}_{00}\ \breve{C}_{10}\ \breve{C}_{01}\ \breve{C}_{11}]$. Therefore, the third column of $\mathbf{X}_s$ corresponding to $\hat{\mathbf{C}}_{01}$ is added to the second column and the third column is deleted. The matrix $\mathbf{X}_s$ then becomes:

$$\mathbf{X}_s = \begin{bmatrix} 0.5000 & -1.7320 & 1.4999 \\ 0.5000 & -0.9237 & 0.0999 \\ 0.5000 & 0.0000 & -1.4999 \\ 0.5000 & -0.1154 & 0.0067 \\ 0.5000 & 0.8083 & -0.0999 \\ 0.5000 & 1.7320 & 1.4999 \end{bmatrix}$$

Finally, delete from $\breve{\mathbf{c}}$ the elements for which $\breve{\mathbf{C}}$ has elements $i < j$. The vector $\hat{\mathbf{c}}$ now has $k(k+1)/2$ elements. For the example data, $\breve{\mathbf{c}}' = [\breve{C}_{00}\ \breve{C}_{10}\ \breve{C}_{11}]$. The vector $\breve{\mathbf{c}}$ can now be calculated by a weighted least-square procedure as:

$$\breve{\mathbf{c}} = (\mathbf{X}_s' \hat{\mathbf{V}}^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s' \hat{\mathbf{V}}^{-1} \breve{\mathbf{g}}$$

For the example data, $\breve{\mathbf{c}}$ calculated using the above equation is:

$$\breve{\mathbf{c}}' = [341.8512\ \ 45.0421\ \ 24.5405]$$

The reduced coefficient matrix $\breve{\mathbf{C}}$ is then constructed from the calculated $\breve{\mathbf{c}}$. Then a row and column of zeros are inserted in positions corresponding to those polynomials not included in to obtain $\breve{\mathbf{C}}$. For the example data, $\breve{\mathbf{C}}$ is now:

$$\begin{bmatrix} 341.8512 & 45.0421 & 0.0 \\ 45.0421 & 24.5405 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}$$

Kirkpatrick *et al*. (1990) presented the following chi-square statistic to test the goodness of fit of the reduced covariance function to $\breve{\mathbf{G}}$:

$$\chi^2(m-p) = (\breve{\mathbf{g}} - \mathbf{X}_s \breve{\mathbf{c}})' \hat{\mathbf{V}}^{-1} (\breve{\mathbf{g}} - \mathbf{X}_s \breve{\mathbf{c}})$$

where $m = t(t + 1)/2$ is the number of degrees of freedom in $\breve{\mathbf{G}}$ and $p = (k(k + 1))/2$ is the number of parameters being fitted. A significant result indicates that the model is inconsistent with the data and a higher order of fit may be needed. For the beef cattle example, the value of χ^2 was 0.2231 with $m = 6$ and $p = 3$. This value of χ^2 was not significant with three degrees of freedom and thus the reduced covariance function was not significantly different from $\breve{\mathbf{G}}$.

Another method of fitting a reduced-order CF, proposed by Mantysaari (1999), involved eigenvalue decomposition of the coefficient matrix. The largest k eigenvalues of $\hat{\mathbf{C}}$ in [7.9], for instance, are kept in a diagonal matrix ($\mathbf{D}_a$) and the matrix $\mathbf{\Phi}$ is replaced by the k corresponding eigenfunctions. Thus $\hat{\mathbf{G}}$ in [7.8] can be approximated as:

$$\hat{\mathbf{G}} \approx \mathbf{\Phi}[\mathbf{v}_1\mathbf{v}_2 \cdots \mathbf{v}_k]\mathbf{D}_a[\mathbf{v}_1\mathbf{v}_2 \cdots \mathbf{v}_k]'\mathbf{\Phi}' = \mathbf{T}\mathbf{D}_a\mathbf{T}'$$

where the $\mathbf{v}_i$ are the eigenvectors of $\hat{\mathbf{C}}$ corresponding to eigenvalues in $\mathbf{D}_a$.

Similarly, if CF has been fitted to the environmental covariance matrix, a similar reduction can be carried out, as follows:

$$\begin{aligned}\mathbf{R} &= \mathbf{\Phi}\mathbf{C}_p\mathbf{\Phi} + \mathbf{I}\sigma^2\varepsilon \\ &= \mathbf{\Phi}[\mathbf{v}_1\mathbf{v}_2 \cdots \mathbf{v}_k]\mathbf{D}_p[\mathbf{v}_1\mathbf{v}_2 \cdots \mathbf{v}_k]'\mathbf{\Phi}' + \mathbf{I}\sigma^2\varepsilon = \mathbf{Q}\mathbf{D}_p\mathbf{Q}' + \mathbf{I}\sigma^2\varepsilon\end{aligned} \quad [7.11]$$

where $\mathbf{D}_p$ contains the k largest eigenvalues of $\mathbf{C}_p$. However, Mantysaari (1999) indicated that, with several biological traits, [7.11] could easily lead to a non-positive definite $\mathbf{C}_p$ and the decomposition might not be possible. He used an expectation maximization (EM) algorithm to fit the CF to the environmental covariance matrix. However, if $\mathbf{C}_p$ has been estimated directly using REML (Meyer and Hill, 1997), the EM algorithm would not be necessary and the covariance matrix for pe can be approximated as $\mathbf{Q}\mathbf{D}_p\mathbf{Q}'$. In addition to reducing the number of equations to k per animal in the MME with this method, the system of equations is very sparse since $\mathbf{D}_a$ or $\mathbf{D}_p$ are diagonal.

7.4 Equivalence of the Random Regression Model to the Covariance Function

Meyer and Hill (1997) indicated that the RR model is equivalent to a covariance function model. The equivalence of the RR model fitting either a parametric curve or Legendre polynomials to the CF model is presented below. Similarly to the model in Section 7.2, the RR model with a parametric curve can be represented as:

$$y_{jt} = F_{jt} + \sum_{m=0}^{f-1} z_m(t)\beta_m + \sum_{m=0}^{k-1} z_m(t)\alpha_{jm} + \sum_{m=0}^{k-1} z_m(t)\lambda_{jm} + e_{jt} \quad [7.12]$$

where y_{jt} is the test day record of cow j made on day t; β_m are fixed regression coefficients; α_{jm} and λ_{jm} are the additive genetic and permanent environmental random regressions for cow j; F_{jt} represents the remaining

fixed effects in the model; $z_m(t)$ is the mth parameter of a parametric function of days in milk and e_{jt} is the random error term. For example, in the model of Jamrozik *et al.* (1997), z was a function of days in milk with five parameters: $z = (1\ c\ c^2\ d\ d^2)$, where $c = t/305$ and $d = \ln(1/c)$, with ln being the natural logarithm. Then the covariance between breeding values u_i and u_l on an animal recorded at DIM t_i and t_l is:

$$\text{cov}(u_i, u_l) = f(t_i, t_l) = \sum_{m=0}^{k-1} \sum_{r=0}^{k-1} z_m(t_i) z_r(t_l) \text{cov}(\alpha_m, \alpha_r) \quad [7.13]$$

However, instead of a parametric curve, assume that orthogonal polynomials such as Legendre polynomials were fitted in an RR model, as described in Section 7.2. Let a_i and a_l represent TD records on days t_i and t_l of animal j standardized to the interval –1 to 1, as outlined in Appendix G. Furthermore, assume that the mth Legendre polynomial of a_i be $\phi_m(a_l)$, for $m = 0, \ldots, k - 1$. The covariance between breeding values u_i and u_l on an animal recorded at DIM a_i and a_l could then be represented as:

$$\text{cov}(u_i, u_l) = f(a_i, a_l) = \sum_{m=0}^{k-1} \sum_{r=0}^{k-1} \phi_m(a_i) \phi_r(a_l) \text{cov}(\alpha_m, \alpha_r) \quad [7.14]$$

The right-hand sides of equations [7.13] and [7.14] are clearly equivalent to the right-hand side of equation [7.6], with cov(α_m, α_r) equal to $\mathbf{C}_{ij}$, the ij element of the coefficient matrix of the covariance function. This equivalence of the RR model with the covariance function is useful when analysing data observed at many ages or time periods, as only k regression coefficients and their $(k(k + 1))/2$ covariances need to be estimated for each source of variation in an RR univariate model.

8 Use of Genetic Markers in Prediction of Breeding Values

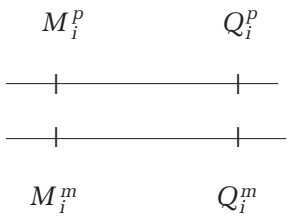
Genetic markers are useful in identifying portions of the chromosomes that are associated with particular quantitative traits. The incorporation of information on marker loci that are linked to quantitative trait loci (QTL), together with phenotypic information in a genetic evaluation procedure, would increase the accuracy of evaluations and therefore of selection. The use of breeding values with marker information incorporated in the selection of animals in a breeding programme is termed marker-assisted selection (MAS). The gains from MAS depend on the amount of genetic variation explained by the marker information and are larger for traits with low heritabilities, and therefore estimated breeding values (BV) from phenotype are of low accuracy (Goddard and Hayes, 2002). Similarly, MAS should result in larger increases in accuracies for traits that are sex-limited, such as milk yield, or measured only in culled animals, for instance, carcass traits. Marker information could be used for the selection or screening of young males before the progeny test in dairy cattle or performance testing in pigs, resulting in the reduction of the generation interval.

Fernando and Grossman (1989) presented a methodology that incorporated marker information into the best linear unbiased prediction (BLUP) procedure for the genetic evaluation of animals. In this chapter, this method is discussed and illustrated. The use of a reduced animal model and other approaches to reduce the number of equations is presented. The extension of the method of Fernando and Grossman (1989) by Goddard (1992) to handle information on QTL bracketed by two markers is examined.

8.1 Defining a Model with Marker Information

Consider a single polymorphic marker locus (ML), which is closely linked to a quantitative trait locus (MQTL). Assume individual i inherited

M_i^p and M_i^m at the ML from its paternal (p) and its maternal (m) parents. Also let Q_i^p and Q_i^m denote alleles at the quantitative trait loci linked to M_i^p and M_i^m, as illustrated below:



Let v_i^p and v_i^m be the genetic additive effects of Q_i^p and Q_i^m, respectively, and u_i the genetic additive effects of the remaining quantitative trait loci not linked to the ML. Then the additive genetic value (a_i) of individual i is:

$$a_i = v_i^p + v_i^m + u_i \quad [8.1]$$

Given only phenotypic information, the usual BLUP equation for additive genetic effects (Section 3.1) is:

$$y_i = x_i\beta + a_i + e_i \quad [8.2]$$

Replacing a_i above by 8.1 gives:

$$y_i = x_i\beta + v_i^p + v_i^m + u_i + e_i \quad [8.3]$$

From Section 2.1, the covariance matrix for u_i, **A**, is the usual relationship matrix (Henderson, 1976), but the covariance for v_i, $\mathbf{G}_v$, depends on both the relationship matrix and marker information. Thus, given **A** and $\mathbf{G}_v$, the BLUP of v_i and u_i can be obtained using the usual mixed model equations (MME). The calculation of **A** and its inverse has been covered in Chapter 3. The calculation of $\mathbf{G}_v$ and its inverse is covered in the next section.

8.2 Calculating the Covariance Matrix ($\mathbf{G}_v$) for MQTL Effects

The matrix $\mathbf{G}_v\sigma_v^2$ represents the covariance between the additive effects of the MQTL alleles. For simplicity, consider only maternal MQTL. Assume two arbitrary individuals b and b' inherit MQTL alleles Q_b^m and $Q_{b'}^m$ with additive effects v_b^m and $v_{b'}^m$ from dams d and d', respectively. The covariance between the additive effects v_b^m and $v_{b'}^m$ for the maternal MQTL in b and b' is:

$$\begin{aligned}\mathrm{cov}(v_b^m, v_{b'}^m) &= \mathrm{cov}(v_b^m, v_{b'}^m \mid Q_b^m \equiv Q_{b'}^m) \cdot P(Q_b^m \equiv Q_{b'}^m) \\ &= \mathrm{var}(v_b^m) \cdot P(Q_b^m \equiv Q_{b'}^m) \\ &= \sigma_v^2 \mathbf{G}_{v(b,b')} \end{aligned} \quad [8.4]$$

where $\mathrm{var}(v_b^m) = \sigma_v^2$ is the variance of the MQTL allele, $P(Q_b^m \equiv Q_{b'}^m)$ is the probability that Q_b^m is identical by descent (IBD) to $Q_{b'}^m$ and the

matrix $\mathbf{G}_{v(b,b')}$ is the covariance matrix for the MQTL between b and b'. Given that b is not a direct descendant of b', Q_b^m can only be identical by descent to $Q_{b'}^m$ in two mutually exclusive manners: (i) if Q_b^m is IBD to $Q_{d'}^p$, the paternal MQTL allele of the dam of b', and b' has inherited $Q_{d'}^p$; or (ii) if Q_b^m is IBD to $Q_{d'}^m$, the maternal MQTL allele of the dam of b', and b' has inherited $Q_{d'}^m$.This is akin to calculating **A** where the relationship, say, between b and b' is evaluated through the relationship of b with the parents of b'.

With marker information available, the conditional probability that b' inherits $Q_{d'}^m$, given that it has inherited $M_{d'}^m$, is $(1-r)$, with r being the recombination rate between the ML and MQTL. Thus, if b' inherits $M_{d'}^m$, the probability in equation [8.4] can be calculated recursively as:

$$P(Q_b^m \equiv Q_{b'}^m) = P(Q_b^m \equiv Q_{d'}^p) \cdot r + P(Q_b^m \equiv Q_{d'}^m) \cdot (1-r) \qquad [8.5]$$

Similarly, given that b' inherits $M_{d'}^p$, then:

$$P(Q_b^m \equiv Q_{b'}^m) = P(Q_b^m \equiv Q_{d'}^p) \cdot (1-r) + P(Q_b^m \equiv Q_{d'}^m) \cdot r \qquad [8.6]$$

If it is not known whether b' inherits $M_{d'}^m$ or $M_{d'}^p$ due to lack of marker information, then $Q_{d'}^m$ and $Q_{d'}^p$ have equal probability of being transmitted to b'. Therefore, r is replaced by 0.5 in equations [8.5] and [8.6].

Using the above information, Fernando and Grossman (1989) developed a tabular method for constructing $\mathbf{G}_v$ which is similar to that for calculating **A**. The rows and columns of $\mathbf{G}_v$ should be such that those for parents precede those for progeny. It should be noted that there are two rows for an individual in $\mathbf{G}_v$, one each for the paternal and maternal MQTL alleles. Let g_{ij} be the ij element of $\mathbf{G}_v$ and i_o^p, i_o^m be the rows of $\mathbf{G}_v$ corresponding to the additive effects of MQTL alleles (v_o^p, v_o^m) of the oth individual. Similarly let i_s^p, i_s^m be the rows for the additive effects of the MQTL alleles (v_s^p, v_s^m) of its sire (s) additive effects and i_d^p, i_d^m be the rows for the effects of the MQTL alleles (v_d^p, v_d^m) of its dam (d). Then the elements of the row i_o^p below the diagonal, using equations [8.4] to [8.6], can be calculated as:

$$g_{i_b^p,j} = (1-\rho_o^p) g_{i_b^p,j} + \rho_o^p g_{i_s^m,j}; \quad \text{for } j = 1,\ldots,i_o^p - 1 \qquad [8.7]$$

with $\rho_o^p = r$ if b inherits M_s^p or $\rho_o^p = (1-r)$ if o inherits M_s^m. Similarly, elements of row i_o^m below the diagonal are:

$$g_{i_o^m,j} = (1-\rho_o^m) g_{i_d^p,j} + \rho_o^m g_{i_d^m,j}; \quad \text{for } j = 1,\ldots,i_o^m - 1 \qquad [8.8]$$

where $\rho_o^m = r$ if o inherits M_d^p or $\rho_o^m = (1-r)$ if o inherits M_d^m. Since $\mathbf{G}_v$ is symmetric, then:

$$g_{j,i_o^p} = g_{i_o^p,j} \quad \text{and} \quad g_{j,i_o^m} = g_{i_o^m,j}$$

It is obvious from [8.4] that, if $o = o'$, that is, the same individual, then $\text{cov}(v_o^m, v_{o'}^m) = \text{var}(v_o^m)$, as $P(Q_o^m \equiv Q_{o'}^m) = 1$. Therefore the diagonal elements

of $\mathbf{G}_v$ equal unity. If it is not possible to determine which of the two marker alleles o inherited from its sire or dam, then ρ_o^p in [8.7] and ρ_o^m in [8.8] are replaced by 0.5.

8.2.1 Numerical application

Example 8.1
Given in the table below are the post-weaning gain data of five calves with the genotype at the marker locus given. The aim at this stage is to construct the covariance matrix $\mathbf{G}_v$ for the MQTL among the five calves assuming a recombination rate (r) between ML and MQTL of 0.1.

				Marker inheritance		
Calf	Sex of calf	Sire	Dam	Sire	Dam	PWG (kg)
1	M	–	–	–	–	6.8
2	F	–	–	–	–	4.5
3	M	1	2	M_1^p	M_2^m	8.5
4	F	1	3	M_1^m	M_3^p	6.0
5	F	4	3	M_4^p	M_3^p	7.0

For ease of illustration, let rows i_o^p and i_o^m for animal o in $\mathbf{G}_v$ be coded as ip and im, respectively. Thus, for example, for animal 1, i_1^p and i_1^m will be coded as $1p$ and $1m$, respectively, for animal 2, i_2^p and i_2^m will be coded as $2p$ and $2m$ and, for animal 5, i_5^p and i_5^m will be coded as $5p$ and $5m$, respectively. The $\mathbf{G}_v$ for the example, therefore, is:

	1*p*	1*m*	2*p*	2*m*	3*p*	3*m*	4*p*	4*m*	5*p*	5*m*
1*p*	1.000	0.000	0.000	0.000	0.900	0.000	0.100	0.810	0.171	0.810
1*m*	0.000	1.000	0.000	0.000	0.100	0.000	0.900	0.090	0.819	0.090
2*p*	0.000	0.000	1.000	0.000	0.000	0.100	0.000	0.010	0.001	0.010
2*m*	0.000	0.000	0.000	1.000	0.000	0.900	0.000	0.090	0.009	0.090
3*p*	0.900	0.100	0.000	0.000	1.000	0.000	0.180	0.900	0.252	0.900
3*m*	0.000	0.000	0.100	0.900	0.000	1.000	0.000	0.100	0.010	0.100
4*p*	0.100	0.900	0.000	0.000	0.180	0.000	1.000	0.162	0.916	0.162
4*m*	0.810	0.090	0.010	0.090	0.900	0.100	0.162	1.000	0.246	0.820
5*p*	0.171	0.819	0.001	0.009	0.252	0.010	0.916	0.246	1.000	0.228
5*m*	0.810	0.090	0.010	0.090	0.900	0.100	0.162	0.820	0.228	1.000

The calculation of $\mathbf{G}_v$ for the first three animals is illustrated as below. For the first two animals, the parents are unknown; therefore:

$$g_{1p,1p} = g_{1m,1m} = g_{2p,2p} = g_{2m,2m} = 0$$

At the ML, animal 3 inherited M_s^p from his father; therefore, for row $3p$ in $\mathbf{G}_v$, corresponding to the effects of the paternal alleles of the MTQL for animal 3, $r = 0.1$. Hence, from equation [8.7]:

$$
\begin{aligned}
g_{3p,1p} &= (1-0.1)g_{1p,1p} + (0.1)g_{1m,1p} = (0.9)1 + (0.1)0 = 0.9\\
g_{3p,1m} &= (1-0.1)g_{1p,1m} + (0.1)g_{1m,1m} = (0.9)0 + (0.1)1 = 0.1\\
g_{3p,2p} &= (1-0.1)g_{1p,2p} + (0.1)g_{1m,2p} = (0.9)0 + (0.1)0 = 0\\
g_{3p,2m} &= (1-0.1)g_{1p,2m} + (0.1)g_{1m,2m} = (0.9)0 + (0.1)0 = 0\\
g_{3p,3p} &= 1.0
\end{aligned}
$$

At the ML, animal 3 inherited M_d^m from his mother; therefore, for row $3m$ in $\mathbf{G}_v$, corresponding to the effects of the maternal alleles of the MQTL for animal 3, $r = 0.9$. Hence, from equation [8.8]:

$$
\begin{aligned}
g_{3m,1p} &= (1-0.9)g_{2p,1p} + (0.9)g_{2m,1p} = (0.1)0 + (0.9)0 = 0\\
g_{3m,1m} &= (1-0.9)g_{2p,1m} + (0.9)g_{2m,1m} = (0.1)0 + (0.9)0 = 0\\
g_{3m,2p} &= (1-0.9)g_{2p,2p} + (0.9)g_{2m,2p} = (0.1)1 + (0.9)0 = 0.1\\
g_{3m,2m} &= (1-0.9)g_{2p,2m} + (0.9)g_{2m,2m} = (0.1)0 + (0.9)1 = 0.9\\
g_{3m,3p} &= (1-0.9)g_{2p,3p} + (0.9)g_{2m,3p} = (0.1)0 + (0.9)0 = 0\\
g_{3m,3m} &= 1.0
\end{aligned}
$$

8.3 An Alternative Approach for Calculating $\mathbf{G}_v$

An alternative recursive method for the calculation of $\mathbf{G}_v$ and its inverse was presented by Van Arendonk *et al*. (1994) using matrix notation. Their method accounts for inbreeding and can be used to calculate a combined numerator relationship matrix ($\mathbf{A}_a$) and its inverse. The matrix $\mathbf{A}_a = \mathbf{A}_u + \mathbf{A}_v$, where $\mathbf{A}_u$ is the numerator relationship matrix for animals for QTL not linked to the marker and $\mathbf{A}_v$ is the relationship matrix for animals for MQTL linked to the marker. The inverse of $\mathbf{A}_a$ is useful for the direct prediction of total additive genetic merit, that is, additive genetic merit with information from markers directly included.

The principles of their methodology are initially illustrated briefly using the calculation of the relationship matrix ($\mathbf{A}$) among animals in the absence of marker information. The representation of the rules for building $\mathbf{A}_i$ for animals 1 to i in matrix form is:

$$
\mathbf{A}_i = \begin{bmatrix} \mathbf{A}_{i-1} & \mathbf{A}_{i-1}\mathbf{s}_i \\ \mathbf{s}_i\mathbf{A}_{i-1} & a_{ii} \end{bmatrix} \qquad [8.9]
$$

where $\mathbf{s}_i$ is the column vector of $i-1$ elements containing two non-zero elements, $\frac{1}{2}$, corresponding to the sire or dam (if known) and zeros elsewhere. $\mathbf{A}_{i-1}$ is the numerator relationship matrix for animals 1 to $(i-1)$ and a_{ii} is the diagonal element of $\mathbf{A}$ for animal i and is equal to $1 + F_i$,

where F_i is the inbreeding coefficient of the ith animal. Using the data in Example 8.1, the **A** matrix, ignoring marker information, is:

$$\mathbf{A} = \begin{bmatrix} 1.000 & 0.000 & 0.500 & 0.750 & 0.625 \\ 0.000 & 1.000 & 0.500 & 0.250 & 0.375 \\ 0.500 & 0.500 & 1.000 & 0.750 & 0.875 \\ 0.750 & 0.250 & 0.750 & 1.250 & 1.000 \\ 0.625 & 0.375 & 0.875 & 1.000 & 1.375 \end{bmatrix}$$

For animal 5, $\mathbf{s}_5' = [0 \quad 0 \quad 0.5 \quad 0.5]$; therefore the column vector above the diagonal for animal 5 ($\mathbf{q}_5$) in **A**, using [8.9], can be calculated as $\mathbf{q}_5 = \mathbf{A}_4 \mathbf{s}_5$. Thus the row vector $\mathbf{q}_5' = \mathbf{s}_5' \mathbf{A}_4 = [0.625 \quad 0.375 \quad 0.875 \quad 1.00]$ and the diagonal element for animal 5, $a_{55} = 1 + 0.5(a_{34}) = 1.375$. Note also, that given $\mathbf{q}_i$, $\mathbf{s}_i$ can be computed as:

$$\mathbf{s}_i = \mathbf{A}_{i-1}^{-1} \mathbf{q}_i \qquad [8.10]$$

This relationship will be used in subsequent sections when it is not possible to calculate $\mathbf{s}_i$ directly.

Given $\mathbf{A}_{i-1}^{-1}$, for animal $i-1$, Tier and Solkner (1993) demonstrated the effect of adding an additional row to **A** on the elements of $\mathbf{A}^{-1}$ as:

$$\mathbf{A}_i^{-1} = \begin{bmatrix} \mathbf{A}_{i-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + (a_{ii} - \mathbf{s}_i' \mathbf{A}_{i-1} \mathbf{s}_i)^{-1} \begin{bmatrix} \mathbf{s}_i \mathbf{s}_i' & -\mathbf{s}_i \\ -\mathbf{s}_i' & 1 \end{bmatrix} \qquad [8.11]$$

When both sire (f) and dam (d) of i are known, $\mathbf{s}_i' \mathbf{A}_{i-1} \mathbf{s} = \frac{1}{4}(a_{ff} + a_{fd} + a_{df} + a_{dd})$ where a_{jj} are the elements of $\mathbf{A}_{i-1}$ for f and d. Since $a_{ii} = (1 + \frac{1}{2} a_{fd})$, then $(a_{ii} - \mathbf{s}_i' \mathbf{A}_{i-1} \mathbf{s})^{-1}$ can be written as $(1 - \frac{1}{4}(a_{ff} + a_{dd}))^{-1}$. The application of [8.11] to calculate $\mathbf{A}^{-1}$ for the pedigree in Example 8.1 is straightforward. For instance, for the first two animals with parents unknown, $\mathbf{A}_2^{-1}$ is an identity matrix of order 2. Then $\mathbf{A}_3^{-1}$ can be calculated using [8.11]. Given that $\mathbf{A}_4^{-1}$ has been calculated, the inverse of **A** for all five animals can be illustrated as follows.

For animal 5, $(a_{55} - \mathbf{s}_5' \mathbf{A}_4 \mathbf{s}_5)^{-1} = (1 - \frac{1}{4}(a_{33} + a_{44}))^{-1} = (1 - \frac{1}{4}(1 + 1.25))^{-1} = 2.286$. Then [8.11] is:

$$\mathbf{A}_5^{-1} = \begin{bmatrix} 2.000 & 0.500 & -0.500 & -1.000 & 0.000 \\ 0.500 & 1.500 & -1.000 & 0.000 & 0.000 \\ -0.500 & -1.000 & 2.500 & -1.000 & 0.000 \\ -1.000 & 0.000 & -1.000 & 2.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix} + (2.286) \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.25 & 0.25 & -0.5 \\ 0.0 & 0.0 & 0.25 & 0.25 & -0.5 \\ 0.0 & 0.0 & -0.5 & -0.5 & 1.0 \end{bmatrix}$$

$$= \begin{bmatrix} 2.000 & 0.500 & -0.500 & -1.000 & 0.000 \\ 0.500 & 1.500 & -1.000 & 0.000 & 0.000 \\ -0.500 & -1.000 & 3.071 & -0.429 & -1.143 \\ -1.000 & 0.000 & -0.429 & 2.571 & -1.143 \\ 0.000 & 0.000 & -1.143 & -1.143 & 2.28 \end{bmatrix}$$

where:

$\mathbf{s}'_5 = (0 \quad 0 \quad 0.5 \quad 0.5)$

Applying equations [8.9], Van Arendonk *et al.* (1994) showed that, when alleles are ordered chronologically, $\mathbf{G}_{v,i}$ can be calculated as:

$$\mathbf{G}_{v,i} = \begin{bmatrix} \mathbf{G}_{v,i-1} & \mathbf{G}_{v,i-1}\mathbf{s}_i \\ \mathbf{s}'_i\mathbf{G}_{v,i-1} & g_{ii} \end{bmatrix} \qquad [8.12]$$

where $\mathbf{s}_i$ is the column vector of $i-1$ elements containing non-zero elements relating allele i to paternal and maternal alleles of parent (if known) and zeros elsewhere; $\mathbf{G}_{v,i-1}$ is the covariance matrix for MQTL for alleles 1 to $(i-1)$ and g_{ii} is the diagonal element of $\mathbf{G}_v$ for the i allele, which is equal to one. Using the same notation for the rows in $\mathbf{G}_v$ shown in Section 8.2.1, $\mathbf{s}_i$ for animals 3, 4 and 5 are: $\mathbf{s}'_{3p} = [(1-r) \;\; r \;\; 0 \;\; 0]$, $\mathbf{s}'_{3m} = [0 \;\; 0 \;\; r \;\; (1-r) \;\; 0]$, $\mathbf{s}'_{4p} = [r \;\; (1-r) \;\; 0 \;\; 0 \;\; 0 \;\; 0]$, $\mathbf{s}'_{4m} = [0 \;\; 0 \;\; 0 \;\; 0 \;\; (1-r) \;\; r \;\; 0]$, $\mathbf{s}'_{5p} = [0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; (1-r) \;\; r]$ and $\mathbf{s}'_{5m} = [0 \;\; 0 \;\; 0 \;\; 0 \;\; (1-r) \;\; r \;\; 0 \;\; 0 \;\; 0]$. Thus $\mathbf{G}_v$ can easily be constructed using [8.12].

8.4 Calculating the Inverse of $\mathbf{G}_v$

Fernando and Grossman (1989) used an approach similar to that for setting up $\mathbf{A}^{-1}$ in calculating the inverse of $\mathbf{G}_v$. They showed that $\mathbf{G}_v$ could be expressed as:

$$\mathbf{G}_v = (\mathbf{Q}^{-1})'\mathbf{H}\mathbf{Q}^{-1}$$

Therefore $\mathbf{G}_v^{-1}$ can be written as:

$$\mathbf{G}_v^{-1} = \mathbf{Q}\mathbf{H}\mathbf{Q}' \qquad [8.13]$$

where $\mathbf{Q} = (\mathbf{I} - \mathbf{P}')$ and $\mathbf{P}$ is a matrix that relates the effect of the MQTL allele of an individual to the paternal and maternal MQTL alleles of its parent. Each row of $\mathbf{P}$ contains only two non-zero elements if the parent is known, otherwise only zeros if the parent is unknown. For instance, for individual i with sire (s) known, row i_o^p will have $(1-\rho_o^p)$ in the column corresponding to i_s^p, and ρ_o^p in the column corresponding to column i_s^m. Similarly, if the dam (d) is known, row i_o^m will contain $(1-\rho_o^m)$ in the column corresponding to i_d^p, and ρ_o^m in the column corresponding to i_d^m. The row of $\mathbf{P}$ for allele i is equal to s_i in equation [8.12]. The matrix $\mathbf{P}$ for the pedigree in Example 8.1 is:

	1*p*	1*m*	2*p*	2*m*	3*p*	3*m*	4*p*	4*m*	5*p*	5*m*
1*p*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1*m*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2*p*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2*m*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3*p*	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3*m*	0.0	0.0	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0
4*p*	0.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4*m*	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0
5*p*	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0
5*m*	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0

The matrix **H** is a diagonal matrix for the covariance of residual effects after adjusting the effect of the MQTL allele of an individual for the effects of the parent's paternal and maternal MQTL alleles. For example, the residual effect (ε_o^p) for a paternal MQTL allele of an individual with sire s known is:

$$(\varepsilon_o^p) = v_o^p - (1-\rho_o^p)v_s^p + \rho_o^p v_s^m$$

and the variance of (ε_o^p) is:

$$\begin{aligned}\mathrm{var}(\varepsilon_o^p) = {} & \mathrm{var}(v_o^p) - (1-\rho_o^p)^2 \cdot \mathrm{var}(v_s^p) - (\rho_o^p)^2 \cdot \mathrm{var}(v_s^m) \\ & -2(1-\rho_o^p)\rho_o^p \cdot \mathrm{cov}(v_s^p, v_s^m)\end{aligned}$$

Since $\mathrm{var}(v_o^p) = \mathrm{var}(v_s^p) = \mathrm{var}(v_s^m) = \sigma_v^2$ and $\mathrm{cov}(v_s^p, v_s^m) = \mathrm{var}(v_s^p) \cdot P(Q_s^p \equiv Q_s^m) = \mathrm{var}(v_s^p) \cdot F_s = \sigma_v^2 F_s$, the above equation can be written as:

$$\begin{aligned}\mathrm{var}(\varepsilon_o^p) &= 2\sigma_v^2(\rho_o^p) - 2\sigma_v^2(\rho_o^p)^2 - 2\sigma_v^2(1-\rho_o^p)\rho_o^p F_s \\ &= 2\sigma_v^2((1-\rho_o^p)\rho_o^p - (1-\rho_o^p)\rho_o^p F_s) \\ &= 2\sigma_v^2(1-\rho_o^p)\rho_o^p\ (1-F_s) \\ \mathrm{var}(\varepsilon_o^p)/\sigma_v^2 &= h_o^p = 2(1-\rho_o^p)\rho_o^p(1-F_s) \qquad [8.14]\end{aligned}$$

where $(1-\rho_o^p)\rho_o^p = (1-r)r$ for $\rho_o^p = r$ or $(1-r)$, F_s is the inbreeding coefficient at the MQTL of the sire and h_o^p is the diagonal element of **H** for the paternal MQTL of individual o. Therefore, if the sire is not inbred, $h_o^p = 2(1-r)r$ with marker information or $h_o^p = 0.5$ with no marker information and $h_o^p = 1$ if the sire is unknown. Similarly, for the maternal MQTL of o:

$$\mathrm{var}(\varepsilon_o^m)/\sigma_v^2 = h_o^m = 2(1-\rho_o^m)\rho_o^m(1-F_d) \qquad [8.15]$$

where $(1-\rho_o^m)\rho_o^m = (1-r)r$ for $\rho_o^m = r$ or $(1-r)$, F_d is the inbreeding coefficient at the MQTL of the dam and h_o^m is the diagonal element of **D** for the paternal MQTL of individual o. Therefore, if the dam is not inbred, $h_o^m = 2(1-r)r$ with marker information or $h_o^m = 0.5$ with no marker information and $h_o^m = \sigma_v^2$ if the dam is unknown.

Equation [8.13] may be written as:

$$\mathbf{G}_v^{-1} = \sum_{j=1}^{n} q_j q_j h_j$$

where n is the number of individuals in the pedigree, q_j is the jth column of **Q** and h_j is the jth diagonal element of **H**. Since $\mathbf{Q} = (1-\mathbf{P}')$, the jth element of q_j (that is, the diagonal element) is unity and q_j has at most only two other non-zero elements. If the sire of o is known, $j = i_o^p$, element $i_s^p = -(1-\rho_o^p)$ and element $i_s^m = -\rho_o^p$. Similarly, if the dam is known, then, for $j = i_o^m$, element $i_d^p = -(1-\rho_o^m)$ and element $i_d^m = -\rho_o^m$. Therefore the contribution corresponding to the paternal and maternal MQTL alleles of an individual to $\mathbf{G}_v^{-1}$ can easily be calculated from parent and marker information.

Fernando and Grossman (1989) gave the following rules for obtaining $\mathbf{G}_V^{-1}$. First, calculate the diagonals of $\mathbf{H}$ using equations [8.14] and [8.15] and its inverse. Secondly, set $\mathbf{G}_V^{-1}$ to zero and for each offspring o, with sire s and dam d, add the following to the indicated elements of $\mathbf{G}_V^{-1}$.

If the sire is known, add:

$(1-\rho_o^p)^2 h_{i_o^p}$ to diagonal element $i_s^p i_s^p$

$-(1-\rho_o^p) h_{i_o^p}$ to elements $i_s^p i_o^p$ and $i_o^p i_s^p$

$(1-\rho_o^p)\rho_o^p h_{i_o^p}$ to elements $i_s^p i_s^m$ and $i_s^m i_s^p$

$(\rho_o^p)^2 h_{i_o^p}$ to diagonal element $i_s^m i_s^m$

$-\rho_o^p h_{i_o^p}$ to elements $i_s^m i_o^p$ and $i_o^p i_s^m$

If the dam is known, add:

$(1-\rho_o^m)^2 h_{i_o^m}$ to diagonal element $i_d^p i_d^p$

$(1-\rho_o^m)\rho_o^m h_{i_o^m}$ to elements $i_d^p i_d^m$ and $i_d^m i_d^p$

$-(1-\rho_o^m) h_{i_o^m}$ to elements $i_d^p i_o^m$ and $i_o^m i_d^p$

$(\rho_o^m)^2 h_{i_o^m}$ to diagonal element $i_d^m i_d^m$

$-\rho_o^m h_{i_o^m}$ to elements $i_d^m i_o^m$ and $i_o^m i_d^m$

And always add:

$h_{i_o^p}$ to element $i_o^p i_o^p$ and $h_{i_o^m}$ to element $i_o^m i_o^m$

Applying these rules, the calculation of the inverse of $\mathbf{G}_V^{-1}$ for the pedigree in Example 8.1 is illustrated. For this pedigree, the matrix $\mathbf{H}$ and its inverse are:

$\mathbf{H}$ = diag(1 1 1 1 0.18 0.18 0.18 0.18 0.1508 0.18) and

$\mathbf{H}^{-1}$ = diag(1 1 1 1 5.556 5.556 5.556 5.556 6.630 5.556)

Note that in calculating the diagonal element for the paternal MQTL of animal 5 ($h_{5p,5p}$), an inbreeding coefficient of 0.162 (covariance between the maternal and paternal MQTL alleles of the sire (animal 4)) has been accounted for. Set $\mathbf{G}_V^{-1}$ with elements represented as $g^{ii,jj}$ to zero and the contribution from the first three animals can be calculated as follows.

For animals 1 and 2, parents are unknown; the diagonal elements are equal to 1 for the MQTL alleles of these animals. Therefore add 1 to $g^{1p,1p}$, $g^{1m,1m}$, $g^{2p,2p}$ and $g^{2m,2m}$, using the same coding as for the rows of $\mathbf{G}_V$ in Section 8.2. For the paternal MQTL allele of animal 3, $\rho_o^p = 0.1$ and $h_{3p,3p}$ equals 5.556. Add $(1-0.1)^2 h_{3p,3p} = 4.50$ to $g^{1p,1p}$, $(1-0.1)0.1(h_{3p,3p}) = 0.5$ to $g^{1p,1m}$, $-(1-0.1)h_{3p,3p} = -5.00$ to $g^{1p,3p}$, $(0.1)^2 h_{3p,3p} = 0.056$ to $g^{1m,1m}$, $(-0.1)h_{3p,3p} = 0.556$ to $g^{1m,3p}$ and $h_{3p,3p}$ to $g^{3p,3p}$. For the maternal allele of

animal 3, $\rho_o^m = 0.9$ and $h_{3m,3m} = 5.556$. Add $(1 - 0.9)^2 h_{3m,3m} = 0.056$ to $g^{2p,2p}$, $(1 - 0.9)0.9(h_{3m,3m}) = 0.5$ to $g^{2p,2m}$, $-(1 - 0.9)h_{3m,3m} = -0.556$ to $g^{2p,3m}$, $(0.9)^2 h_{3m,3m} = 4.50$ to $g^{2m,2m}$, $(-0.9)h_{3m,3m} = -0.500$ to $g^{2m,3m}$ and $h_{3m,3m}$ to $g^{3m,3m}$. Applying the rules to all animals in the pedigree gives $\mathbf{G}_v^{-1}$ as:

	1*p*	1*m*	2*p*	2*m*	3*p*	3*m*	4*p*	4*m*	5*p*	5*m*
1*p*	5.556	1.000	0.000	0.000	–5.000	0.000	–0.556	0.000	0.000	0.000
1*m*	1.000	5.556	0.000	0.000	–0.556	0.000	–5.000	0.000	0.000	0.000
2*p*	0.000	0.000	1.056	0.500	0.000	–0.556	0.000	0.000	0.000	0.000
2*m*	0.000	0.000	0.500	5.500	0.000	–5.000	0.000	0.000	0.000	0.000
3*p*	–5.000	–0.556	0.000	0.000	14.556	1.000	0.000	–5.000	0.000	–5.000
3*m*	0.000	0.000	–0.556	–5.000	1.000	5.667	0.000	–0.556	0.000	–0.556
4*p*	–0.556	–5.000	0.000	0.000	0.000	0.000	10.925	0.597	–5.967	0.000
4*m*	0.000	0.000	0.000	0.000	–5.000	–0.556	0.597	5.622	–0.663	0.000
5*p*	0.000	0.000	0.000	0.000	0.000	0.000	–5.967	–0.663	6.630	0.000
5*m*	0.000	0.000	0.000	0.000	–5.000	–0.556	0.000	0.000	0.000	5.556

Similarly, the inverse of $\mathbf{G}_{v,i}^{-1}$ can be obtained using [8.11] (Van Arendonk *et al*., 1994) as:

$$\mathbf{G}_{v,i}^{-1} = \begin{bmatrix} \mathbf{G}_{v,i-1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + (g_{ii} - \mathbf{s}_i' \mathbf{G}_{v,i-1} \mathbf{s}_i)^{-1} \begin{bmatrix} \mathbf{s}_i \mathbf{s}_i' & -\mathbf{s}_i \\ -\mathbf{s}_i' & 1 \end{bmatrix} \qquad [8.16]$$

The application of [8.16] for the calculation of $\mathbf{G}_v^{-1}$ is briefly illustrated. It has been shown earlier that $\mathbf{G}_v^{-1}$ for the MQTL alleles of the first two animals is an identity matrix of order 4. The matrix $\mathbf{G}_v^{-1}$ with the paternal MQTL allele of animal 3 added can be computed as:

$$\mathbf{G}_{v,3p}^{-1} = \begin{bmatrix} \mathbf{G}_{v,2m}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + (1 - 0.82)^{-1} \begin{bmatrix} 0.81 & 0.09 & 0 & 0 & -0.9 \\ 0.09 & 0.01 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0.9 & -0.1 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 5.500 & 0.500 & 0.0 & 0.0 & -5.000 \\ 0.500 & 1.056 & 0.0 & 0.0 & -0.556 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ -0.500 & -0.556 & 0.0 & 0.0 & 5.556 \end{bmatrix}$$

The computation of $\mathbf{G}_v$ and G_v^{-1} in Sections 8.2 and 8.4 has assumed that paternal and maternal origin of marker alleles can be determined and marker information is complete. When these assumptions do not hold, approximate methods are described in Appendix H.

8.5 Prediction of Breeding Values with Marker Information

The model [8.3] for breeding value prediction with marker information can be written in matrix notation as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Wv} + \mathbf{e} \qquad [8.17]$$

where **y** is the vector of observation, **β** is the vector of fixed effects, **u** is the random vector for additive genetic effects due to loci not linked to ML, **v** is the random vector with allelic effects at the MQTL and **e** is random residual effects. The matrices **X**, **Z** and **W** are incidence matrices. $\text{Var}(u) = \mathbf{A}_u\sigma_u^2$, $\text{var}(v) = \mathbf{G}_v\sigma_v^2$, $\text{var}(e) = \text{I}\sigma_e^2$ and $\text{cov}(u, v) = \text{cov}(u, e) = \text{cov}(v, e) = 0$.

The MME for the above linear model are:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}_u^{-1}\alpha_1 & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \mathbf{G}_v^{-1}\alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\text{v}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix} \qquad [8.18]$$

where:

$\alpha_1 = \sigma_e^2/\sigma_u^2$ and $\alpha_2 = \sigma_e^2/\sigma_v^2$

8.5.1 An illustration

Example 8.2

Using the data for Example 8.1, the breeding value of animals for QTL not linked to ML (simply referred to subsequently as breeding values), additive MQTL effects are predicted for the beef calves and sex effects are estimated. It is assumed that $\sigma_u^2 = 0.3, \sigma_v^2 = 0.05$ and $\sigma_e^2 = 0.6$. Therefore $\alpha_1 = 0.6/0.3 = 2$ and $\alpha_2 = 0.6/0.05 = 12$. The parameters are expressed as a proportion of the phenotypic variance. Note that the total genetic variance $\sigma_a^2 = (\sigma_u^2 + 2\sigma_v^2) = 0.3 + 2(0.05) = 0.40$. Thus 40% of the phenotypic variance is due to additive genetic variance, of which 25% can be explained by the MQTL.

The matrix **X** is formed as discussed in Example 3.1, **Z** is an identity matrix and the matrix **W** is:

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The matrices $\mathbf{A}_u^{-1}$ and $\mathbf{G}_v^{-1}$ have been calculated for the Example data. The remaining matrices in the MME are calculated through matrix

multiplication and addition. The MME are too large to be shown, but solving the equations by direct inversion gives the following results:

Effects	Solutions
Sex[a]	
1	7.357
2	5.529
Animal	Breeding values
1	0.092
2	–0.091
3	0.341
4	0.329
5	0.515
MQTL alleles of animals	Additive effects
1*p*	0.064
1*m*	0.011
2*p*	–0.065
2*m*	–0.011
3*p*	0.083
3*m*	–0.004
4*p*	0.028
4*m*	0.076
5*p*	0.043
5*m*	0.086

[a]1 = male, 2 = female.

The additive genetic effects of the MQTL accounted for about 45% of the total genetic merit of animals 1 and 2 but only about 20% for animals 3 and 5.

In Germany, with Holstein dairy cattle, the method in Example 8.2 has been used for incorporating QTL information into routine estimation of breeding values (Szyda *et al.*, 2003). In their study, 13 markers were used for routine genotyping of animals, and regions representing QTL for milk, protein, fat yields and somatic cell counts have been identified on several chromosomes. The QTL information has been incorporated into BLUP, analysing daughter yield deviation (DYD) as the dependent variable. As a percentage of the polygenic variance, the variances of the MQTL in their study varied from 3 to 5% for milk, fat and protein yields in the first lactation.

8.6 Reduced Animal Model with Marker Information

The main advantage of a reduced animal model (RAM), as indicated in Section 3.4, is the reduction in the number of equations in the MME to be

solved. Fitting an animal model with marker information incorporated, using the method of Fernando and Grossman (1989), results in the total number of equations for animals being $k(2m + 1)$, where k is the number of animals and m the number of MQTL. For a large k, m or both, solving such a system may not always be feasible. Cantet and Smith (1991) presented a RAM, based on the methodology of Fernando and Grossman(1989), which would reduce the total number of equations for animals from $k(2m + 1)$ to $q(2m + 1)$, where q is the number of parents. The application of RAM involves partitioning the data vector, $\mathbf{y}$, into records of animals that are parents ($\mathbf{y}_p$) and those without progeny ($\mathbf{y}_n$; non-parents). With a conformable partitioning of $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$, $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{e}$, then [8.17] can be written as:

$$\begin{pmatrix}\mathbf{y}_p\\ \mathbf{y}_n\end{pmatrix}=\begin{pmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{pmatrix}\boldsymbol{\beta}+\begin{pmatrix}\mathbf{Z}_p & 0\\ 0 & \mathbf{Z}_n\end{pmatrix}\begin{pmatrix}\mathbf{u}_p\\ \mathbf{u}_n\end{pmatrix}+\begin{pmatrix}\mathbf{W}_p & 0\\ 0 & \mathbf{W}_n\end{pmatrix}\begin{pmatrix}\mathbf{v}_p\\ \mathbf{v}_n\end{pmatrix}+\begin{pmatrix}\mathbf{e}_p\\ \mathbf{e}_n\end{pmatrix} \qquad [8.19]$$

The vectors $\mathbf{u}_n$ and $\mathbf{v}_n$ are expressed as linear functions of $\mathbf{u}_p$ and $\mathbf{v}_p$ to obtain the RAM. The equations relating progeny MQTL effects to MQTL of parents were given by Fernando and Grossman (1989) as:

$$v_o^p = (1-\rho_o^p)v_s^p + \rho_o^p v_s^m + \varepsilon_o^p$$
$$v_o^m = (1-\rho_o^m)v_d^p + \rho_o^p v_d^m + \varepsilon_o^m$$

Therefore, for non-parents ($\mathbf{v}_n$), the vector of MQTL effects can be written as a function of that of parents ($\mathbf{v}_p$) as:

$$\mathbf{v}_n = \mathbf{P}\mathbf{v}_p + \boldsymbol{\varepsilon} \qquad [8.20]$$

where the vector $\boldsymbol{\varepsilon}$ has elements i and $i + 1$ equal to ε_o^p and ε_o^m, respectively. As shown in Section 8.4, var($\boldsymbol{\varepsilon}$) is diagonal and is equal to $\mathbf{H}\sigma_v^2$. The matrix $\mathbf{P}$ of order $2(k - q)$ by $2q$ is as defined in Section 8.4 and the ith row contains at most two non-zero elements in the columns corresponding to the effects of the paternal and maternal MQTL of the parent if known. The value of these non-zero elements has been defined in Section 8.4. For instance, assume that only the first three calves in the pedigree (for Example 8.1) were being considered in an analysis. Then calf 3 is a non-parent and the matrix $\mathbf{W}$ is:

$$\mathbf{W}=\begin{pmatrix}\mathbf{W}_p & 0\\ 0 & \mathbf{W}_n\end{pmatrix}=\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\end{pmatrix}$$

And, for $r = 0.1$, $\mathbf{P}$ is:

$$\mathbf{P}=\begin{pmatrix}0.9&0.1&0&0\\0&0&0.1&0.9\end{pmatrix}$$

The derivation of the equations for RAM for estimating $\mathbf{u}_p$ and back-solving for $\mathbf{u}_n$ has been presented in Chapter 3, Section 3.4. From [3.16], an expression for $\mathbf{u}_p$ in matrix notation can be written as:

$$\mathbf{u}_p = \mathbf{Z}_1\mathbf{u}_p + \boldsymbol{\phi}$$

where $\boldsymbol{\phi}$ is the vector of Mendelian sampling.

Substituting the above equation and [8.20] into [8.19] gives the following RAM equations:

$$\begin{pmatrix}\mathbf{y}_p\\ \mathbf{y}_n\end{pmatrix}=\begin{pmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{pmatrix}\boldsymbol{\beta}+\begin{pmatrix}\mathbf{Z}_p & 0\\ \mathbf{Z}_n\mathbf{Z}_1 & \mathbf{Z}_n\end{pmatrix}\begin{pmatrix}\mathbf{u}_p\\ \boldsymbol{\phi}\end{pmatrix}+\begin{pmatrix}\mathbf{W}_p & 0\\ \mathbf{W}_n\mathbf{P} & \mathbf{W}_n\end{pmatrix}\begin{pmatrix}\mathbf{v}_p\\ \boldsymbol{\varepsilon}\end{pmatrix}+\begin{pmatrix}\mathbf{e}_p\\ \mathbf{e}_n\end{pmatrix} \quad [8.21]$$

or:

$$\begin{pmatrix}\mathbf{y}_p\\ \mathbf{y}_n\end{pmatrix}=\begin{pmatrix}\mathbf{X}_p\\ \mathbf{X}_n\end{pmatrix}\boldsymbol{\beta}+\begin{pmatrix}\mathbf{Z}_p\\ \mathbf{Z}_1\end{pmatrix}(\mathbf{u}_p)+\begin{pmatrix}\mathbf{W}_p\\ \mathbf{T}\end{pmatrix}(\mathbf{v}_p)+\begin{pmatrix}\mathbf{e}_p\\ \mathbf{e}^*+\mathbf{W}_n\boldsymbol{\varepsilon}\end{pmatrix} \quad [8.22]$$

where $\mathbf{e}^*$, $\mathbf{Z}_1$ are as defined in [3.21] and [3.22], respectively, and $\mathbf{T}=\mathbf{W}_n\mathbf{P}$. Note that, in the above equation, $\mathbf{Z}_n\mathbf{Z}_1$ in [8.21] = $\mathbf{Z}_1$ in [8.22] since $\mathbf{Z}_n$ is an identity matrix.

The variance for the vector of residuals can be expressed as:

$$\text{var}\begin{pmatrix}\mathbf{e}_p\\ \mathbf{e}^*+\mathbf{W}_n\boldsymbol{\varepsilon}\end{pmatrix}=\begin{pmatrix}\mathbf{R}_p & 0\\ 0 & \mathbf{R}_n\end{pmatrix}=\begin{pmatrix}\mathbf{I} & 0\\ 0 & \mathbf{I}+\mathbf{D}\alpha_1^{-1}+\mathbf{F}\alpha_2^{-1}\end{pmatrix}\sigma_e^2$$

with $\alpha_1^{-1}=\sigma_a^2/\sigma_e^2$ and $\alpha_2^{-1}=\sigma_v^2/\sigma_e^2$, $\mathbf{D}$ as defined in [3.22] and $\mathbf{F}=\mathbf{W}_n\mathbf{H}\mathbf{W}_n'$. The matrix $\mathbf{F}$ is a diagonal matrix with the ith diagonal element equal to $\text{var}(\varepsilon_o^p)+\text{var}(\varepsilon_o^m)$, which is equal to the sum of the diagonal elements of row i and $i+1$ of $\mathbf{H}$. Thus the diagonal elements of $\mathbf{F}$ are equal to:

$2r(1-r)(2-f_s-f_d)$	if both sire and dam of the non-parent are known
$2r(1-r)(1-f_s)+1$	if only the sire is known
$2r(1-r)(1-f_d)+1$	if only the dam is known
2	if both parents are unknown

The MME for [8.22] with $\mathbf{R}_p$ factored out are:

$$\begin{bmatrix}\mathbf{X}_p'\mathbf{X}_p+\mathbf{X}_n'\mathbf{R}_s^{-1}\mathbf{X}_n & \mathbf{X}_p'\mathbf{Z}_p+\mathbf{X}_n'\mathbf{R}_s^{-1}\mathbf{Z}_1 & \mathbf{X}_p'\mathbf{W}_p+\mathbf{X}_n'\mathbf{R}_s^{-1}\mathbf{T}\\ \mathbf{Z}_p'\mathbf{X}_p+\mathbf{Z}_n'\mathbf{R}_s^{-1}\mathbf{X}_n & \mathbf{Z}_p'\mathbf{Z}_p+\mathbf{Z}_1\mathbf{R}_s^{-1}\mathbf{Z}_1+\mathbf{A}_{up}^{-1}\alpha_1 & \mathbf{Z}_p'\mathbf{W}_p+\mathbf{Z}_1\mathbf{R}_s^{-1}\mathbf{T}\\ \mathbf{W}_p'\mathbf{X}+\mathbf{T}'\mathbf{R}_s^{-1}\mathbf{X} & \mathbf{W}_p'\mathbf{Z}_p+\mathbf{T}'\mathbf{R}_s^{-1}\mathbf{Z}_1 & \mathbf{W}_p'\mathbf{W}_p+\mathbf{T}'\mathbf{R}_s^{-1}\mathbf{T}+\mathbf{G}_{vp}^{-1}\alpha_2\end{bmatrix}$$

$$\begin{bmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\mathbf{u}}_p\\ \hat{\mathbf{v}}_p\end{bmatrix}=\begin{bmatrix}\mathbf{X}_p'\mathbf{y}_p+\mathbf{X}_n'\mathbf{R}_s^{-1}\mathbf{y}_n\\ \mathbf{Z}_p'\mathbf{y}_p+\mathbf{Z}_1'\mathbf{R}_s^{-1}\mathbf{y}_n\\ \mathbf{W}_p'\mathbf{y}_p+\mathbf{T}'\mathbf{R}_s^{-1}\mathbf{y}_n\end{bmatrix} \quad [8.23]$$

where $\mathbf{R}_s^{-1}=(1+\mathbf{D}\alpha_1^{-1}+\mathbf{F}\alpha_2^{-1})^{-1}$ and the matrices $\mathbf{A}_{up}^{-1}$ and $\mathbf{G}_{vp}^{-1}$ are formed only for individuals who are parents.

8.6.1 Numerical example

Example 8.3

Using the data in Example 8.1 and the genetic parameters, a reduced animal model is used to estimate the sex effects of calves and predict breeding values of animals and the paternal and maternal additive effects of MQTL alleles of animals.

The only non-parent animal in the data set is calf 5 with both parents known. Note that the sire of calf 5 is inbred. Therefore the diagonal element of $\mathbf{D}$ for calf 5 accounting for inbreeding equals 0.5(1 – 0.125) = 0.4375. Also

the diagonal element of $\mathbf{F}$ for calf 5 = $2r(1-r)(2-\mathbf{F}_s-\mathbf{F}_d) = 0.2(0.1)\times(0.9)(2-0.162-0) = 0.3308$. Therefore $\mathbf{R}_s^{-1} = (1 + 0.4375(0.5) + 0.3308(0.0833))^{-1} = 0.8024$.

The design matrix $\mathbf{W}_p$ is as in Example 8.1 but formed only for parents, $\mathbf{Z}_p = \mathbf{I}_4$ and $\mathbf{Z}_1 = [0\ \ 0\ \ 0.5\ \ 0.5]$. The matrix $\mathbf{P}$ relates the MQTL alleles of calf 5 to those of its parents and is:

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 \\ 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 & 0 \end{pmatrix}$$

Since $\mathbf{W}_n$ for calf 5 equals [1 1], then $\mathbf{W}_n\mathbf{P} = \mathbf{T} = [0\ \ 0\ \ 0\ \ 0\ \ 0.9\ \ 0.1\ \ 0.9\ \ 0.1]$. The matrices in the MME can be calculated through matrix multiplication from the design matrices outlined. For instance, the matrix $\mathbf{W}_p'\mathbf{W}_p + \mathbf{T}'\mathbf{R}_s^{-1}\mathbf{T}$ is:

$$\mathbf{W}_p'\mathbf{W}_p + \mathbf{T}'\mathbf{R}_s^{-1}\mathbf{T} = \begin{bmatrix} 1.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 1.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.650 & 1.072 & 0.650 & 0.072 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.072 & 1.008 & 0.072 & 0.008 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.650 & 0.072 & 1.650 & 1.072 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.072 & 0.008 & 1.072 & 1.008 \end{bmatrix}$$

The matrices $\mathbf{A}_{up}^{-1}$ and $\mathbf{G}_{vp}^{-1}$ in the MME have been calculated, accounting for inbreeding as described in Section 8.4. The number of equations in the MME for the reduced animal model are 14 compared to 17 for the animal model in Example 8.1. The MME are too large to be shown but solving the equations gives the following solutions:

Effects	Solutions
Sex[a]	
1	7.357
2	5.529
Animal	Breeding values
1	0.092
2	–0.091
3	0.341
4	0.329
MQTL alleles of animals	Additive effects
1*p*	0.064
1*m*	0.011
2*p*	–0.065
2*m*	–0.011
3*p*	0.083
3*m*	–0.004
4*p*	0.028
4*m*	0.076

[a]1 = male, 2 = female.

As expected, solutions for sex of calf effect and parents were the same as those obtained with the animal model in Example 8.1.

8.6.2 Back-solving for solutions of non-parents

After obtaining solutions for fixed effects, parental breeding values and parental MQTL effects, the breeding values and additive MQTL effects for non-parents can be calculated. From [8.21] the equations for $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\varepsilon}}$ are:

$$\mathbf{Z}'_n\mathbf{X}'_n\hat{\boldsymbol{\beta}} + \mathbf{Z}'_n\mathbf{Z}_1\hat{\mathbf{u}}_p + (\mathbf{Z}'_n\mathbf{Z}'_n + \mathbf{D}^{-1}\alpha_1)\hat{\boldsymbol{\phi}} + \mathbf{Z}'_n\mathbf{W}_n\mathbf{P}\hat{\mathbf{v}}_p + \mathbf{Z}'_n\mathbf{W}_n\hat{\boldsymbol{\varepsilon}} = \mathbf{Z}'_n\mathbf{y}_n$$

$$\mathbf{W}'_n\mathbf{X}'_n\hat{\boldsymbol{\beta}} + \mathbf{W}'_n\mathbf{Z}_1\hat{\mathbf{u}}_p + \mathbf{W}'_n\mathbf{Z}_n\hat{\boldsymbol{\phi}} + \mathbf{W}'_n\mathbf{W}_n\mathbf{P}\hat{\mathbf{v}}_p + (\mathbf{W}'_n\mathbf{W}_n + \mathbf{H}^{-1}\alpha_2)\hat{\boldsymbol{\varepsilon}} = \mathbf{W}'_n\mathbf{y}_n$$

Assuming that non-parents have one record each ($\mathbf{Z}_n = \mathbf{I}$) and moving terms which are not directly associated with $\boldsymbol{\phi}$ and $\boldsymbol{\varepsilon}$ to the right-hand side gives:

$$\begin{pmatrix} \mathbf{I} + \mathbf{D}^{-1}\alpha_1 & \mathbf{W}_n \\ \mathbf{W}'_n & \mathbf{W}'_n\mathbf{W}_n + \mathbf{H}^{-1}\alpha_2 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\phi}} \\ \hat{\boldsymbol{\varepsilon}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{W}'_n \end{pmatrix} (\mathbf{y}_n - \mathbf{X}_n\hat{\boldsymbol{\beta}} - \mathbf{Z}_1\hat{\mathbf{u}}_p - \mathbf{T}\hat{\mathbf{v}}_p) \quad [8.24]$$

On absorbing the equations for $\hat{\boldsymbol{\phi}}$, the equation for $\hat{\boldsymbol{\varepsilon}}$ becomes:

$$\begin{aligned} &(\mathbf{W}'_n[\mathbf{I} - (\mathbf{I} + \mathbf{D}^{-1}\alpha_1)^{-1}]\mathbf{W}_n + \mathbf{H}^{-1}\alpha_2)\hat{\boldsymbol{\varepsilon}} \\ &\quad = \mathbf{W}'_n(\mathbf{I} - (\mathbf{I} + \mathbf{D}^{-1}\alpha_1)(y_n - \mathbf{X}_n\hat{\boldsymbol{\beta}} - \mathbf{Z}_1\hat{\mathbf{u}}_p - \mathbf{T}\hat{\mathbf{v}}_p) \end{aligned} \quad [8.25]$$

Let $\mathbf{B} = \mathbf{I} - (\mathbf{I} + \mathbf{D}^{-1}\alpha_1)^{-1}$ in [8.5], then $\mathbf{B}$ is a diagonal matrix with element b_{ii} equal to:

$$b_{ii} = 1 - (1 + (\alpha_1/d_{ii}))^{-1} = \alpha_1/(\alpha_1 + d_{ii})$$

where d_{ii} is the diagonal element of $\mathbf{D}$. The matrix $\mathbf{W}'_n\mathbf{B}\mathbf{W}_n$ in [8.25] is then block diagonal, since $\mathbf{W}_n$ has rows with two consecutive elements equal to 1 and the remaining elements are equal to zero. Each block is of order 2 by 2 with all elements equal to b_{ii}. Adding $\mathbf{H}^{-1}\alpha_2$ to this 2 by 2 block diagonal matrix gives the coefficient matrix on the left-hand side of [8.25]. Thus solutions for $\hat{\boldsymbol{\varepsilon}}$ can be obtained by solving $(s - p)$ systems of order 2 where s is total number of animals. For the ith animal, [8.25] can be written as:

$$\begin{pmatrix} b_{ii} + \dfrac{\alpha_2}{2r(1-r)(1-f_s)} & b_{ii} \\ b_{ii} & b_{ii} + \dfrac{\alpha_2}{2r(1-r)(1-f_d)} \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_i^p \\ \hat{\varepsilon}_i^m \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} b_{ii}y_{ci} \quad [8.26]$$

where y_{ci} = is the element i of $\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1\hat{\mathbf{u}}_p - \mathbf{T}\mathbf{v}_p$.

After solving for $\hat{\boldsymbol{\varepsilon}}$, the solution $\hat{\boldsymbol{\phi}}$ can be obtained by solving the first equation in [8.24] as:

$$\begin{aligned} \hat{\boldsymbol{\phi}} &= (\mathbf{I} + \mathbf{D}^{-1}\alpha_1)^{-1}\mathbf{y}_c - \mathbf{W}_n\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\phi}}_i = (\mathbf{I} + \mathbf{D}^{-1}\alpha_1)^{-1}y_{ci} - \hat{\varepsilon}_i^p - \hat{\varepsilon}_i^m \\ &= (d_{ii}/(d_{ii} + \alpha_1))y_{ci} - \hat{\varepsilon}_i^p - \hat{\varepsilon}_i^m \end{aligned} \quad [8.27]$$

Solutions for additive genetic effects not linked to MQTL for non-parents are then obtained as:

$$\hat{\mathbf{u}}_n = \mathbf{Z}_1\hat{\mathbf{u}}_p + \boldsymbol{\phi}_n \quad [8.28]$$

and additive MQTL effects as:

$$\hat{\mathbf{v}}_n = \mathbf{P}\hat{\mathbf{v}}_p + \hat{\boldsymbol{\varepsilon}} \quad [8.29]$$

In Example 8.2, parents of the only non-parent (calf 5) are known; therefore:

$$b_{55} = 2/(2 + 0.4375) = 0.8205$$

For calf 5, $b_{55}y_{c5}$ in [8.26] is:

$$\begin{aligned}&0.8205(y_5 - \hat{\beta}_2 - 0.5(u_4 + u_3) - (1-r)\hat{v}_{4p} - r(\hat{v}_{4m}) - (1-r)\hat{v}_{3p} - r(\hat{v}_{3m}))\\ &= 0.8205(7.0 - 5.529 - 0.5(0.341 + 0.329) - 0.9(0.028) - 0.1(0.076)\\ &\quad -0.9(0.083) - 0.1(-0.004))\\ &= 0.8205(1.10289) = 0.8442\end{aligned}$$

Then equation [8.26] becomes:

$$\begin{pmatrix} 0.8205 + \dfrac{12}{0.15084} & 0.8205 \\ 0.8205 & 0.8205 + \dfrac{12}{0.16} \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_5^p \\ \hat{\varepsilon}_5^m \end{pmatrix} = \begin{pmatrix} 0.8442 \\ 0.8442 \end{pmatrix}$$

The solutions for ε_5^p and ε_5^m are 0.0104 and 0.0124, respectively, from the above equation. Therefore the solutions for the paternal and maternal MQTL effects of calf 5 from [8.29] are:

$$\hat{v}_{5p} = (1-r)\hat{v}_{4p} + r(\hat{v}_{4m}) + \hat{\varepsilon}_5^p = 0.9(0.028) + 0.1(0.076) + 0.0104 = 0.043$$
$$\hat{v}_{5m} = (1-r)\hat{v}_{3p} + r(\hat{v}_{3m}) + \hat{\varepsilon}_5^m = 0.9(0.083) + 0.1(-0.004) + 0.0124 = 0.086$$

Using [8.27], the solution for $\hat{\phi}_5$ is:

$$\begin{aligned}\hat{\phi}_5 &= [0.4375/(0.4375 + 2)](y_c - \hat{\varepsilon}_5^p - \hat{\varepsilon}_5^p)\\ &= 0.179(1.029 - 0.0104 - 0.0124) = 0.180\end{aligned}$$

Therefore, the breeding value of calf 5 from [8.28] is:

$$\hat{u}_5 = 0.5(\hat{u}_4 + \hat{u}_5) + \hat{\phi}_5 = 0.5(0.341 + 0.329) + 0.180 = 0.515$$

which is the same as that obtained from the animal model.

8.7 Directly Predicting the Additive Genetic Merit at the MQTL

In Section 8.6, the application of a RAM in order to reduce the number of equations for the prediction of breeding values was presented. However, with RAM, three equations are still required for each animal that is a parent. Another approach to reduce the number of equations in the MME is to

directly predict the combined additive genetic effects of the paternal and maternal alleles at the MQTL of an individual. The number of equations per animal would therefore be two, one for the additive genetic effects not linked to the MQTL and the other for MQTL. This implies predicting the additive genetic effects at the MQTL at the animal level; therefore a covariance matrix ($\mathbf{A}_v$) for the MQTL at the animal level is needed. The covariance matrix $\mathbf{A}_v$ can be obtained from $\mathbf{G}_v$ as $\mathbf{A}_v = \frac{1}{2}\mathbf{B}\mathbf{G}_v\mathbf{B}'$; where $\mathbf{B} = \mathbf{I}_n \otimes [1\ \ 1]$, where n is the number of animals and $\otimes$ denotes the Kronecker product. For Example 8.1, the matrix $\mathbf{B} = \mathbf{W}$ in Section 8.4 and $\mathbf{A}_v$ is:

$$\mathbf{A}_v = \begin{bmatrix} 1.000 & 0.000 & 0.500 & 0.950 & 0.945 \\ 0.000 & 1.000 & 0.500 & 0.050 & 0.055 \\ 0.500 & 0.500 & 1.000 & 0.590 & 0.631 \\ 0.950 & 0.050 & 0.590 & 1.162 & 1.072 \\ 0.945 & 0.055 & 0.631 & 1.072 & 1.228 \end{bmatrix}$$

Equation [8.11] can be used to obtain the inverse of $\mathbf{A}_v$. However, the vector $\mathbf{s}_i$, containing the contributions from ancestors, is needed and this can be computed using [8.10]. The vector $\mathbf{s}_i$ for the ith animal needed to calculate $\mathbf{A}_v^{-1}$ is shown in Table 8.1.

The inverse of $\mathbf{A}_v$ is:

$$\mathbf{A}_v^{-1} = \begin{bmatrix} 4.966 & 0.286 & -0.148 & -2.723 & -1.382 \\ 0.286 & 1.519 & -1.068 & 0.013 & 0.249 \\ -0.148 & -1.068 & 2.245 & -0.298 & -0.732 \\ -2.723 & 0.013 & -0.298 & 5.978 & -2.971 \\ -1.382 & 0.249 & -0.732 & -2.971 & 4.836 \end{bmatrix}$$

The model for the prediction now becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{q} + \mathbf{e} \qquad [8.30]$$

where all terms are as defined in [8.18] except that $\mathbf{W}$ is now identical to $\mathbf{Z}$ and relates additive genetic effects at the MQTL to animals. Both matrices $\mathbf{Z}$ and $\mathbf{W}$ are identity matrices and are of the order of animals. The vector $\mathbf{q}$ is the vector of additive genetic effects at the MQTL and is equal to the

Table 8.1. Vector ($\mathbf{s}_i$) with contributions at the MQTL from ancestors (animals 1 to 4) to animals 2 to 5, using the pedigree in Example 8.1.

	Elements in $\mathbf{s}_i$ relating to animal			
Animal	1	2	3	4
2	0.0000			
3	0.5000	0.5000		
4	0.8600	–0.0400	0.1800	
5	0.2857	–0.0514	0.1514	0.6143

sum of the additive genetic effects of the paternal and maternal alleles for the animal. The covariance matrix of $\mathbf{q} = 2\mathbf{A}_v\sigma_v^2 = \mathbf{A}_v\sigma_q^2$, since $\sigma_q^2 = 2\sigma_v^2$. The MME for the above model are:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}_u^{-1}\alpha_1 & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \mathbf{A}_v^{-1}\alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{q}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix} \qquad [8.31]$$

with:

$\alpha_1 = \sigma_e^2/\sigma_u^2$ and $\alpha_2 = \sigma_e^2/\sigma_q^2$

8.7.1 An illustration

Example 8.4

Using the same data set as in Example 8.2 and the same genetic parameters, the prediction of additive genetic effects breeding values at the QTL not linked to the MQTL, and combined additive genetic effect of the MQTL at the animal level, is illustrated.

From the parameters, $\alpha_1 = 0.6/0.3 = 2$ and $\alpha_2 = 0.6/0.10 = 6$. The design matrices **X** and **Z** are as defined in Example 8.2 and **W** is now equal to **Z**. The MME will be too large to show but the matrix $\mathbf{W'R}^{-1}\mathbf{W} + \mathbf{A}_v^{-1}\alpha_2$ is:

$$\mathbf{W'R}^{-1}\mathbf{W} + \mathbf{A}_v^{-1}\alpha_2 = \begin{bmatrix} 30.796 & 1.716 & -0.888 & -16.338 & -8.292 \\ 1.716 & 10.114 & -6.408 & 0.078 & 1.494 \\ -0.888 & -6.408 & 14.470 & -1.788 & -4.392 \\ -16.338 & 0.078 & -1.788 & 36.868 & -17.826 \\ -8.292 & 1.494 & -4.392 & -17.826 & 30.016 \end{bmatrix}$$

Solving the MME gave the following solutions:

Effects	Solutions
Sex[a]	
1	7.356
2	5.529
Animal	Additive genetic effects not linked to MQTL
1	0.091
2	–0.091
3	0.341
4	0.329
5	0.515
Animal	Combined additive genetic effects at the MQTL
1	0.076
2	–0.076
3	0.079
4	0.104
5	0.130

[a]1 = male, 2 = female.

The solutions for the additive effect at the MQTL are the same as the sum of estimated effects in Examples 8.2 and 8.3. Although the number of equations in the MME is 12 compared with 14 for the RAM, the difference between the two models will generally be data dependent. However, application of this model may be limited to populations of small size as the tabular method of calculating $\mathbf{A}_v$ and its inverse may not be computationally feasible in large populations.

8.8 Predicting Total Additive Genetic Merit

Van Arendonk *et al.* (1994) showed that total additive genetic merit ($\mathbf{a}$) for animals that includes marker information could be predicted directly. This implies that only a single equation is needed for an animal in the MME for the prediction of breeding values with marker information included. Let equation [8.17] be written as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e} \qquad [8.32]$$

where $\mathbf{a} = \mathbf{u} + \mathbf{Kv}$, with $\mathbf{u}$ and $\mathbf{v}$ as defined in [8.17]. The matrix $\mathbf{K}$, which relates allelic effects to animals, is identical to $\mathbf{W}$ in equation [8.17] when all animals have observations. The covariance matrix of $\mathbf{a}$ ($\mathbf{V}_a$) is:

$$\begin{aligned}
\mathbf{V}_a &= \text{var}(\mathbf{u} + \mathbf{Kv}) \\
&= \text{var}(\mathbf{u}) + \mathbf{K}\,\text{var}(\mathbf{v})\mathbf{K}' \\
&= \mathbf{A}_u\sigma_u^2 + \mathbf{KG}_v\mathbf{K}'\sigma_v^2 \\
&= \mathbf{A}_u\sigma_u^2 + 2\mathbf{A}_v\sigma_v^2 \\
\mathbf{V}_a &= \mathbf{A}_u\sigma_u^2 + \mathbf{A}_v\sigma_q^2
\end{aligned}$$

The combined numerator relationship matrix among animals with marker information included ($\mathbf{A}_a$) is:

$$\mathbf{A}_a = \mathbf{A}_u\sigma_u^2/\sigma_a^2 + \mathbf{A}_v\sigma_q^2/\sigma_a^2 \qquad [8.33]$$

with:

$$\sigma_a^2 = \sigma_u^2 + \sigma_q^2$$

The MME for [8.32] are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}_a^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \qquad [8.34]$$

where:

$$\alpha = \sigma_e^2/\sigma_a^2$$

The use of equation [8.34] would require the inverse of $\mathbf{A}_a$ to be calculated. Initially $\mathbf{A}_a$ is computed using [8.33], $\mathbf{A}_a^{-1}$ can be calculated using [8.11], with the vector $\mathbf{s}_i$ containing the contributions from ancestors computed

using [8.10]. The calculation of both matrices is illustrated in the following example.

8.8.1 Numerical application

Example 8.5
Using the same data set as in Example 8.2 and the same genetic parameters, the total additive genetic effects of animals, which included marker information, are directly predicted. From the genetic parameters in Example 8.2, $\sigma_a^2 = \sigma_u^2 + \sigma_q^2 = 0.3 + 0.1 = 0.4$ and $\sigma_e^2 = 0.6$; therefore, $\alpha = \sigma_e^2/\sigma_a^2 = 0.6/0.4 = 1.5$. The **Z** matrix in [8.32] is now an identity matrix considering animals with records.

The matrix $\mathbf{A}_a$ below was calculated as the sum of $\mathbf{A}_{u(0.3/0.4)}$ and $\mathbf{A}_{v(0.1/0.4)}$. The matrices $\mathbf{A}_u$ and $\mathbf{A}_v$ have been calculated in Examples 8.2 and 8.4.

$$\mathbf{A}_a = \begin{bmatrix} 1.000 & 0.000 & 0.500 & 0.800 & 0.705 \\ 0.000 & 1.000 & 0.500 & 0.200 & 0.295 \\ 0.500 & 0.500 & 1.000 & 0.710 & 0.814 \\ 0.800 & 0.200 & 0.710 & 1.228 & 1.018 \\ 0.705 & 0.295 & 0.814 & 1.018 & 1.338 \end{bmatrix}$$

The vector $\mathbf{s}_i$ for the ith animal needed to calculate $\mathbf{A}_a^{-1}$ is shown in Table 8.2.

The matrix $\mathbf{A}_a^{-1}$ calculated using [8.11] is:

$$\mathbf{A}_a^{-1} = \begin{bmatrix} 2.2641 & 0.4854 & -0.4101 & -1.2080 & -0.1314 \\ 0.4854 & 1.5007 & -1.0218 & 0.0030 & 0.0327 \\ -0.4101 & -1.0218 & 2.7536 & -0.3673 & -0.9544 \\ -1.2080 & 0.0030 & -0.3673 & 2.9811 & -1.4088 \\ -0.1314 & 0.0327 & -0.9544 & -1.4088 & 2.4619 \end{bmatrix}$$

The MME [8.32] for the Example data are as follows:

$$\begin{bmatrix} 2.000 & 0.000 & 1.000 & 0.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 3.000 & 0.000 & 1.000 & 0.000 & 1.000 & 1.000 \\ 1.000 & 0.000 & 4.396 & 0.728 & -0.615 & -1.812 & -0.197 \\ 0.000 & 1.000 & 0.728 & 3.251 & -1.533 & 0.005 & 0.049 \\ 1.000 & 0.000 & -0.615 & -1.533 & 5.130 & -0.551 & -1.432 \\ 0.000 & 1.000 & -1.812 & 0.005 & -0.551 & 5.472 & -2.113 \\ 0.000 & 1.000 & -0.197 & 0.049 & -1.432 & -2.113 & 4.693 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \end{bmatrix} = \begin{bmatrix} 15.3 \\ 17.5 \\ 6.8 \\ 4.5 \\ 8.5 \\ 6.0 \\ 7.0 \end{bmatrix}$$

Table 8.2. Vector ($\mathbf{s}_i$) with contributions from ancestors (animals 1 to 4) to animals 2 to 5, using the pedigree in Example 8.1.

	Elements in $\mathbf{s}_i$ relating to animal			
Animal	1	2	3	4
2	0.0000			
3	0.5000	0.5000		
4	0.5900	–0.0100	0.4200	
5	0.0534	–0.0133	0.3877	0.5722

Solving the MME gave these results:

Effects	Solutions
Sex[a]	
1	7.356
2	5.529
Animal	Total additive genetic merit including marker information
3	0.167
4	–0.167
5	0.419
6	0.432
7	0.645

[a]1 = male, 2 = female.

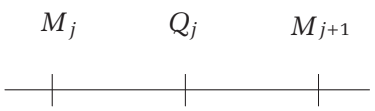
The application of [8.32] is valuable as only one equation is fitted per animal but its application to a large data set may be limited because of the tabular method of calculating the relationship matrix needed and its inverse.

8.9 Analysis of Data with QTL Bracketed by Two Markers

This section deals with the extension of the model of Fernando and Grossman (1989) by Goddard (1992) to handle situations in which MQTL is bracketed between two markers. The use of marker information when MQTL is bracketed between two markers should enhance the accuracy of estimated breeding values compared with information with a single marker.

8.9.1 Basic model

Consider a chromosome with a series of marked loci with at most one QTL located between each pair of markers:



Each animal inherits two alleles at the Q_j locus, one from its sire and the other from its dam. A marker haplotype consisting of the marker alleles at M_j and M_{j+1} would be associated with each of the MQTL alleles. Let the *j*th chromosome segment that animal *i* inherited from its sire be of the marker haplotype (*kl*) and the value of the MQTL allele be $v_{ij(kl)}$ or simply $v_{ij(p)}$. Similarly, let the value of the MQTL allele from its dam be $v_{ij(m)}$. Summed over all chromosome segments, the breeding value of animal $i(a_i)$ is:

$$a_i = u_i + \sum_j v_{ij(p)} + \sum_j v_{ij(m)}$$

Similarly to [8.3], the model for the phenotypic record of the animal is:

$$y_i = x_i\beta + u_i + \sum_j v_{ij(p)} + \sum_j v_{ij(m)} + e$$

or in matrix notation the model is:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \sum_j \mathbf{W}_j\mathbf{v}_j + \mathbf{e}$$

The terms are as defined in [8.17]. The vector $\mathbf{v}_j$ contains the effects of the paternal and maternal MQTL alleles at each locus. The summation is over chromosome segments bounded by markers. The variance of $\mathbf{u}$ and $\mathbf{v}_j$ is as defined in [8.18], such that:

$$\text{var}(\mathbf{v}_j) = \mathbf{G}_{vj}J^2_{vj}$$

Assuming $j = 2$, the BLUP equations for the above model are:

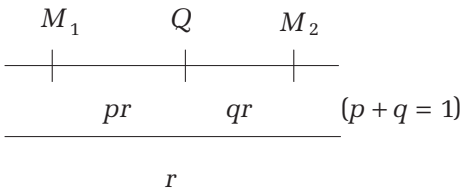
$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W}_1 & \mathbf{X'W}_2 \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z'W}_1 & \mathbf{Z'W}_2 \\ \mathbf{W}_1'\mathbf{X} & \mathbf{W}_1'\mathbf{Z} & \mathbf{W}_1'\mathbf{W}_1' + \mathbf{G}_{v1}^{-1}\alpha_2 & \mathbf{W}_1'\mathbf{W}_2 \\ \mathbf{W}_2'\mathbf{X} & \mathbf{W}_2'\mathbf{Z} & \mathbf{W}_2'\mathbf{W}_1 & \mathbf{W}_2'\mathbf{W}_2 + \mathbf{G}_{v2}^{-1}\alpha_3 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{v}}_1 \\ \hat{\mathbf{v}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W}_1'\mathbf{y} \\ \mathbf{W}_2'\mathbf{y} \end{pmatrix} \quad [8.35]$$

where:

$$\alpha_1 = \sigma^2_e/\sigma^2_u, \quad \alpha_2 = \sigma^2_e/\sigma^2_{v1} \quad \text{and} \quad \alpha_3 = \sigma^2_e/\sigma^2_{v2}$$

8.9.2 Calculating the covariance matrix, G

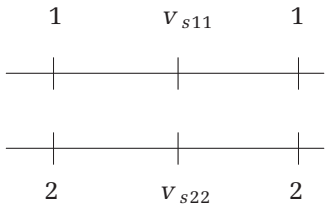
Consider a single MQTL bounded by two marker loci with marker distances as follows:



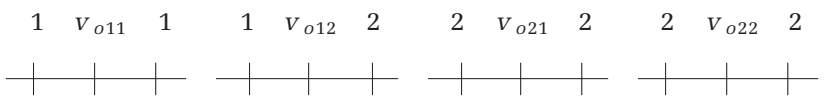
With the assumption of no crossover, the recombination rates are (Haldane, 1919) between:

M_1 and $M_2 = a = 0.5(1 - e^{-2r})$
M_1 and $Q = b = 0.5(1 - e^{-2pr})$
Q and $M_2 = c = 0.5(1 - e^{-2qr})$

Similarly to the situation with a single marker, the variance of **v** depends on the relationship among the **v** terms. The MQTL alleles in the progeny can be expressed in terms of parental MQTL. Thus, given, for instance, that the genotype of the sire is:



The sire will produce the following four types of gametes on the basis of marker haplotypes:



Assuming no double recombination between markers, the frequency, means and approximate means for the four gametes (Goddard, 1992) are:

Haplotype		Frequency	Mean	Approximate mean
1	1	$\frac{1}{2}(1 - a)$	$[(1 - b)(1 - c)/1 - a]v_{s11} + [bc/1 - a]v_{s22}$	v_{s11}
1	2	$\frac{1}{2}a$	$[(1 - b)c/a]v_{s11} + [b(1 - c)/a]v_{s22}$	$qv_{s11} + pv_{s22}$
2	1	$\frac{1}{2}a$	$[b(1 - c)/a]v_{s11} + [(1 - b)c/a]v_{s22}$	$pv_{s11} + qv_{s22}$
2	2	$\frac{1}{2}(1 - a)$	$[bc/1 - a]v_{s11} + [(1 - b)(1 - c)/1 - a]v_{s22}$	v_{s22}

Given, for instance, that $r = 0.2$, $p = 0.8$ and $q = 0.2$, then a, b and c are 0.1649, 0.1370 and 0.0385, respectively. The means for the haplotypes are $0.99v_{s11}$ and $0.01v_{s11}$ for (1 1), $0.2v_{s11}$ and $0.8v_{s11}$ for (1 2), $0.8v_{s11}$ and $0.2v_{s11}$ for (2 1) and $0.01(v_{s11})$ and $0.99(v_{s11})$ for (2 2). The approximate means are very similar to these estimates. The maximum errors associated with the above approximate means are when $p = q = 0.5$ for haplotypes (1 1) and (2 2)

(Goddard, 1992). Using the approximate means, the value of the MQTL in each gamete can be written in terms of the parental MQTL as:

$$\begin{pmatrix} v_{o11} \\ v_{o12} \\ v_{o21} \\ v_{o22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ q & p \\ p & q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_{s11} \\ v_{s22} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix} \qquad [8.36]$$

where ε_{ij} is the deviation of each gamete from the mean of the haplotype. Since v_{o11} is identical to v_{s11} and v_{o22} to v_{s22} with the approximate means, then $\varepsilon_{11} = \varepsilon_{22} = 0$.

Similarly to [8.20], equation [8.36] may be expressed as:

$$\mathbf{v} = \mathbf{P}\mathbf{v} + \boldsymbol{\varepsilon}$$

where $\mathbf{P}$ is as defined in [8.20] and has at most two non-zero elements which sum to unity. Thus:

$$\mathbf{v} = (\mathbf{I} - \mathbf{P})^{-1}\boldsymbol{\varepsilon}$$

Therefore:

$$\mathbf{G} = \text{var}(\mathbf{v}) = (\mathbf{I} - \mathbf{P})^{-1}\ \text{var}(\boldsymbol{\varepsilon})(\mathbf{I} - \mathbf{P})'^{-1}$$

and:

$$\mathbf{G}^{-1} = (\mathbf{I} - \mathbf{P})'\mathbf{H}^{-1}(\mathbf{I} - \mathbf{P}) \qquad [8.37]$$

where $\mathbf{H}\sigma_v^2 = \text{var}(\boldsymbol{\varepsilon})$ and $\mathbf{H}$ is a diagonal matrix. Since $\varepsilon_{11} = \varepsilon_{22} = 0$, $\text{var}(\varepsilon_{11}) = \text{var}(\varepsilon_{22}) = 0$. The main interest, therefore, is in calculating $\text{var}(\varepsilon_{12})$ and $\text{var}(\varepsilon_{21})$. The calculation of either $\text{var}(\varepsilon_{12})$ or $\text{var}(\varepsilon_{21})$ is similar to that for $\text{var}(\varepsilon)$ in Section 8.4. For instance, for the oth progeny:

$$\begin{aligned} \varepsilon_{o12} &= v_{o12} - q(v_{s11}) - p(v_{s22}) \\ \text{var}(\varepsilon_{o12}) &= \text{var}(v_{o12}) - q^2\ \text{var}(v_{s11}) - p^2\ \text{var}(v_{s22}) - 2qp\,\text{cov}(v_{s11}, v_{s22}) \\ &= \sigma_v^2 - (1-p)^2\sigma_v^2 + p^2\sigma_v^2 - 2(1-p)p\sigma_v^2 F_s \\ &= 2\sigma_v^2((1-p)p - (1-p)pF_s) \\ &= 2\sigma_v^2(1-p)p(1-F_s) = 2\sigma_v^2 pq(1-F_s) \\ \text{var}(\varepsilon_{o12})/\sigma_v^2 &= \mathbf{H} = 2pq(1-F_s) \end{aligned}$$

Therefore, if the sire is not inbred, the diagonal element of $\mathbf{H}$ for progeny o(h_{oo}) with the allele v_{o12} equals $2pq$. If the sire is unknown, $h_{oo} = 1$. Similarly, for a progeny o with allele v_{o21}, $h_{oo} = 2qp$ if the sire is known, otherwise 1 if the sire is unknown.

The matrix $\mathbf{G}$ can be calculated using rules similar to those defined in Section 8.3. The relationship of the MQTL paternal allele of a progeny o with MQTL alleles of individuals 1 to $(o-1)$ can be calculated using [8.7], with $\rho_o^p = p$ when o inherits marker haplotype v_{s12} or $\rho_o^p = (1-p)$ when o inherits marker haplotype v_{s21}. Similarly, for the maternal MQTL allele,

[8.8] can be used, with $\rho_o^m = p$ when o inherits marker haplotype v_{m12} or $\rho_o^m = (1-p)$ when o inherits marker haplotype v_{m21}.

Using [8.37], Goddard (1992) derived the following rules for calculating $\mathbf{G}_v^{-1}$:

1. Replace v_{o11} with v_{s11} in all equations and then delete the row and column for v_{o11} in $\mathbf{G}^{-1}$. Similarly replace v_{o22} with v_{s22}. Set $\mathbf{G}^{-1}$ to zero.
2. For progeny allele v_{o12}, add:

$q/2p$ to the element corresponding to (v_{s11}, v_{s11})
$p/2q$ to the element corresponding to (v_{s22}, v_{s22})
$1/2pq$ to the element corresponding to (v_{o12}, v_{o12})
$-1/2p$ to the element corresponding to (v_{s11}, v_{o12}) and (v_{o12}, v_{s11})
$-1/2q$ to the element corresponding to (v_{s22}, v_{o12}) and (v_{o12}, v_{s22})
$-1/2$ to the element corresponding to (v_{s11}, v_{s22}) and (v_{s22}, v_{s11})

3. For progeny allele v_{o21}, replace p with q and v_{o12} with v_{o21} in the rules above.
4. For an allele v_{s11} without known parents, add 1 to the element corresponding to (v_{s11}, v_{s11}).

Goddard (1992) indicated that the use of the approximate means to calculate **P** implies that v_{s11} and v_{o11} are forced to be identical even if double crossover occurs. Therefore it might be desirable to use a correlation (r) slightly less than unity between v_{s11} and v_{o11}. This is achieved by using:

$$v_{o11} = (1 - r^2/4)v_{s11} + r^2/4v_{s22} + \varepsilon_{11}$$

Then the row and column for v_{o11} are retained in $\mathbf{G}_v^{-1}$ and, in the above rules, v_{o12} is replaced by v_{o11} and p by $r^2/4$.

8.9.3 An illustration

Example 8.5
Consider that the first four calves in Example 8.3 have the following genotype at two linked loci.

			Genotype at the two linked markers	
Animal	Sire	Dam	Marker 1	Marker 2
1	–	–	11	22
2	–	–	33	44
3	1	3	12	44
4	4	3	21	14

Assuming no double crossing over, the same pedigree structure, genetic parameters and letting p and q be equal to 0.8 and 0.2, respectively, predict

the effects of the sex of the calf, additive genetic effects (breeding values) not linked to the MQTL for animals and additive genetic effects for the MQTL alleles of animals.

The alleles at the MQTL can be defined from the genotypes at the two linked marker loci. Thus the paternal and maternal MQTL alleles for animal 1 will be v_{s11} and v_{s22}, respectively. Correspondingly, those for animal 4 will be v_{o21} and v_{o14}, respectively. As in Example 8.3, $\alpha_1 = 0.6/0.3 = 2$ and $\alpha_2 = 0.6/0.05 = 12$. With the assumption of no double crossing over, for calf 3, $v_{o44} = v_{m44}$ (calf 2); therefore, the row and column for v_{o44} are deleted from $\mathbf{G}_v$ and MME.

The design matrix $\mathbf{Z}$ is an identity matrix of order four and $\mathbf{W}$ is:

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{W'W} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The covariance matrix $\mathbf{G}_v$ is:

$$\mathbf{G}_v = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.200 & 0.800 & 0.040 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.800 & 0.200 & 0.160 \\ 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.800 \\ 0.200 & 0.800 & 0.000 & 0.000 & 1.000 & 0.320 & 0.200 \\ 0.800 & 0.200 & 0.000 & 0.000 & 0.320 & 1.000 & 0.064 \\ 0.040 & 0.160 & 0.000 & 0.800 & 0.200 & 0.064 & 1.000 \end{bmatrix}$$

The calculation of $\mathbf{G}_v$ with elements $g(i,j)$ for the first few animals is as follows. For the first two animals, both parents are unknown; therefore the diagonal element of $\mathbf{G}$ for either the paternal or maternal allele is one for these animals. Calf 3 inherited marker haplotype v_{s12} from its sire; therefore $\rho_o^p = p$ in [8.7]. Assuming the same coding for $\mathbf{G}_v$ as in Example 8.1, then

$$g_{(3p3p,1p1p)} = (1-p)g_{(1p1p,1p1p)} + pg_{(1m1m,1p1p)} = q(1) + p(0) = q = 0.2$$
$$g_{(3p3p,1m1m)} = (1-p)g_{(1p1p,1m1m)} + pg_{(1m1m,1m1m)} = q(0) + p(1) = p = 0.8$$
$$g_{(3p3p,2p2p)} = (1-p)g_{(1p1p,2p2p)} + pg_{(1m1m,2p2p)} = q(0) + p(0) = 0$$
$$g_{(3p3p,2m2m)} = (1-p)g_{(1p1p,2m2m)} + pg_{(1m1m,2m2m)} = q(0) + p(0) = 0$$

The marker haplotype inherited by calf 4 from its sire is v_{s21}; therefore $\rho_o^p = q$ in [8.7]. Thus:

$$g_{(4p4p,1p1p)} = (1-q)g_{(1p1p,1p1p)} + qg_{(1m1m,1p1p)} = p(1) + q(0) = p = 0.8$$
$$g_{(4p4p,1m1m)} = (1-q)g_{(1p1p,1m1m)} + qg_{(1m1m,1m1m)} = p(0) + q(1) = q = 0.2$$
$$g_{(4p4p,2p2p)} = (1-q)g_{(1p1p,2p2p)} + qg_{(1m1m,2p2p)} = p(0) + q(0) = 0$$

$$g_{(4p4p,2m2m)} = (1-q)g_{(1p1p,2m2m)} + qg_{(1m1m,2m2m)} = p(0) + q(0) = 0$$
$$g_{(4p4p,3m3m)} = (1-q)g_{(1p1p,3m3m)} + qg_{(1m1m,3m3m)} = p(q) + q(p) = 2pq = 0.32$$

The inverse of $\mathbf{G}_V$ is:

$$\mathbf{G}_V^{-1} = \begin{bmatrix} 3.125 & 1.000 & 0.000 & 0.000 & -0.625 & -2.500 & 0.000 \\ 1.000 & 3.125 & 0.000 & 0.000 & -2.500 & -0.625 & 0.000 \\ 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 3.000 & 0.500 & 0.000 & -2.500 \\ -0.625 & -2.500 & 0.000 & 0.500 & 3.250 & 0.000 & -0.625 \\ -2.500 & -0.625 & 0.000 & 0.000 & 0.000 & 3.125 & 0.000 \\ 0.000 & 0.000 & 0.000 & -2.500 & -0.625 & 0.000 & 3.125 \end{bmatrix}$$

The matrix $\mathbf{G}^{-1}$ was computed using the rules outlined earlier. Thus, for the first two animals (first four alleles), add 1 to the diagonal elements since parents of both calves are unknown. For the paternal allele of calf 3, add $1/2pq$ to the diagonal element $(3p3p,3p3p)$ of G^{-1}, $q/2p$ to element $(1p1p,1p1p)$, $p/2q$ to element $(1m1m,1m1m)$, $-1/2p$ to elements $(1p1p,3p3p)$ and $(3p3p,1p1p)$, $-1/2q$ to elements $(1m1m,3p3p)$ and $(3p3p,1m1m)$ and 0.5 to elements $(1p1p,1m1m)$ and $(1m1m,1p1p)$.

The matrix $\mathbf{A}^{-1}$ for the example data has been given in Example 8.3; therefore the MME can easily be set up from the design matrices and inverse of the covariance matrices given. Solving the MME by direct inversion gave the following results:

Effects	Solutions
Sex of calf[a]	
1	7.475
2	5.091
Breeding values for animals	
1	0.034
2	–0.034
3	0.246
4	0.280
Additive effects for animals at the MQTL	
1*p*	–0.008
1*m*	0.005
2*p*	–0.047
2*m*	0.049
3*p*	0.024
4*p*	0.010
4*m*	0.059

[a]1 = male, 2 = female.

A similar model to that in Example 8.5 has been used by Boichard *et al.* (2002) for incorporating MQTL information into genetic evaluation for milk production traits in young bulls.

8.10 Reduced Animal Model

The use of the animal model (equation [8.35]) implies that, given t QTL loci, there are one u effect and two tv effects to be fitted per animal. The use of the reduced animal model could be very advantageous in reducing the number of equations as only parents are considered. The fitting of a RAM with MQTL information from multiple markers is the same as described in Section 8.6; therefore only a brief outline is presented. With RAM, MQTL additive genetic effects for non-parents are expressed as a function of MQTL effects of parents. For this model, the equation for a non-parent is:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{u}_p + \sum_j \mathbf{P}\mathbf{v}_{(p)j} + \mathbf{e}^*$$

where $\mathbf{Z}_1$ is as defined in Section 8.6, $\mathbf{P}$ is as defined in [8.20] and:

$$\mathbf{e}^* = \boldsymbol{\phi} + \sum_j \mathbf{W}_{nj}\boldsymbol{\varepsilon}_j + \mathbf{e}$$

Then:

$$\text{var}(\mathbf{e}^*) = \mathbf{I} + \mathbf{D}\alpha_1 + \sum_j \mathbf{W}_n\mathbf{H}\mathbf{W}_n'\alpha_2$$

with terms defined as in [8.22] and the matrix $\Sigma_j \mathbf{W}_n\mathbf{H}\mathbf{W}_n'$ for the jth locus is diagonal with the following elements:

$2p(1-p)(2-f_s-f_d)$ if both sire and dam of the non-parent are known
$2p(1-p)(1-f_s)+1$ when only the sire is known
$2p(1-p)(1-f_d)+1$ when only the dam is known
2 if both the sire and dam of the non-parents are unknown

The equations for parents are the same as in Section 8.6 and the MME similar to those in equation [8.23].

9 Non-additive Animal Models

The models considered in the previous chapters have dealt with only additive genetic effects. Henderson (1985) provided a statistical framework for modelling additive and non-additive genetic effects when there is no inbreeding. This chapter covers some of such models. The ability to separate non-additive genetic effects implies removal of some of the confounding that would otherwise bias the results from the analysis. Moreover, the availability of estimates of non-additive genetic effects for individuals could be used in mate selection, which would maximize the use of both additive and non-additive genetic variance. In this chapter, the prediction of dominance and epistatic effects using mixed model methodology is discussed. In practice, the application of non-additive models in genetic evaluation has been limited due to lack of genetic parameters and partly due to the fact that these effects tend to be highly confounded with others, such as common maternal environment.

9.1 Dominance Relationship Matrix

Dominance genetic effects result from the action of pairs of alleles at a locus on a trait. If two animals have the same set of parents or grandparents, it is possible that they possess the pair of alleles in common. The dominance relationship between two such animals represents the probability that they have the same pair of alleles in common. Thus, for a group of animals, the dominance genetic relationship matrix (**D**) among them can be set up. The dominance relationship between an individual x with parents s and d and an individual y with parents f and m in a non-inbred population can be calculated (Cockerham, 1954) as:

$$d_{xy} = 0.25(u_{sf}u_{dm} + u_{sm}u_{df}) \qquad [9.1]$$

where u_{ij} represents the additive genetic relationship between *i* and *j*. For instance, for two full-sibs with parents unrelated to each other:

$$d = 0.5[(1)(1) + (0)(0)] = 0.25$$

with the assumption that there is no common environmental variance.

Thus **D** can be generated from the additive genetic relationship. However, the prediction of dominance effects requires the inverse of **D**. This could be obtained by calculating **D** by [9.1] and inverting it; this is not computationally feasible with large data sets. Hoeschele and VanRaden (1991) developed a methodology for obtaining a rapid inversion of **D** and this is presented in Section 9.3. Initially, the principles involved in using $\mathbf{D}^{-1}$ from [9.1] for the prediction of dominance effects are discussed.

9.2 Animal Model with Dominance Effect

The model with dominance included is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wd} + \mathbf{e} \qquad [9.2]$$

where **y** = vector of observation, **b** = vector of fixed effects, **a** = vector for random animal additive genetic effects, **d** = vector of random dominance effects, and **e** = random residual error.

It is assumed that:

$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2, \quad \text{var}(\mathbf{d}) = \mathbf{D}\sigma_d^2 \quad \text{and} \quad \text{var}(\mathbf{e}) = \sigma_e^2$$
$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}' + \mathbf{WDW}' + \mathbf{I}\sigma_e^2$$

The mixed model equations (MME) to be solved for the best linear unbiased prediction (BLUP) of **a** and **d** and best linear unbiased estimator (BLUE) of **b** are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{W} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\alpha_1 & \mathbf{Z}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{Z} & \mathbf{W}'\mathbf{W} + \mathbf{D}^{-1}\alpha_2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix} \qquad [9.3]$$

with $\alpha_1 = \sigma_e^2/\sigma_a^2$ and $\alpha_2 = \sigma_e^2/\sigma_d^2$. However, we are interested in the total genetic merit (**g**) of the animal, which is $\mathbf{g} = \mathbf{a} + \mathbf{d}$. The MME could be modified such that the total genetic merit is solved for directly. Since $\mathbf{g} = \mathbf{a} + \mathbf{d}$, then:

$$\text{var}(\mathbf{g}) = \mathbf{G} = \mathbf{A}\sigma_a^2 + \mathbf{D}\sigma_d^2$$

The MME become:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}\sigma_e^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \qquad [9.4]$$

The individual components of **g** can be obtained as:

$$\hat{\mathbf{a}} = \sigma_a^2 \mathbf{A}\mathbf{G}^{-1}\hat{\mathbf{g}} \quad \text{and}$$

$$\hat{\mathbf{d}} = \sigma_d^2 \mathbf{D}\mathbf{G}^{-1}\hat{\mathbf{g}}$$

9.2.1 Solving for animal and dominance genetic effects separately

Example 9.1
Suppose the data below were the weaning weight for some piglets in a herd.

Pig	Sire	Dam	Sex[a]	Weaning weight (kg)
5	1	2	1	17.0
6	3	4	1	20.0
7	6	5	1	18.0
8	0	5	1	13.5
9	3	8	2	20.0
10	3	8	2	15.0
11	6	8	2	15.0
12	6	8	2	19.5

[a]1 = female, 2 = male (throughout chapter)

The aim is to estimate sex effects and predict solutions for animal and dominance genetic effects, assuming that $\sigma_e^2 = 120$, $\sigma_a^2 = 90$, $\sigma_d^2 = 80$. This has been illustrated below, solving for animal and dominance effects separately [9.3]. From the above parameters, $\alpha_1 = 1.333$ and $\alpha_2 = 1.5$.

Setting up the MME

The matrix **X** relates records to sex effects. Its transpose, considering only animals with records, is:

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

The matrices **Z** and **W** are both identity matrices since each animal has one record. The transpose of the vector of observations, $\mathbf{y}' = [17\ 20\ 18\ 1.5\ 20\ 15\ 25\ 19.5]$.

The other matrices in the MME, apart from $\mathbf{A}^{-1}$ and $\mathbf{D}^{-1}$, can be obtained through matrix multiplication from the matrices already calculated. The inverse of the additive relationship matrix is set up, using rules

outlined in Section 2.3.1. Using equation [9.1], the dominance relationship matrix is:

$$\mathbf{D} = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.063 & 0.063 & 0.125 & 0.125 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.063 & 0.000 & 1.000 & 0.250 & 0.125 & 0.125 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.063 & 0.000 & 0.250 & 1.000 & 0.125 & 0.125 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.125 & 0.000 & 0.125 & 0.125 & 1.000 & 0.250 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.125 & 0.000 & 0.125 & 0.125 & 0.250 & 1.000
\end{bmatrix}$$

and its inverse is:

$$\mathbf{D}^{-1} = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.028 & 0.000 & -0.032 & -0.032 & -0.096 & -0.096 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.032 & 0.000 & 1.084 & -0.249 & -0.080 & -0.080 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.032 & 0.000 & -0.249 & 1.084 & -0.080 & -0.080 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.096 & 0.000 & -0.080 & -0.080 & 1.092 & -0.241 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.096 & 0.000 & -0.080 & -0.080 & -0.241 & 1.092
\end{bmatrix}$$

The matrices $\mathbf{A}^{-1}\alpha_1$ and $\mathbf{D}^{-1}\alpha_2$ are added to $\mathbf{Z}'\mathbf{Z}$ and $\mathbf{W}'\mathbf{W}$ in the MME. The MME are of the order 26 by 26 and are too large to be presented. However, the solutions to the MME by direct inversion of the coefficient matrix are:

Effects	Solutions	
Sex		
1	16.980	
2	20.030	
Animal	BV[a]	DV[a]
1	–0.160	0.000
2	–0.160	0.000

(*Continued*)

Effects	Solutions	
3	0.059	0.000
4	0.819	0.000
5	–0.320	0.136
6	1.259	0.705
7	0.555	0.237
8	–0.998	–0.993
9	–0.350	0.000
10	–1.350	–1.333
11	1.061	1.428
12	–0.039	–0.038

[a]BV, DV = solutions for random animal and dominance effects, respectively.

The results indicate that males were heavier than females by about 3.05 kg at weaning. The breeding value for animal i, $\hat{a}_i$, from the MME can be calculated using [3.8] except that yield deviation is corrected not only for fixed effects but also for dominance effect. Thus the solution for animal 6 can be calculated as:

$$\begin{aligned}\hat{a}_6 &= n_1((\hat{a}_3 + \hat{a}_4)/2) + n_2(y_6 - \hat{b}_1 - \hat{d}_6) + n_3(2\hat{a}_{12} - \hat{a}_8) \\ &\quad + n_3(2\hat{a}_{11} - \hat{a}_8) + n_3(2\hat{a}_7 - \hat{a}_5) \\ &= n_1(0.059 + 0.819)/2 + n_2(20 - 16.980 - 0.705) + n_3(2(-0.039) - (-0.998)) \\ &\quad + n_3(2(1.061) - (-0.998)) + n_3(2(0.555) - (-0.320)) \\ &= 1.259\end{aligned}$$

where $n_1 = 2\alpha_1/wt$, $n_2 = 1/wt$, $n_3 = 0.5\alpha_1/wt$, with wt equal to the sum of the numerator of n_1, n_2 and 3(n_3).

The solution for dominance effect of animal i from the MME is:

$$\hat{d}_i = \left[-\alpha_2 \left(\sum_j c_{ij} \hat{d}_j \right) + (y_i - \hat{b}_k - \hat{a}_i) \right] \Big/ (n + c_{ii}\alpha_2)$$

where c_{ij} is the inverse element of **D** between animal i and j, and n is the number of records. For instance, the dominance effect of animal 6 is:

$$\hat{d}_6 = (0 + (20 - 16.980 - 1.259))/(1 + 1.5) = 0.705$$

The dominance effect for an individual represents interactions of pairs of genes from both parents and Mendelian sampling; it therefore gives an indication of how well the genes from two parents combine. This could be used in the selection of mates.

9.2.2 Solving for total genetic merit directly

Example 9.2
Using the same data and genetic parameters as in Example 9.1, solving directly for total genetic merit ($\hat{\mathbf{a}} + \hat{\mathbf{d}}$) applying equation [9.4] is illustrated.

SETTING UP THE MME

The design matrices **X** and **Z** are exactly the same as in equation [9.3]. However, in [9.4], $\mathbf{G} = \mathbf{A}\sigma_a^2 + \mathbf{D}\sigma_d^2$. The matrix **D** has been given earlier and **A** can be calculated as outlined in Section 2.1. Then $\mathbf{G}^{-1}\sigma_e^2$ is added to $\mathbf{Z}'\mathbf{Z}$ to obtain the mixed model equations [9.4]. Solving the mixed model equations by direct inversion of the coefficient matrix gives the following solutions:

Effects	Solutions
Sex	
1	16.980
2	20.030
Animal + dominance	
1	–0.160
2	–0.160
3	0.059
4	0.819
5	–0.184
6	1.963
7	0.792
8	–1.991
9	–0.349
10	–2.683
11	2.489
12	–0.078

The vector of solutions for additive genetic effects can then be calculated as $\hat{\mathbf{a}} = \sigma_a^2\mathbf{A}\mathbf{G}^{-1}\hat{\mathbf{g}}$ and as $\hat{\mathbf{d}} = \sigma_d^2\mathbf{D}\mathbf{G}^{-1}\hat{\mathbf{g}}$ dominance effects, as mentioned earlier. It should be noted that the sum of $\hat{a}_i$ and d_i for animal i in Example 9.1 equals the solution for animal i above, indicating that the two sets of results are equivalent. The advantage of using [9.4] is the reduction in the number of equations to be solved.

9.3 Method for Rapid Inversion of the Dominance Matrix

Hoeschele and VanRaden (1991) developed a method for computing directly the inverse of the dominance relationship matrix for populations

that are not inbred, by including sire and dam or sire and maternal grandsire subclass effects in the model. However, only the inclusion of sire and dam subclasses is considered in this text. Dominance effects result from interaction of pairs of genes and are not inherited through individuals. Since animals receive half of their genes from the sire and half from the dam, the dominance effect of an individual could be expressed as:

$$d = f_{S,D} + \varepsilon \qquad [9.5]$$

where f represents the average dominance effect of many hypothetical full-sibs produced by sire (S) and dam (D) and ε is the Mendelian sampling deviation of the individual from the S by D subclass effect. Variance of S by D subclass effects, σ_f^2, is equal to the covariance among full-sibs due to dominance, that is, $\sigma_f^2 = 0.25\sigma_d^2$; therefore $\text{var}(\varepsilon) = 0.75\sigma_d^2$. On the basis of [9.5] Hoeschele and VanRaden developed simple recurrence formulae for dominance effects using pairs of animals (sire and dam) and interaction between their parents.

For a particular sire and dam subclass (f_{SD}), the combination effect results from the interactions between the sire and the parents of D, interactions of the dam with the parents of S and interactions of the parents of S with the parents of D. Thus:

$$\begin{aligned} f_{SD} &= 0.5(f_{S,SD} + f_{S,DD} + f_{SS,D} + f_{DS,D}) \\ &\quad - 0.25(f_{SS,SD} + f_{SS,DD} + f_{DS,SD} + f_{DS,DD}) + e \end{aligned} \qquad [9.6]$$

where SS and DS denote sire, and dam of sire, respectively, and SD and DD corresponding parents for the dam. Equation [9.6] can also be obtained by regressing f_{SD} on its parent subclass effects as:

$$f_{SD} = \mathbf{b}'\mathbf{f}_{par} + e$$

where $\mathbf{f}_{par}$ is a vector of eight parent subclasses in [9.6] and $\mathbf{b}$ is a vector of corresponding partial regression coefficients with:

$$\mathbf{b}' = \text{cov}(f_{SD}, \mathbf{f}_{par})/\text{var}(\mathbf{f}_{par}) \qquad [9.7]$$

and:

$$\text{var}(e) = \sigma_f^2 - \mathbf{b}'\,\text{var}(\mathbf{f}_{par})\mathbf{b} \qquad [9.8]$$

The covariance between subclasses [9.7], for instance, between f_{SD} and f_{PM} is:

$$\text{cov}(f_{SD}, f_{PM}) = (a_{SP}a_{DM} + a_{SM}a_{DP})\sigma_f^2 \qquad [9.9]$$

with a_{ij} being the additive relationship between i and j. Thus:

$$\begin{aligned} \text{cov}(f_{SD}, f_{SS}, f_{DD}) &= (a_{S,SS}a_{D,DD} + a_{S,DD}a_{D,SS})\sigma_f^2 \\ &= (0.5(0.5)) + (0(0)) = 0.25\sigma_f^2 \end{aligned}$$

and:

$$\text{cov}(f_{SD}, f_{S,SD}) = (a_{SS}a_{D,SD} + a_{S,SD}a_{D,S})\sigma_f^2 = (1(0.5)) + (0(0)) = 0.5\sigma_f^2$$

If the nine subclasses in equation [9.6] are identified by 1, 2, 3, 4, 5, 6, 7, 8 and 9 (that is, $f_{SD} = 1$, $f_{S,SD} = 2$, etc.), the covariances between f_{SD} and its parent subclasses ($\text{cov}(f_{SD}, \mathbf{f}_{par})/\sigma_f^2$) using [9.9] are:

$$\begin{array}{cccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1[0.5 & 0.5 & 0.5 & 0.5 & 0.25 & 0.25 & 0.25 & 0.25] \end{array} \qquad [9.10]$$

and the relationship matrix among parent subclasses ($\text{var}(\mathbf{f}_{par})/\sigma_f^2$) (subclasses 2 to 9) using [9.9] is:

$$\begin{bmatrix} 1.0 & 0.0 & 0.25 & 0.25 & 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 1.0 & 0.25 & 0.25 & 0.0 & 0.5 & 0.0 & 0.5 \\ 0.25 & 0.25 & 1.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.25 & 0.25 & 0.0 & 1.0 & 0.0 & 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.5 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \qquad [9.11]$$

From the two matrices above ([9.10] and [9.11]), the regression coefficients [9.7] are:

$$\mathbf{b}' = [0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad -0.25 \quad -0.25 \quad -0.25 \quad -0.25] \qquad [9.12]$$

which are identical to the coefficients in equation [9.6]. It should be noted that there is no need to add more remote ancestors of *S* and *D* as the partial regression of these is zero.

9.3.1 Inverse of the relationship matrix of subclass effects

The recurrences in equation [9.6] could be represented as:

$$\mathbf{f} = \mathbf{Q}\mathbf{f} + \boldsymbol{\varepsilon} \qquad [9.13]$$

where $\mathbf{f}$ is the vector of sire by dam subclasses and the row *i* of $\mathbf{Q}$ contains the elements of $\mathbf{b}$ from equation [9.7] in columns pertaining to identified parent subclasses of subclass *i*. The relationship matrix for subclasses in $\mathbf{f}$ is $\mathbf{F} = \text{var}(\mathbf{f})/\sigma_f^2$. From equation [9.13]:

$$\mathbf{f} = (\mathbf{I} - \mathbf{Q})^{-1}\boldsymbol{\varepsilon}$$

The variance–covariance of $\mathbf{f}$ is:

$$\text{var}(\mathbf{f}) = \mathbf{F}\sigma_f^2 = (\mathbf{I} - \mathbf{Q}')^{-1}\mathbf{R}(\mathbf{I} - \mathbf{Q})^{-1}\sigma_f^2$$

with:

$$\mathbf{R}\sigma_f^2 = \text{var}(\boldsymbol{\varepsilon})$$

Therefore:

$$\mathbf{F}^{-1} = (\mathbf{I} - \mathbf{Q}')^{-1}\mathbf{R}^{-1}(\mathbf{I} - \mathbf{Q}) \qquad [9.14]$$

The diagonal elements **R** can be obtained from equation [9.8]. The off-diagonals are zeros if all ancestors subclasses providing relationship ties are included in **f**. To ensure a diagonal **R**, Hoeschele and VanRaden specified two conditions to be used in deciding which subclasses should be included in **f** as known. These are:

1. A subclass should remain in **f** if any of its parent subclasses remain in **f**.
2. A subclass should remain in **f** if **f** contains two or more of its immediate progeny subclasses.

Equation [9.14] implies that $\mathbf{F}^{-1}$ can be calculated from a list of subclasses and their parent subclass effects by computing for the *i*th subclass, r^{ii} (the diagonal element *i* of $\mathbf{R}^{-1}$) and c_i (the *i*th row of $(\mathbf{I} - \mathbf{Q})$). Then the contribution of the *i*th subclass to $\mathbf{F}^{-1}$ is calculated as $c_i c_i' r^{ii}$. In summary, the following procedure could therefore be used to calculate $\mathbf{F}^{-1}$:

1. List animals and their sires and dams. Parents not in the list of animals with more than one progeny should be added to the list while those with one progeny may be treated as unknown.
2. Form a list of all filled (*S* and *D* known) subclasses and add ancestor subclasses that provide ties. Ancestors are identified by listing subclasses for the sire with parents of the dam and for the dam with parents of the sire for each filled subclass and then repeating this process for the subclasses just added until no further ancestors are known. The same sex subclasses of animal *i* with animal *j* and of animal *j* with animal *i* should be treated as identical when listing ancestor subclasses. The list of subclasses is sorted such that progeny subclass precedes its parent subclasses. Commencing with the oldest ancestor subclass, subclasses could be regarded as unknown if they are not filled, have no known parents and provide no ties for at least two filled descendant subclasses.

The number of connections provided by an ancestor subclass may be approximately determined from counts formed when ancestor subclasses are being identified originally. Progeny subclass (f_{SD}) would contribute 1 to parent subclasses of type $f_{S,SD}$ and $f_{SS,D}$ but −1 to parent subclasses of type $f_{SS,SD}$. The subtraction of 1 is due to the fact that $f_{S,SD}$ and $f_{SS,D}$ are regarded as progeny subclasses of $f_{SS,SD}$ and both may have come from one f_{SD}. It should be noted, however, that some subclasses which should be deleted for having a count of less than two may be needed in order to achieve a diagonal **R**. Thus, if both $f_{S,SD}$ and $f_{SS,D}$ are known, for instance, it may be necessary to add back subclasses of type $f_{SS,SD}$ if they have been deleted for a count of less than two.
3. Go through the list of all subclasses and calculate contributions (coefficients) of each subclass *i* to $\mathbf{F}^{-1}$ as $r^{11} c_i c_i'$. The vector $\mathbf{c}_i$ contains non-zero coefficients, which is equal to 1 in subclass *i* and equal to **–b** for parent subclasses, with **b** computed as in equation [9.7].
4. Sort the coefficients by columns within rows and sum those with identical columns and rows to obtain $\mathbf{F}^{-1}$.

9.3.2 Prediction of dominance effects

So far, the discussion has been on the inverse of the relationship matrix for subclass effects but the major interest is the prediction of dominance effects.

Since the inheritance of dominance effects is from subclass effects, dominance effects can be predicted by the inclusion of the inverse of the relationship matrix ($\mathbf{D}_*$) among dominance effects and subclass effects in the mixed model equations. From [9.5] and [9.13] the dominance ($\mathbf{d}$) and subclass effect ($\mathbf{f}$) may be predicted as:

$$\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix} = \begin{bmatrix}0 & \mathbf{S}\\ 0 & \mathbf{Q}\end{bmatrix}\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{e}\end{bmatrix}$$

with:

$$\text{var}\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix} = \mathbf{D}_*\sigma_d^2$$

and:

$$\text{var}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}0.75\mathbf{I} & 0\\ 0 & 0.25\mathbf{R}\end{bmatrix}\sigma_d^2$$

where $\mathbf{S}$ is the incidence matrix relating $\mathbf{d}$ to $\mathbf{f}$, and $\boldsymbol{\beta}$ equals $\mathbf{d}$ minus $\mathbf{Sf}$. Therefore:

$$\begin{bmatrix}\mathbf{d}\\ \mathbf{f}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & -\mathbf{S}\\ 0 & \mathbf{I}-\mathbf{Q}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{e}\end{bmatrix}$$

and the inverse of $\mathbf{D}_*$ can be computed as:

$$\mathbf{D}_*^{-1} = \begin{bmatrix}\mathbf{I} & 0\\ -\mathbf{S}' & \mathbf{I}-\mathbf{Q}'\end{bmatrix}\begin{bmatrix}(4/3)\mathbf{I} & 0\\ 0 & 4\mathbf{R}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I} & -\mathbf{S}\\ 0 & \mathbf{I}-\mathbf{Q}\end{bmatrix} \qquad [9.15]$$

From the above, the inverse of $\mathbf{D}_*$ is similar to $\mathbf{F}^{-1}$ with coefficients of $\frac{4}{3}$ on the diagonals of dominance effects, $-\frac{4}{3}$ of off-diagonals linking dominance to subclass effects and the coefficients contributed by the subclass effects are multiplied by 4. The matrix $\mathbf{D}_*^{-1}$ can then be included in the mixed model equations, resulting in the prediction of both dominance and subclass effects. The only disadvantage is that the inclusion of subclass effects in the MME will increase the order of equations but the method can easily be applied to large data sets.

9.3.3 Calculating the inverse of the relationship matrix among dominance and subclass effects for example data

Example 9.3

Using the pedigree information in Example 9.1, the calculations of $\mathbf{F}^{-1}$ and $\mathbf{D}_*^{-1}$ are illustrated.

SETTING UP $\mathbf{F}^{-1}$

Application of rules 1 and 2 in Section 9.3.1 for calculating $\mathbf{F}^{-1}$ generated Table 9.1. Creating a list of filled subclasses in the first pass (pass 1) through the pedigree in reverse order generated subclasses A to E (sorted by sire) in Table 9.1. Passes 2 and 3 through this list identified all ancestor subclasses (subclasses F to N). Counts to determine whether ancestor subclasses are treated as known or unknown were calculated as specified earlier. Subclasses of the types $f_{S,SD}$ and $f_{SS,SD}$ received a count of 1 and –1, respectively, from progeny subclass f_{SD}. Thus subclass $f_{3,5}$ received a count of 1 from each of its progeny subclasses, $f_{3,8}$ and $f_{6,5}$, and a count of –1 from $f_{6,8}$. Again $f_{4,1}$ received 1 each from $f_{6,1}$ and $f_{4,5}$ and –1 from $f_{6,5}$. Proceeding through the ancestor subclasses (F to N), those with a count of 1 and with less than two progeny subclasses known are regarded as unknown. Only the ancestor subclass $f_{3,5}$ is regarded as known because two of its progeny subclasses ($f_{3,8}$ and $f_{6,5}$) were known although it had a count of 1.

Using rule 3, the contribution of subclass i regarded as known (subclasses 1 to 6 (see Table 9.1)) to $\mathbf{F}^{-1}$ is then calculated as $\mathbf{c}_i \mathbf{c}_i' r^{ii}$. For example, for the subclass $f_{6,8}$ (subclass 1), three parent subclasses are known, 2, 3 and 6, which are of the subclass type, $f_{S,DD}$, $f_{SS,D}$ and $f_{SS,DD}$, respectively. Therefore $\mathbf{b}_1' = [0.5 \quad 0.5 \quad -0.25]$, $\mathbf{c}_1' = [1 - \mathbf{b}_1'] = [1 - 0.5 - 0.5 - 0.25]$. The matrix $\mathbf{F}_1$, the

Table 9.1. List of filled sire × dam subclasses and ancestor subclasses.

	Sire × dam subclass		Pass subclass	Counts from progeny			Known parent
Φ	S	D	added	subclasses	Status	φ	subclasses
A	6	8	1		KN	1	2 3 6
B	6	5	1	1	KN	2	3 6
C	3	8	1	1	KN	3	6
D	3	4	1		KN	4	
E	1	2	1		KN	5	
F	4	8	2	1	UK		
G	3	5	2	1 + 1 – 1 = 1	KN	6	
H	6	1	2	1	UK		
I	6	2	2	1	UK		
J	4	5	2	1 + 1 – 1 = 1	UK		
K	3	1	3	1 + 1 – 1 = 1	UK		
L	3	2	3	1 + 1 – 1 = 1	UK		
M	4	1	3	1 + 1 – 1 = 1	UK		
N	4	2	3	1 + 1 – 1 = 1	UK		

Φ = consecutive label for subclasses; S, sire; D, dam; KN, known, UK, regarded as unknown; φ = consecutive number for known subclasses.

relationship among parent subclasses 2, 3 and 6 (see [9.13]) is:

$$\mathbf{F}_1 = \begin{array}{c} \\ 2 \\ 3 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 3 & 6 \end{array} \\ \begin{bmatrix} 1.00 & 0.25 & 0.50 \\ 0.25 & 1.00 & 0.50 \\ 0.50 & 0.50 & 1.00 \end{bmatrix} \end{array}$$

The contribution of $f_{6,8}$ to $\mathbf{F}^{-1}$ therefore is:

$$\mathbf{c}_1\mathbf{c}_1' r^{11} = \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 6 \end{array} \\ \begin{bmatrix} 1.00 & -0.50 & -0.50 & 0.25 \\ -0.50 & 0.25 & 0.25 & -0.125 \\ -0.50 & 0.25 & 0.25 & -0.125 \\ 0.25 & -0.125 & 0.125 & 0.0625 \end{bmatrix} \end{array} 1.778$$

where $r^{11} = 1/(1-(\mathbf{b}_1'\mathbf{F}_1\mathbf{b}_1)) = 1/(1-0.4375) = 1.778$ (see equation [9.8]).

Processing of all subclasses gives $\mathbf{F}^{-1}$ as:

$$\mathbf{F}^{-1} = \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1.778 & -0.889 & -0.889 & 0.000 & 0.000 & 0.445 \\ -0.889 & 1.778 & 0.445 & 0.000 & 0.000 & -0.889 \\ -0.889 & 0.445 & 1.778 & 0.000 & 0.000 & -0.889 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 \\ 0.445 & -0.889 & -0.889 & 0.000 & 0.000 & 1.778 \end{bmatrix} \end{array}$$

The methodology can be verified by calculating the dominance relationship matrix among animals as $\mathbf{D} = (0.25)\mathbf{SFS}' + \mathbf{I}(0.75)$, which should give the same **D** as that calculated using [9.1]. **S**, as defined earlier, relates dominance effects to subclass effects. For the example pedigree:

$$\mathbf{S}' = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccccccc} 5 & 6 & 7 & 9 & 10 & 11 & 12 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

and:

$$\mathbf{D} = (0.25)\mathbf{SFS}' + \mathbf{I}(0.75)$$

$$
\begin{array}{c}
\begin{matrix} 5 & 6 & 7 & 9 & 10 & 11 & 12 \end{matrix} \\
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0.0625 & 0.0625 & 0.125 & 0.125 \\
0 & 0 & 0.0625 & 1 & 0.25 & 0.125 & 0.125 \\
0 & 0 & 0.0625 & 0.25 & 1 & 0.125 & 0.125 \\
0 & 0 & 0.125 & 0.125 & 0.125 & 1 & 0.25 \\
0 & 0 & 0.125 & 0.125 & 0.125 & 0.25 & 1
\end{bmatrix}
\end{array}
$$

which is the same as the **D** (Section 9.2.1) calculated from the pedigree using [9.1].

Let $\mathbf{D}_*^{-1}$ be partitioned as:

$$\mathbf{D}_*^{-1} = \begin{bmatrix} \mathbf{D}_{*11}^{-1} & \mathbf{D}_{*12}^{-1} \\ \mathbf{D}_{*21}^{-1} & \mathbf{D}_{*22}^{-1} \end{bmatrix}$$

where $\mathbf{D}_{*11}^{-1}$ is the top 12 by 12 block for dominance effects for animals, $\mathbf{D}_{*22}^{-1}$ is the bottom 6 by 6 block for subclass effects and $\mathbf{D}_{*12}^{-1}$ is the block for dominance by subclass effects. For the example data using [9.15], the sub-matrices of $\mathbf{D}_*^{-1}$ are:

$$\mathbf{D}_{*11}^{-1} = \text{diag}(4/3, 4/3, 4/3, 4/3, 4/3, 4/3, 4/3, 4/3, 4/3, 4/3, 4/3, 4/3)$$

$$
\mathbf{D}_{*21}^{-1} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.333 & -2.333 \\
0 & 0 & 0 & 0 & 0 & 0 & -1.333 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.333 & -1.333 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1.333 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1.333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

$\mathbf{D}_{*12}^{-1}$ is the transpose of $\mathbf{D}_{*21}^{-1}$

$$
\mathbf{D}_{*22}^{-1} = \begin{bmatrix}
7.111 & -3.556 & -3.556 & 0 & 0 & 1.778 \\
-3.556 & 7.111 & 1.778 & 0 & 0 & -3.556 \\
-3.556 & 1.778 & 7.111 & 0 & 0 & -3.556 \\
0 & 0 & 0 & 4.0 & 0 & 0 \\
0 & 0 & 0 & 0 & 4.0 & 0 \\
1.778 & -3.556 & -3.556 & 0 & 0 & 7.111
\end{bmatrix}
$$

The matrix $\mathbf{D}_*^{-1}$ can be included in the usual mixed model equations for the prediction of dominance and subclass effects.

9.4 Epistasis

Epistasis refers to the interaction among additive and dominance genetic effects, for instance, additive by additive, additive by dominance, additive by additive by dominance, etc. The epistasis relationship matrix can be derived from **A** and **D** as:

A#A for additive by additive
D#D for dominance by dominance
AA#D for additive by additive by dominance

where # represents the Hadamard product of the two matrices. The *ij* element of the Hadamard product of the two matrices is the product of the *ij* elements of the two matrices. Thus, if $\mathbf{M} = \mathbf{A}\#\mathbf{B}$, then $m_{ij} = (a_{ij})(b_{ij})$ where the matrices **A** and **B** should be of the same order.

The model (equation [9.2]) can be expanded to include epistatic effects as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wd} + \mathbf{Sep} + \mathbf{e}$$

where **ep** is the vector of interaction (epistatic) effects. The evaluation can be carried out as described in Section 9.2 but the major limitation is obtaining the inverse of the epistatic relationship matrix for large data sets. However, VanRaden and Hoeschele (1991) presented a rapid method for obtaining the inverse of epistatic relationship matrix when epistasis results from interactions between additive by additive ($\mathbf{A} \times \mathbf{A}$) genetic effects and the population is inbred or is not. The approach is similar to the method described for obtaining the inverse of the dominance relationship matrix and it involves including sire × dam subclasses; consequently, the details of the method are not covered in this text. The method involves calculating the inverse of **U**, the relationship matrix among epistatic and subclass effects, and $\mathbf{U}^{-1}$ is then included in the usual mixed model equations for the prediction of epistatic and sire × dam subclass effects.

The rules for obtaining $\mathbf{U}^{-1}$ for a population that is not inbred are given in the next section, with an illustration.

9.4.1 Rules for the inverse of the relationship matrix for epistatic and subclass effects

The inverse of **U** can be computed by going through a list of individuals and their parents and sire × dam subclasses. See rules 1 and 2 in Section 9.3.1 on how such a list should be set up. The contribution of individual *i* in the list to $\mathbf{U}^{-1}$ is computed by the following rules:

1. For an individual *i* with sire (*s*) and dam (*d*) and subclass effects (*s*, *d*) known, the contribution to $\mathbf{U}^{-1}$ is

$$\begin{array}{c} \begin{matrix} c & s & d & (s,d) \end{matrix} \\ \begin{bmatrix} 16 & -4 & -4 & -16 \\ -4 & 1 & 1 & 4 \\ -4 & 1 & 1 & 4 \\ -16 & 4 & 4 & 16 \end{bmatrix} \end{array} (1/12) \qquad [9.16]$$

2. For an individual with both parents known but subclass effects treated as unknown, the contribution to $\mathbf{U}^{-1}$ is:

$$\begin{matrix} c & s & d \end{matrix}$$
$$\begin{bmatrix} 16 & -4 & -4 \\ -4 & 1 & 1 \\ -4 & 1 & 1 \end{bmatrix}(1/14) \qquad [9.17]$$

3. If only one parent, say, s_p is known, the contribution is:

$$\begin{matrix} c & s \end{matrix}$$
$$\begin{bmatrix} 16 & -4 \\ -4 & 1 \end{bmatrix}(1/15) \qquad [9.18]$$

4 If both parents and subclass are unknown, add 1 to the individual diagonal.

5. For sire × dam subclasses, the contribution of the *i*th subclass to $\mathbf{U}^{-1}$ is the same as for the inverse of the dominance matrix (see rule 3 in Section 9.3.1) except that the coefficients are multiplied by 8.

6. Sort coefficients by row and by columns within row, and sum coefficients with identical rows and columns to obtain $\mathbf{U}^{-1}$.

The method can be verified by inverting $\mathbf{U}^{-1}$ to form $\mathbf{U}$. The animal by animal submatrix of $\mathbf{U}$ should be equal to the epistatic relationship matrix calculated as $\mathbf{A\#A}$.

9.4.2 Calculating the inverse relationship matrix for epistasis and the subclass matrix for an example pedigree

Example 9.4

The calculation of $\mathbf{U}^{-1}$ is illustrated below using the pedigree information in Example 9.1.

The identification of sire and dam subclasses and their ancestor subclasses treated as known has been discussed in Section 9.3.3. Thus the list of animals and known subclasses is:

Animal	Sire	Dam
1	0	0
2	0	0
3	0	0
4	0	0
5	1	2
6	3	4
7	6	5
8	0	5
9	3	8

(*Continued*)

(*Continued*)

Animal	Sire	Dam
10	3	8
11	6	8
12	6	8

Subclasses	Parent subclasses
6,8	6,5, 3,8, 3,5
6,5	3,8, 3,5
3,8	3,5
3,4	
1,2	
3,5	

In setting up $\mathbf{U}^{-1}$, animals 1 to 12 have been regarded as rows 1 to 12 while subclasses have been assigned rows 13 (subclass (6,8)) to 18 (subclass (3,5)). The first four animals have both parents and sire–dam subclasses unknown and therefore each contributes 1 to their respective diagonals. The parents of animals 5, 6, 7, 10, 11 and 12 and their sire × dam subclass effects are known, therefore the contributions of each of these animals to $\mathbf{U}^{-1}$ are computed using rule 1 in Section 9.4.1. For animals 8 and 9, only one of their parents is known and rule 3 is applicable when processing these animals. The calculation of the contributions of subclass effects has been given in Section 9.3.3 (Example 9.3); these are multiplied by 8, as mentioned earlier. After processing all animals and subclass effects, the top 12 by 12 submatrix of $\mathbf{U}^{-1}$ (block for animals only) is:

$$\begin{bmatrix}
1.083 & 0.083 & 0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.083 & 1.083 & 0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.250 & 0.083 & 0.000 & -0.333 & 0.000 & 0.167 & -0.333 & -0.333 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.083 & 1.083 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
-0.333 & -0.333 & 0.000 & 0.000 & 1.483 & 0.083 & -0.333 & -0.267 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & -0.333 & -0.333 & 0.083 & 1.583 & -0.333 & 0.167 & 0.000 & 0.000 & -0.333 & -0.333 \\
0.000 & 0.000 & 0.000 & 0.000 & -0.333 & -0.333 & 1.333 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.167 & 0.000 & -0.267 & 0.167 & 0.000 & 1.400 & -0.333 & -0.333 & -0.333 & -0.333 \\
0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 1.333 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 0.000 & 1.333 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 0.000 & -0.333 & 0.000 & 0.000 & 1.333 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.333 & 0.000 & -0.333 & 0.000 & 0.000 & 0.000 & 1.333
\end{bmatrix}$$

The top 12 by 12 submatrix of the inverse $\mathbf{U}^{-1}$ is the epistatic relationship matrix for the animals and is:

$$\begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.250 & 0.000 & 0.063 & 0.063 & 0.016 & 0.016 & 0.016 & 0.016 \\
0.000 & 1.000 & 0.000 & 0.000 & 0.250 & 0.000 & 0.063 & 0.063 & 0.016 & 0.016 & 0.016 & 0.016 \\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.250 & 0.063 & 0.000 & 0.250 & 0.250 & 0.063 & 0.063 \\
0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.250 & 0.063 & 0.000 & 0.000 & 0.000 & 0.063 & 0.063 \\
0.250 & 0.250 & 0.000 & 0.000 & 1.000 & 0.000 & 0.250 & 0.250 & 0.063 & 0.063 & 0.063 & 0.063 \\
0.000 & 0.000 & 0.250 & 0.250 & 0.000 & 1.000 & 0.250 & 0.000 & 0.063 & 0.063 & 0.250 & 0.250 \\
0.063 & 0.063 & 0.063 & 0.063 & 0.250 & 0.250 & 1.000 & 0.063 & 0.063 & 0.063 & 0.141 & 0.141 \\
0.063 & 0.063 & 0.000 & 0.000 & 0.250 & 0.000 & 0.063 & 1.000 & 0.250 & 0.250 & 0.250 & 0.250 \\
0.016 & 0.016 & 0.250 & 0.000 & 0.063 & 0.063 & 0.063 & 0.250 & 1.000 & 0.250 & 0.141 & 0.141 \\
0.016 & 0.016 & 0.250 & 0.000 & 0.063 & 0.063 & 0.063 & 0.250 & 0.250 & 1.000 & 0.141 & 0.141 \\
0.016 & 0.016 & 0.063 & 0.063 & 0.063 & 0.250 & 0.141 & 0.250 & 0.141 & 0.141 & 1.000 & 0.250 \\
0.016 & 0.016 & 0.063 & 0.063 & 0.063 & 0.250 & 0.141 & 0.250 & 0.141 & 0.141 & 0.250 & 1.000
\end{bmatrix}$$

It is equal to the epistatic relationship matrix calculated as $\mathbf{A}\#\mathbf{A}$. The matrix $\mathbf{U}^{-1}$ can then be incorporated in the usual mixed model equations for the prediction of epistatic and subclass effects.

10 Analysis of Ordered Categorical Traits

Some traits of economic importance in animal breeding, such as calving ease or litter size, are expressed and recorded in a categorical fashion. For instance, in the case of calving ease, births may be assigned to one of several distinct classes, such as difficult, assisted and easy calving, or litter size in pigs might be scored 1, 2, 3 or more piglets born per sow. Usually, these categories are ordered along a gradient. In the case of calving ease, for example, the responses are ordered along a continuum measuring the ease with which birth occurred. These traits are therefore termed ordered categorical traits. Such traits are not normally distributed and animal breeders have usually attributed the phenotypic expression of categorical traits to an underlying continuous unobservable trait which is normally distributed, referred to as the liability (Falconer and MacKay, 1996). The observed categorical responses are therefore due to animals exceeding particular threshold levels (t_i) of the underlying trait. Thus, with m categories of responses, there are $m-1$ thresholds such that $t_1 < t_2 < t_3, \ldots, t_{m-1}$. For traits such as survival to a particular age or stage, the variate to be analysed is coded 1 (survived) or 0 (not survived) and there is basically only one threshold.

Linear and non-linear models have been applied for the genetic analysis of categorical traits with the assumption of an underlying normally distributed liability. Usually, the non-linear (threshold) models are more complex and have higher computing requirements. The advantage of the linear model is the ease of implementation as programmes used for analysis of quantitative traits could be utilized without any modifications. However, Fernando *et al.* (1983) indicated that some of the properties of best linear unbiased prediction (BLUP) do not hold with categorical traits. Such properties include the invariance of BLUP to certain types of culling (selection) and the ability of BLUP to maximize the probability of correct pairwise ranking. Also Gianola (1982) indicated that the variance of a

categorical trait is a function of its expectation and the application of a linear model that has fixed effects in addition to an effect common to all observations results in heterogeneity of variance.

In a simulation study, Meijering and Gianola (1985) demonstrated that, with no fixed effects and a constant or variable number of offspring per sire, an analysis of a binary trait with either a linear or non-linear model gave similar sire rankings. This was independent of the heritability of the liability or incidence of the binary trait. However, with the inclusion of fixed effects and a variable number of progeny per sire, the non-linear model gave breeding values which were more similar to the true breeding values compared with the linear model. The advantage of the threshold model increased as the incidence of the binary trait and its heritability decreased. Thus, for traits with low heritability and low incidence, a threshold model might be the method of choice.

The principles required to apply a linear model for the analysis of categorical traits are the same as discussed in the previous chapters; therefore the main focus of this chapter is on threshold models, assuming a normal distribution for the liability. Cameron (1997) illustrated the analysis of a binary trait with a threshold model, using a logit function. In this chapter sample data used for the illustration of the threshold model have also been analysed with a linear model.

10.1 The Threshold Model

10.1.1 Defining some functions of the normal distribution

The use of the threshold model involves the use of some functions of the normal distribution and these are briefly defined. Assume the number of lambs born alive to ewes in the breeding season is scored using four categories. The distribution of the liability for the number of lambs born alive with three thresholds (t_i) can be illustrated as in Fig. 10.1, where N_j is the number of ewes with the jth number of lambs and these exceed the threshold point t_{j-1}, with $j > 1$ and $j \leq m - 1$.

With the assumption that the liability (l) is normally distributed ($l \sim N(0,1)$), the height of the normal curve at t_j ($\phi(t_j)$) is:

$$\phi(t_j) = \exp(-0.5t_j^2)/\sqrt{2\pi} \qquad [10.1]$$

For instance, given that $t_j = 0.779$, then $\phi(0.779) = 0.2945$.

The function $\Phi()$ is the standard cumulative distribution function of the normal distribution. Thus $\Phi(k)$ or Φ_k gives the areas under the normal curve up to and including the kth category. Given that there are m categories, then $\Phi_k = 1$ when the kth category equals m. For a variable x, for instance, drawn from a normal distribution, the value Φ_x can be computed, using a subroutine from the IMSL (1980) library. Thus, if $x = 0.560$, then $\Phi(0.560) = 0.7123$.

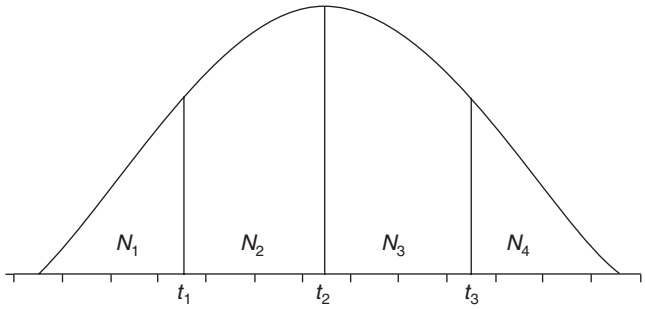


Fig. 10.1. The distribution of liability for number of lambs born alive with four categories and three thresholds.

$P(k)$ defines the probability of a response being observed in category k assuming a normal distribution. This is also the same probability that a response is between the thresholds defined by category k. Thus $P(k)$ or P_k may be calculated as $P(k) = \Phi(k) - \Phi(k-1)$ with $\Phi(k-1) = 0$, when $k = 1$; or, expressed in terms of thresholds defining the category k, $P_k = \Phi t_k - \Phi t_{(k-1)}$. For instance, in Fig. 10.1, the probability of response in the k category (P_k) can be computed as:

$$P_1 = \Phi(t_1) \quad [10.2]$$

$$P_2 = \Phi(t_2) - \Phi(t_1) \quad [10.3]$$

$$P_3 = \Phi(t_3) - \Phi(t_2)$$

$$P_4 = 1 - \Phi(t_4)$$

10.1.2 Data organization and the threshold model

Usually the data are organized into an s by m contingency table (Table 10.1), where the s rows represent individuals or herd–year subclasses of effects, such as herd, and the m columns indicate ordered categories of response. If the rows represent individuals, then all n_{jk} will be zero except one and the $n_{j.} = 1$, for $j = 1, \ldots, s$.

The linear model for the analysis of the liability is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of liability on a normal scale, $\mathbf{b}$ and $\mathbf{u}$ are vectors of fixed and random (sire or animal) effects, respectively, and $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating data to fixed effects and response effects, respectively. Since $\mathbf{y}$ is not observed, it is not possible to solve for $\mathbf{u}$ using the usual mixed model equations. Given that $\mathbf{H}' = [\mathbf{t}', \mathbf{b}', \mathbf{u}']$, where $\mathbf{t}$ is the vector for the threshold effects, Gianola and Foulley (1983) proceeded to

Table 10.1. Ordered categorical data arranged as an *s* by *m* contingency table.

	Categories[a]					
Subclasses	1	2	⋯	*k*	⋯	*m*	Totals[b]
1	n_{11}	n_{12}	⋯	n_{1k}	⋯	n_{1m}	$n_{1.}$
2	n_{21}	n_{22}	⋯	n_{2k}	⋯	n_{2m}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮		⋮	⋮
j	n_{j1}	n_{j2}	⋯	n_{jk}	⋯	n_{jm}	$n_{j.}$
⋮	⋮	⋮	⋮	⋮		⋮	⋮
s	n_{s1}	n_{s2}	⋯	n_{sk}	⋯	n_{sm}	$n_{s.}$

[a] n_{jk} = number of counts in category *k* of response in row *j*.
[b] $n_{j.} = \Sigma_{k=1}^{m} n_{jk}$.

find the estimator $\hat{\mathbf{H}}$ that maximizes the log of the posterior density L(**H**). The resulting set of equations involved in the differentiation was not linear with respect to **H**. They therefore provided the following non-linear iterative system of equations based on the first and second derivatives, assuming a normal distribution, to obtain solutions for $\Delta\mathbf{t}$, $\Delta\mathbf{b}$ and $\Delta\mathbf{u}$

$$\begin{bmatrix} \mathbf{Q} & \mathbf{L'X} & \mathbf{L'Z} \\ \mathbf{X'L} & \mathbf{X'WX} & \mathbf{X'WZ} \\ \mathbf{Z'L} & \mathbf{Z'WX} & \mathbf{Z'WZ} + \mathbf{A}^{-1}\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{b} \\ \Delta\mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{p} \\ \mathbf{X'v} \\ \mathbf{Z'v} - A^{-1}G^{-1}u \end{bmatrix} \qquad [10.4]$$

with $\mathbf{G} = \mathbf{I}\sigma_s^2$ or $\mathbf{I}\sigma_u^2$ if a sire or an animal model is being fitted in a univariate situation. They presented equations for the calculation of the matrices in equation [10.4], which are outlined below. The calculation of most of these matrices involves P_{jk} (see equation [10.2]) and it is initially described. P_{jk}, the response in the *k*th category under the conditions of the *j*th row, is:

$$P_{jk} = \Phi(t_k - a_j) - \Phi(t_{k-1} - a_j); \quad k = 1, m-1; \; j = 1, \ldots, s \qquad [10.5]$$

where $a_j = (x_j b + z_j u)$, with x_j and z_j being the *j*th row of **X** and **Z**, respectively. This equation is not different from that in Section 10.1.1 but it shows that the distribution of response probabilities by category is a function of the distance between a_j and the threshold. Similarly, the height of the normal curve at t_k (equation [10.1]) under the conditions of the *j*th row becomes:

$$\phi_{jk} = \phi(t_k - a_j) \qquad [10.6]$$

The formulae for computing the various matrices and vectors in [10.4] are outlined below.

The *j*th element of vector **v** can be calculated as:

$$\mathbf{v}_j = \sum_{k=1}^{m} n_{jk} \left(\frac{\phi_{j(k-1)} - \phi_{jk}}{p_{jk}} \right) \qquad [10.7]$$

The elements of the matrix **W**, which is a weighting factor, are computed as:

$$w_{jj} = n_{j.} \sum_{k=1}^{m} \frac{(\phi_{j(k-1)} - \phi_{jk})^2}{p_{jk}} \quad [10.8]$$

The matrix **Q** is an $(m-1)$ by $(m-1)$ banded matrix and the diagonal elements are calculated as:

$$q_{kk} = \sum_{j=1}^{s} n_{j.} \frac{P_{jk} - P_{j(k+1)}}{P_{jk} P_{j(k+1)}} \phi_{jk}^2, \quad \text{for } k = 1 \text{ to } (m-1) \quad [10.9]$$

and the off-diagonal elements are:

$$q_{(k+1)k} = -\sum_{j=1}^{s} n_{j.} \frac{\phi_{j(k+1)} \phi_{jk}}{P_{j(k+1)}}, \quad \text{for } k = 1 \text{ to } (m-2) \quad [10.10]$$

with the element $q_{k(k+1)} = q_{(k+1)k}$.

The matrix **L** is of order s by $(m-1)$ and its jkth element is calculated as:

$$l_{jk} = -n_{j.} \phi_{jk} \left(\frac{\phi_{jk} - \phi_{j(k-1)}}{P_{jk}} - \frac{\phi_{j(k+1)} - \phi_{jk}}{P_{j(k+1)}} \right) \quad [10.11]$$

The vector **p** is accumulated over all subclasses and its elements are:

$$P_k = \left\{ \sum_{j=1}^{s} \left[\frac{n_{jk}}{P_{jk}} - \frac{n_{j(k+1)}}{P_{j(k+1)}} \right] \phi_{jk} \right\}; \quad k = 1, m-1 \quad [10.12]$$

The remaining matrices in equation [10.4] can be computed by matrix multiplication.

10.1.3 Numerical example

Example 10.1

The analysis of categorical traits is illustrated below, using the calving ease data described by Gianola and Foulley (1983), but with a relationship matrix included for the sires and the age of dam effect omitted from the model. The data consisted of calving ease scores from 28 male and female calves born in 2 herd–years from cows mated to four sires. Cows were scored for calving ease using three ordered categories: 1 = normal birth, 2 = slight difficulty and 3 = extreme difficulty. The data set is presented in Table 10.2.

The following pedigree was assumed for the four sires:

Animal	Sire	Dam
1	0	0
2	0	0
3	1	0
4	3	0

Table 10.2. Distribution of calving ease score by herd–year and sex of calf subclasses.

Herd	Sex of calf	Sire of calf	Category of response[a] 1	2	3	Total
1	M	1	1	0	0	1
1	F	1	1	0	0	1
1	M	1	1	0	0	1
1	F	2	0	1	0	1
1	M	2	1	0	1	2
1	F	2	3	0	0	3
1	M	3	1	1	0	2
1	F	3	0	1	0	1
1	M	3	1	0	0	1
2	F	1	2	0	0	2
2	M	1	1	0	0	1
2	M	1	0	0	1	1
2	F	2	1	0	1	2
2	M	2	1	0	0	1
2	F	3	0	1	0	1
2	M	3	0	0	1	1
2	M	4	0	1	0	1
2	F	4	1	0	0	1
2	F	4	2	0	0	2
2	M	4	2	0	0	2

[a] 1= normal birth, 2 = slight difficulty, 3 = extreme difficulty.

The sire variance used in the analysis was assumed to be $\frac{1}{19}$. In the underlying scale, residual variance equals one; therefore $\sigma_e^2/\sigma_s^2 = 4 - h^2/h^2 = 19$. Thus the σ_s^2 assumed corresponded to a heritability of 0.20 on the underlying scale.

The vectors of solutions in [10.4] for the example data are:

$\mathbf{t}' = (t_1\ t_2)$, since there are two thresholds
$\mathbf{b}' = (h_1\ h_2\ \eta_1\ \eta_2)$
$\mathbf{u}' = (u_1\ u_2\ u_3\ u_4)$

where h_i and η_i represent solutions for level i of the herd–year and the sex of calf effects, respectively, and $\mathbf{u}$ is the vector of solutions for sires.

The inverse of the relationship for the assumed pedigree is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.3333 & 0.0000 & -0.6667 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ -0.6667 & 0.0000 & 1.6667 & -0.6667 \\ 0.0000 & 0.0000 & -0.6667 & 1.3333 \end{bmatrix}$$

For the example data, the transpose of matrix **X,** which relates subclasses to herd–year and sex of calf effects, and that of matrix **Z**, which

relates subclasses to sires, are:

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

and:

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Starting values for **t**, **b** and **u** are needed to commence the iterative process. Let $\mathbf{b} = \mathbf{u} = 0$, but starting values for t_i can be computed from the proportion of records in all categories of response preceding t_i. In Example 10.1, there is only one category before t_1 and 0.679 of the records are in this category. The first two categories precede t_2 and 0.857 of the records are observed in the two categories. Using these proportions, the values of **t** can be obtained from the usual table of standardized normal deviates of the normal distribution. From these proportions, $t_1 = 0.468$ and $t_2 = 1.080$ and these were used as starting values. However, using various starting values of **t**, Gianola and Foulley (1983) demonstrated that the system of equations converged rapidly. It seems, therefore, that the system of equations is not very sensitive to starting values for **t**. The calculations of the various matrices in the equations have been illustrated below using solutions obtained after the first iteration. The solutions obtained at the end of the first iteration and the updated estimates for the effects (which are now the starting values for the second iteration) are:

Solutions at the end of iteration one	Updated[a] estimates after iteration one
$\Delta t_1 = -0.026992$	$t_1 = 0.441008$
$\Delta t_2 = -0.035208$	$t_2 = 1.044792$
$\Delta \hat{h}_1 = 0.000000$	$\hat{h}_1 = 0.000000$
$\Delta \hat{h}_2 = 0.286869$	$\hat{h}_2 = 0.286869$
$\Delta \eta_1 = 0.000000$	$\eta_1 = 0.000000$
$\Delta \eta_2 = -0.358323$	$\eta_2 = -0.358323$
$\Delta u_1 = -0.041528$	$u_1 = -0.041528$
$\Delta u_2 = 0.057853$	$u_2 = 0.057853$
$\Delta u_3 = 0.039850$	$u_3 = 0.039850$
$\Delta u_4 = -0.065178$	$u_4 = -0.065178$

[a] The updated estimates were obtained as the sum of the starting values and the solutions at the end of the first iteration.

The following steps are involved in calculating P_{jk}, which is required to calculate subsequent matrices in the set of equations shown in [10.4] using the example data. In each round of iteration and for each subclass, that is, for $j = 1, \ldots s$:

1. Initially calculate $(t_k - a_j)$ in equation [10.5] for $k = 1, \ldots m - 1$. Therefore:

$$d_{jk} = (t_k - a_j) = t_k - x_j - z_j \quad \text{for } k = 1, \ldots m - 1$$

where x_j and z_j are the jth rows of **X** and **Z**.

For the example data in the second iteration:

$$d_{11} = t_1 - \hat{h}_1 - \hat{\eta}_1 - \hat{u}_1$$

$$d_{11} = 0.441008 - 0 - 0 - (-0.041528) = 0.482536$$

$$d_{12} = t_2 - \hat{h}_1 - \hat{\eta}_1 - \hat{u}_1$$

$$d_{12} = 1.044792 - 0 - 0 - (-0.041528) = 1.086320$$

$$d_{21} = t_1 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1$$

$$d_{21} = 0.441008 - 0 - (-0.358323) - (-0.041528) = 0.840859$$

$$d_{22} = t_2 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1$$

$$d_{22} = 1.044792 - 0 - (-0.358323) - (-0.041528) = 1.444643$$

$$\vdots$$

$$d_{201} = t_1 - \hat{h}_2 - \hat{\eta}_1 - \hat{u}_4$$

$$d_{201} = 0.441008 - 0.286869 - 0 - (-0.065178) = 0.219317$$

$$d_{202} = t_1 - \hat{h}_2 - \hat{\eta}_1 - \hat{u}_4$$

$$d_{202} = 1.044792 - 0.286869 - 0 - (-0.065178) = 0.823101$$

2. Using the values of d_{jk} computed above, calculate ϕ_{jk} (see equation [10.6]) and Φ_{jk}, for $k = 0, \ldots, m$. Note that, in all cases, when $k = 0$, $\phi_{jk} = \Phi_{jk} = 0$ and, when $k = m$, $\phi_{jk} = 0$ and $\Phi_{jk} = 1$.

In the second round of iteration for the example data:

$$\phi_{11} = \phi(0.482536) = 0.355099 \quad \text{and} \quad \Phi_{11} = \Phi(0.482536) = 0.685288$$

$$\phi_{12} = \phi(1.086320) = 0.221135 \quad \text{and} \quad \Phi_{12} = \Phi(1.086320) = 0.861331$$

$$\phi_{21} = \phi(0.840859) = 0.280142 \quad \text{and} \quad \Phi_{21} = \Phi(0.840859) = 0.799787$$

$$\phi_{22} = \phi(1.444643) = 0.140516 \quad \text{and} \quad \Phi_{22} = \Phi(1.444643) = 0.925721$$

$$\vdots$$

$$\phi_{201} = \phi(0.219317) = 0.389462 \quad \text{and} \quad \Phi_{201} = \Phi(0.219317) = 0.586799$$

$$\phi_{202} = \phi(0.823101) = 0.284311 \quad \text{and} \quad \Phi_{202} = \Phi(0.823101) = 0.794775$$

3. Then calculate P_{jk} as $\Phi_{jk} - \Phi_{j(k-1)}$ for $k = 1, \ldots m$.
In the second round of iteration, for Example 10.1:

$$P_{11} = \Phi_{11} - \Phi_{10} = 0.685288 - 0 = 0.685288$$

$$P_{12} = \Phi_{12} - \Phi_{11} = 0.861331 - 0.685288 = 0.176044$$

$$P_{13} = \Phi_{13} - \Phi_{13} = 1.0 - 0.861331 = 0.138669$$

$$P_{21} = \Phi_{21} - \Phi_{20} = 0.799787 - 0 = 0.799787$$

$$P_{22} = \Phi_{22} - \Phi_{21} = 0.925721 - 0.799787 = 0.125934$$

$$P_{23} = \Phi_{23} - \Phi_{22} = 1.0 - 0.925721 = 0.074279$$

$$\vdots$$

$$P_{201} = \Phi_{201} - \Phi_{200} = 0.586799 - 0 = 0.586799$$

$$P_{202} = \Phi_{202} - \Phi_{201} = 0.794775 - 0.586799 = 0.207976$$

$$P_{203} = \Phi_{203} - \Phi_{202} = 1.0 - 0.794775 = 0.205225$$

The calculation of the remaining matrices in the mixed model equations (MME) can now be illustrated for the example data. The first elements of **W** using equation [10.8] for the example data are:

$$w_{11} = 1\left[\frac{(0 - 0.355099)^2}{0.685288} + \frac{(0.355099 - 0.221135)^2}{0.176044} + \frac{(0.221135 - 0)^2}{0.138669}\right]$$
$$= 0.638589$$

and:

$$\mathbf{W} = \text{diag}[0.638589 \;\; 0.518748 \;\; 0.638589 \;\; 0.554385 \;\; 1.332860$$
$$1.663156 \;\; 1.323206 \;\; 0.548036 \;\; 0.661603 \;\; 1.233768$$
$$0.710404 \;\; 0.710404 \;\; 1.293402 \;\; 0.728641 \;\; 0.641496$$
$$0.725614 \;\; 0.705526 \;\; 0.609417 \;\; 1.218834 \;\; 1.411052]$$

For the vector **v**, the first element can be calculated from equation [10.7] as:

$$v_1 = \frac{1(0 - 0.355099)}{0.685288} + \frac{0(0.355099 - 0.221135)}{0.176044} + \frac{0(0.221135 - 0)}{0.138669}$$
$$= -0.518175$$

and the transpose of **v** is:

$$\mathbf{v}' = [-0.518175 \;\; -0.350270 \;\; -0.518175 \;\; 1.012257 \;\; 0.943660$$
$$-1.179520 \;\; 0.120754 \;\; 1.029729 \;\; -0.561257 \;\; -0.963633$$
$$-0.677635 \;\; 1.366976 \;\; 1.039337 \;\; -0.737615 \;\; 0.751341$$
$$1.304294 \;\; 0.505592 \;\; -0.470090 \;\; -0.940181 \;\; -1.327414]$$

The matrix **L** is order 20 by 2 for the example data. The elements in the first row of **L** from equation [10.11] can be calculated as:

$$l_{11} = (-1)(0.355099)\left[\frac{(0.355099 - 0)}{0.685288} - \frac{(0.221135 - 0.355099)}{0.176044}\right] = -0.454223$$

$$l_{12} = (-1)(0.221135)\left[\frac{(0.221135 - 0.355099)}{0.176044} - \frac{(0 - 0.221135)}{0.138669}\right] = -0.184365$$

The matrix **L** has not been shown because it is too large but the elements of the last row, l_{201} and l_{202}, are –0.910795 and –0.500257, respectively.

The elements of **Q** calculated using equations [10.9] and [10.10] are:

$$q_{11} = \frac{1\,(0.355099)^2(0.685287 + 0.176044)}{(0.685286 + 0.176044)} + \cdots$$
$$\frac{1\,(0.280142)^2(0.799787 + 0.125934)}{(0.799787 + 0.125934)} + \cdots$$
$$\frac{2\,(0.389462)^2(0.586799 + 0.207976)}{(0.586799 + 0.207976)} = 25.072830$$

$$q_{12} = \frac{-[1\,(0.355099)(0.221135)}{0.176044} + \frac{1\,(0.280142)(0.140516)}{0.125934} + \cdots$$
$$\frac{2(0.389462)(0.284311)}{0.207976} = -12.566598$$

$$q_{22} = \frac{1\,(0.221135)^2(0.176044 + 0.138669)}{(0.176044 + 0.138669)} + \cdots$$
$$\frac{1\,(0.140516)^2(0.125934 + 0.074279)}{(0.125934 + 0.074279)} + \cdots$$
$$\frac{2\,(0.284311)^2(0.207976 + 0.205225)}{(0.207976 + 0.205225)} = 17.928093$$

Since **Q** is symmetric, $q_{21} = q_{12}$.

Lastly, the elements of **p** can be calculated using equation [10.12] as:

$$p_1 = 0.355099\left(\frac{1}{0.685287} - \frac{0}{0.176044}\right) + \cdots$$
$$0.280142\left(\frac{1}{0.799787} - \frac{0}{0.125934}\right) + \cdots$$
$$0.389462\left(\frac{2}{0.586799} - \frac{0}{0.207976}\right) = -0.288960$$

$$p_2 = 0.221135\left(\frac{0}{0.176044} - \frac{0}{0.138669}\right) + \cdots$$
$$0.140516\left(\frac{0}{0.125934} - \frac{0}{0.074279}\right) + \cdots$$
$$0.284311\left(\frac{0}{0.207976} - \frac{0}{0.205225}\right) = 0.458984$$

The matrices in equation [10.4] can now be obtained by matrix multiplication and $\mathbf{A}^{-1}$ is added to $\mathbf{Z'WZ}$. The matrix $\mathbf{Z'WZ} + \mathbf{A}^{-1}\mathbf{G}^{-1}$ is illustrated below:

$$\mathbf{Z'WZ} + \mathbf{A}^{-1}\mathbf{G}^{-1} = \begin{bmatrix} 29.783773 & 0.000000 & -12.666731 & 0.000000 \\ 0.000000 & 24.572445 & 0.000000 & 0.000000 \\ -12.666731 & 0.000000 & 35.566685 & -12.666731 \\ 0.000000 & 0.000000 & -12.666731 & 29.278162 \end{bmatrix}$$

Then equation [10.4] is:

$$\begin{bmatrix} 25.073 & -12.567 & -5.733 & -6.733 & -6.366 & -6.140 & -3.123 & -3.977 & -2.699 & -2.707 \\ -12.567 & 17.928 & -2.146 & -3.215 & -3.220 & -2.141 & -1.327 & -1.595 & -1.201 & -1.238 \\ -5.733 & -2.146 & 7.879 & 0.000 & 4.595 & 3.284 & 1.796 & 3.550 & 2.533 & 0.000 \\ -6.773 & -3.215 & 0.000 & 9.989 & 4.992 & 4.997 & 2.655 & 2.022 & 1.367 & 3.945 \\ -6.366 & -3.220 & 4.595 & 4.992 & 9.586 & 0.000 & 2.698 & 2.062 & 2.710 & 2.117 \\ -6.140 & -2.141 & 3.284 & 4.997 & 0.000 & 8.281 & 1.753 & 3.511 & 1.190 & 1.828 \\ -3.123 & -1.327 & 1.796 & 2.655 & 2.698 & 1.753 & 29.784 & 0.000 & -12.667 & 0.000 \\ -3.977 & -1.595 & 3.550 & 2.022 & 2.062 & 3.511 & 0.000 & 24.572 & 0.000 & 0.000 \\ -2.699 & -1.201 & 2.533 & 1.367 & 2.710 & 1.190 & -12.667 & 0.000 & 35.567 & -12.667 \\ -2.707 & -1.238 & 0.000 & 3.945 & 2.117 & 1.828 & 0.000 & 0.000 & -12.667 & 29.278 \end{bmatrix}$$

$$\begin{bmatrix} \Delta\hat{t}_1 \\ \Delta\hat{t}_2 \\ \Delta\hat{h}_1 \\ \Delta\hat{h}_2 \\ \Delta\hat{\eta}_1 \\ \Delta\hat{\eta}_2 \\ \Delta\hat{u}_1 \\ \Delta\hat{u}_2 \\ \Delta\hat{u}_3 \\ \Delta\hat{u}_4 \end{bmatrix} = \begin{bmatrix} -0.289 \\ 0.459 \\ -0.021 \\ -0.149 \\ -0.099 \\ -0.071 \\ -0.104 \\ -0.021 \\ 0.031 \\ -0.076 \end{bmatrix}$$

The equations were solved with the solutions for $\Delta\hat{h}_1$ and $\Delta\hat{\eta}_1$ set to zero. The equations converged rapidly, and solutions at various different

iteration numbers and the final solutions are given below. Solutions from an analysis using a linear model with an α value of 19 are also shown:

	Iteration number				Solution from
Effects	1	2	3	7	linear models
Threshold					
1	0.4410	0.4375	0.4378	0.4378 ± 0.44[a]	–
2	1.0448	1.0661	1.0675	1.0675 ± 0.47	–
Herd–year					
1	0.0000	0.0000	0.0000	0.0000 ± 0.00	0.0
2	0.2869	0.2763	0.2774	0.2774 ± 0.49	1.0604
Sex of calf					
M	0.0000	0.0000	0.0000	0.0000 ± 0.00	0.0
F	–0.3583	–0.3577	–0.3589	–0.3590 ± 0.48	0.5193
Sire					
1	–0.0415	–0.0431	–0.0434	–0.0434 ± 0.22	0.2229
2	0.0579	0.0586	0.0592	0.0592 ± 0.21	0.2751
3	0.0399	0.0410	0.0412	0.0412 ± 0.22	0.3162
4	–0.0652	–0.0653	–0.0660	–0.0660 ± 0.22	0.0985

[a] Standard errors.

The standard errors associated with the results from the last iteration were computed from the square root of the diagonals of the generalized inverse. Sire rankings from the linear model were similar to those from the threshold model except for sires 2 and 3, which ranked differently.

Usually of interest is calculating the probability of response in a given category under specific conditions. For instance, the proportion of calving in the *j*th category of response, considering only female calves in herd–year–season 1 for sire 1, can be estimated as:

$$P_{11} = \Phi(t_1 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1) = \Phi(0.4378 - 0 - (-0.3590) - (-0.0434))$$
$$= \Phi(0.8402) = 0.800$$

$$P_{12} = \Phi(t_2 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1) - \Phi(t_1 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1)$$
$$= \Phi(1.0675 - 0 - (-0.3590) - (-0.0434)) - \Phi(0.800)$$
$$= \Phi(1.4699) - \Phi(0.800) = 0.129$$

$$P_{13} = 1 - \Phi(t_2 - \hat{h}_1 - \hat{\eta}_2 - \hat{u}_1) = 1 - \Phi(1.4699) = 0.071$$

Calculating this probability distribution by category of response for all sires gives the following:

	Probability in category of response		
	1	2	3
Sire 1	0.800	0.129	0.071
Sire 2	0.770	0.145	0.086
Sire 3	0.775	0.142	0.083
Sire 4	0.803	0.129	0.068

The results indicate that the majority of heifers calving in herd–year–season subclass 1 for all four sires were normal, with a very low proportion of extreme difficulties.

Since sires are used across herds, the interest might be the probability distribution of heifer calvings for each sire across all herds and sexes. Such a probability for each sire in category 1 of response per herd–year–sex subclass (Z_{1kji}) can be calculated as follows:

$$Z_{1kji} = \Phi(t_1 - (\hat{h}_k + \eta_j + \hat{u}_i)); \quad k = 1,2;\ i = 1, \ldots, 4$$

Since there are four herd–year–sex subclasses, the probability for sire i in category 1 (S_{1i}), can be obtained by weighting Z_{1kji} by factors that sum up to one. Thus:

$$S_{1i} = \sum_{i=1}^{4} \sum_{k=1}^{2} \sum_{m=1}^{2} a_{km} Z_{1ikm}$$

where $a_{km} = a_{11} + a_{12} + a_{21} + a_{22} = 1$. In the example data, $a_{11} = a_{12} = a_{21} = a_{22} = 0.25$.

Similarly, the probability for each sire in category 2 of response per herd–year–sex subclass (Z_{2kji}) can be calculated as:

$$Z_{2kji} = Z_{2kji^*} - Z_{1kji}$$

where:

$$Z_{1kji^*} = \Phi(t_2 - (\hat{h}_k + \eta_j + \hat{u}_i)); \quad k = 1,2;\ j = 1,2;\ i = 1, \ldots, 4$$

Finally, the probability for each sire in category 3 of response per herd–year–sex subclass (Z_{3kji}) can be calculated as:

$$Z_{3kji} = 1 - Z_{2kji^*}$$

For Example 10.1, the probability distribution of heifer calvings for each sire across all herds and sexes in all categories is as follows:

	Probability in category of response		
	1	2	3
Sire 1	0.695	0.175	0.131
Sire 2	0.659	0.188	0.153
Sire 3	0.665	0.186	0.149
Sire 4	0.702	0.172	0.126

10.2 Joint Analysis of Quantitative and Binary Traits

Genetic improvement may be based on selecting animals on an index that combines both quantitative and categorical traits. Optimally, a joint analysis of the quantitative and categorical traits is required in the prediction of breeding values in such a selection scheme to adequately account for selection. A linear multivariate model might be used for such analysis. However, such an analysis suffers from the limitations associated with the use of a linear model for the analysis of discrete traits mentioned in Section 10.1. In addition, such a multivariate linear model will not properly account for the correlated effects of the quantitative traits on the discrete trait.

Foulley *et al.* (1983) presented a method of analysis to handle the joint analysis of quantitative and binary traits using a Bayesian approach. It involves fitting a linear model for the quantitative traits and a non-linear model for the binary trait. This section presents this methodology and illustrates its application to an example data set.

10.2.1 Data and model definition

Assume that a quantitative trait, such as birth weight, and a binary trait, such as calving difficulty (easy versus difficult calving), are being analysed. As in Section 10.1.2, the data for calving difficulty could be represented in an s by 2 contingency table:

	Response category	
Row	Easy calving	Difficult calving
1	n_{11}	$n_{1.} - n_{11}$
2	n_{21}	$n_{2.} - n_{21}$
⋮	⋮	⋮
j	n_{j1}	$n_{j.} - n_{j1}$
⋮	⋮	⋮
s	n_{s1}	$n_{s.} - n_{s1}$

where the s rows refer to conditions affecting an individual or grouped records. Note that n_{i1} or $n_{i.} - n_{i1}$ in the above table can be null, as responses in the two categories are mutually exclusive, but $n_{i.} \neq 0$.

Assume that a normal function has been used to describe the probability of response for calving ease. Let $\mathbf{y}_1$ be the vector for observations for the quantitative trait, such as birth weight, and $\mathbf{y}_2$ be the vector of the underlying variable for calving difficulty. The model for trait 1 would be:

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{e}_1 \qquad [10.13]$$

and for the underlying variable for trait 2:

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}_2 \qquad [10.14]$$

where $\boldsymbol{\beta}_1$ and $\mathbf{u}_1$ are vectors of fixed effect and sire solutions for trait 1 and $\mathbf{X}_1$ and $\mathbf{Z}_1$ are the usual incidence matrices. The matrices $\mathbf{X}_2$ and $\mathbf{Z}_2$ are incidence matrices for the liability. The matrix $\mathbf{Z}_2 = \mathbf{Z}_1$ and $\mathbf{X}_2 = \mathbf{X}_1\mathbf{H}$, where $\mathbf{H}$ is an identity matrix if all factors affecting the quantitative traits also affect the liability. However, if certain fixed effects affecting the quantitative trait have no effect on the liability, $\mathbf{H}$ is obtained by deleting the columns of an identity matrix of appropriate order corresponding to such effects. It is assumed that:

$$\text{var}\begin{pmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{pmatrix} = \begin{pmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{pmatrix}$$

$$\text{var}\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{pmatrix} = \mathbf{A} \otimes \mathbf{G} \qquad [10.15]$$

where $\mathbf{G}$ is the genetic covariance matrix for both traits and $\mathbf{A}$ is the numerator relationship matrix.

Let $\boldsymbol{\theta}' = [\boldsymbol{\beta}_1, \boldsymbol{\tau}, \mathbf{u}_1, \mathbf{v}]$, the vector of location parameters in [10.13] and [10.14] to be estimated, where $\boldsymbol{\tau} = \boldsymbol{\beta}_2 - b\mathbf{H}\boldsymbol{\beta}_1$ and $\mathbf{v} = \mathbf{u}_2 - b\mathbf{u}_1$, where b is the residual regression coefficient of the underlying variate on the quantitative trait. The calculation of b is illustrated in the subsequent section. Since the residual variance of liability is unity, the use of b is necessary to properly adjust the underlying variate for the effect of the residual covariance between both traits. The use of b can be thought as correcting calving difficulty for other 'risk' factors affecting calving and, in this example, the birth weight of the calf. Thus [10.15] may be written as:

$$\text{var}\begin{pmatrix}\mathbf{u}_1\\ \mathbf{u}_2 - b\mathbf{u}_1\end{pmatrix} = \begin{pmatrix}\mathbf{u}_1\\ \mathbf{v}\end{pmatrix} = \mathbf{A} \otimes \mathbf{G}_c$$

where:

$$\mathbf{G}_c = \begin{pmatrix}\mathbf{I} & \mathbf{0}\\ -b\mathbf{I} & \mathbf{I}\end{pmatrix}\begin{pmatrix}g_{11} & gg_{12}\\ g_{21} & gg_{22}\end{pmatrix}\begin{pmatrix}\mathbf{I} & -b\mathbf{I}\\ \mathbf{0} & \mathbf{I}\end{pmatrix} \qquad [10.16]$$

with g_{ij} being the elements of $\mathbf{G}$.

Using a Bayesian approach, Foulley *et al.* (1983) calculated the mode of the posterior density of $\boldsymbol{\theta}$ by equating the derivatives of the log-posterior density of $\boldsymbol{\theta}$ to zero. The resulting systems of equations were not linear in $\boldsymbol{\theta}$. They set up the following iterative system of equations for $\boldsymbol{\theta}$ to be estimated:

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}_1^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{X}_1 & \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{Z}_1 + \mathbf{A}^{-1}\mathbf{g}_c^{11} & \mathbf{0} & \mathbf{A}^{-1}\mathbf{g}_c^{12} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_2'\mathbf{W}^{[i-1]}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{W}^{[i-1]}\mathbf{Z}_2 \\ \mathbf{0} & \mathbf{A}^{-1}\mathbf{g}_c^{21} & \mathbf{Z}_2'\mathbf{W}^{[i-1]}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{W}^{[i-1]}\mathbf{Z}_2 + \mathbf{A}^{-1}\mathbf{g}_c^{22} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}^{[i]} \\ \hat{\mathbf{u}}^{[i]} \\ \boldsymbol{\Delta\tau}^{[i]} \\ \boldsymbol{\Delta v}^{[i]} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{X}_1'\mathbf{R}_1^{-1}\mathbf{y}_1 \\ \mathbf{Z}_1'\mathbf{R}_1^{-1}\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{q}^{[i-1]} \\ \mathbf{Z}_2'\mathbf{q}^{[i-1]} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{A}^{-1}\mathbf{g}_c^{12}\mathbf{v}^{[i-1]} \\ \mathbf{0} \\ \mathbf{A}^{-1}\mathbf{g}_c^{22}\mathbf{v}^{[i-1]} \end{pmatrix} \qquad [10.17]$$

The matrices and vectors in [10.17] have been defined earlier, apart from $\mathbf{q}$ and $\mathbf{W}$.

Initially, P_{jk}, the probability of response in the category k, given the conditions in the jth row, is defined for the category trait. With only two categories of response for calving difficulty, then from [10.5]:

$$P_{j1} = \Phi(t - a_j) \quad \text{and} \quad P_{j2} = 1 - P_{j1}$$

with a_j regarded as the mean of the liability in the jth row or as defined in [10.5].

However with only one threshold, the value of t by itself is of no interest; the probability of response in the first category for the jth row can then be written as:

$$P_{j1} = \Phi(t - a_j) = \Phi(\mu_j)$$

where μ_j can be defined as the expectation of $\mathbf{y}_{2j}$ given $\boldsymbol{\beta}$, $\mathbf{u}$ and $\mathbf{y}_{1j}$, and this is worked out in the next section.

The vector $\mathbf{q}$ is of order s by 1 with elements:

$$\mathbf{q}_j = -\{n_{j1}d_{j1} + (n_{j.} - n_{j1})d_{j2}\}, \quad j = 1, \ldots, s \qquad [10.18]$$

where $d_{j1} = -\phi(\mu_j)/P_{j1}$ and $d_{j2} = \phi\mu_j/(1 - P_{j1})$, with P_{j1} calculated as $\Phi(\mu_j)$.

$\mathbf{W}$ is an s by s diagonal matrix with the following elements:

$$w_{jj} = \mu_j q_j + n_{j1}d_{j1}^2 + (n_{j.} - n_{j1})d_{j2}^2, \quad j = 1, \ldots, s \qquad [10.19]$$

Calculating $\boldsymbol{\mu}$ *and the residual regression coefficient*

From [10.14], the model for the jth row of the contingency table may be written as:

$$\mathbf{y}_{2j} = \mathbf{x}_{2j}'\boldsymbol{\beta}_2 + \mathbf{z}_{2j}'\mathbf{u}_2 + \mathbf{e}_{2j}$$

where $\mathbf{x}_{2j}'$ and $\mathbf{z}_{2j}'$ are the jth row vectors of $\mathbf{X}_2$ and $\mathbf{Z}_2$, respectively. Similarly, observations for trait 1, corresponding to the jth row of

the contingency table, may be modelled as:

$$\mathbf{y}_{1j} = \mathbf{x}'_{1j}\boldsymbol{\beta}_1 + \mathbf{z}'_{1j}\mathbf{u}_1 + \mathbf{e}_{1j}$$

Let μ_j be the expectation of $\mathbf{y}_{2j}$ given $\boldsymbol{\beta}$, $\mathbf{u}$ and $\mathbf{y}_{1j}$. Thus:

$$\mu_j = \mathrm{E}(\mathbf{y}_{2j}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{y}_{1j}) = \mathbf{x}'_{2j}\boldsymbol{\beta}_2 + \mathbf{z}'_{2j}\mathbf{u}_2 + \mathrm{E}(e_{2j}|e_{1j}) \qquad [10.20]$$

given that e_{2j} is only correlated with e_{1j}. Assuming e_{2j} and e_{1j} are bivariately normally distributed:

$$\begin{aligned}\mathrm{E}(e_{2j}|e_{1j}) &= \frac{\sigma_{e(2,1)}}{\sigma^2_{e1}}(e_{1j}) \\ &= r_{12}\left(\frac{\sigma_{e2}}{\sigma_{e1}}\right)e_{1ij}\end{aligned} \qquad [10.21]$$

where σ^2_{ei} is the residual variance of trait i, $\sigma_{ei,k}$, and r_{ik} are the residual covariance and correlation between traits i and k, and σ_{ei} is the residual standard deviation of the ith trait. Similarly:

$$\begin{aligned}\mathrm{var}(\mathrm{y}_{2j}|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{y}_{1j}) &= \mathrm{var}(e_{2j}|e_{1j}) \\ &= \sigma^2_{e2}(1 - r^2_{12})\end{aligned}$$

Since the unit of the conditional distribution of the underlying trait, given $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u}_1, \mathbf{u}_2$, and y_{1j}, is the standard deviation, then from the above equation, $\sigma_{e2} = 1/\sqrt{(1 - r^2_{12})}$. Therefore [10.21] can be written as:

$$\begin{aligned}\mathrm{E}(e_{2j}|e_{1j}) &= r_{12}\left(\frac{1}{\sigma_{e1}}\right)\frac{1}{\sqrt{1 - r^2_{12}}}e_{1j} \\ &= be_{1j}\end{aligned} \qquad [10.22]$$

In general [10.20] can be expressed as:

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + b\mathbf{e}_1 \\ &= \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + b(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \mathbf{Z}_1\mathbf{u}_1)\end{aligned} \qquad [10.23]$$

The above equation may be written as:

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{X}_2(\boldsymbol{\beta}_2 - b\mathbf{H}\boldsymbol{\beta}_1) + \mathbf{Z}_2(\mathbf{u}_2 - b\mathbf{u}_1) + b\mathbf{y}^*_1 \\ \boldsymbol{\mu} &= \mathbf{X}_2\boldsymbol{\tau} + \mathbf{Z}_2\mathbf{v} + b\mathbf{y}^*_1\end{aligned} \qquad [10.24]$$

with the solutions of factors affecting calving difficulty corrected for the residual relationship between the two traits and $\mathbf{y}^*_1 = (\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 - \mathbf{Z}_1\mathbf{u}_1)$ or $\mathbf{y}^*_1$ may be calculated as $\mathbf{y}^*_1 = (\mathbf{y}_1 - \bar{y}_1)$, where $\bar{y}_1$ is the mean of $\mathbf{y}_1$.

10.2.2 Numerical application

Example 10.2

The bivariate analysis of a quantitative trait and a binary trait is illustrated using the data presented by Foulley *et al.* (1983), but with a sire–maternal grandsire relationship matrix included for the sires and pelvic opening omitted from the analysis. The data consisted of birth weight (BW) and calving difficulty (CD) on 47 Blonde d'Aquitaine heifers, with information on region of origin, sire of the heifer, calving season and sex of the calf included. Calving difficulty was summarized into two categories: easy or difficult calving. The data set is presented below:

Heifer origin	Sire	Season	Sex of calf[a]	BW	CD[b]	Heifer origin	Sire	Season	Sex of calf[a]	BW	CD[b]
1	1	1	1	41.0	E	1	4	2	1	47.0	D
1	1	1	1	37.5	E	1	4	2	2	51.0	D
1	1	1	2	41.5	E	1	4	2	2	39.0	E
1	1	2	2	40.0	E	2	4	1	1	44.5	E
1	1	2	2	43.0	E	1	5	1	1	40.5	E
1	1	2	2	42.0	E	1	5	1	2	43.5	E
1	1	2	2	35.0	E	1	5	2	1	42.5	E
2	1	1	2	46.0	E	1	5	2	1	48.8	D
2	1	1	2	40.5	E	1	5	2	1	38.5	E
2	1	2	2	39.0	E	1	5	2	1	52.0	E
1	2	1	1	41.4	E	1	5	2	2	48.0	E
1	2	1	1	43.0	D	2	5	1	2	41.0	E
1	2	2	2	34.0	E	2	5	1	1	50.5	D
1	2	2	1	47.0	D	2	5	2	1	43.7	D
1	2	2	1	42.0	E	2	5	2	1	51.0	D
2	2	2	1	44.5	E	1	6	1	2	51.6	D
2	2	2	1	49.0	E	1	6	1	1	45.3	D
1	3	1	1	41.6	E	1	6	1	2	36.5	E
2	3	1	1	36.0	E	1	6	2	1	50.5	E
2	3	1	2	42.7	E	1	6	2	1	46.0	D
2	3	2	2	32.5	E	1	6	2	1	45.0	E
2	3	2	2	44.4	E	1	6	2	2	36.0	E
2	3	2	1	46.0	E	2	6	1	2	43.5	E
						2	6	1	2	36.5	E

[a] 1, male; 2, female.
[b] CD, calving difficulty; D, difficult, E, easy.

A summary of the data, in terms of marginal means of calving variables by level of factors considered, is shown in the following table:

Factor		Number	Birth weight	Frequency CD[a]
Heifer origin	1	30	43.02	0.267
	2	17	43.02	0.176
Calving season	1	20	42.23	0.200
	2	27	43.61	0.259
Sex of calf	M	25	–	0.360
	F	22	–	0.091
Sire of heifer	1	10	40.55	0.000
	2	7	42.99	0.286
	3	6	40.53	0.000
	4	4	45.38	0.500
	5	11	45.46	0.364
	6	9	43.43	0.333

[a] Frequency of calving difficulty.

The following sire–maternal grandsire relationship matrix was assumed among the sires:

Bull	Sire	Maternal grandsire
1	0	0
2	0	0
3	1	0
4	2	1
5	3	2
6	2	3

The inverse of the sire–maternal grandsire relationship matrix obtained for the above pedigree using the rules in Section 2.4 is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.424 & 0.182 & -0.667 & -0.364 & 0.000 & 0.000 \\ 0.182 & 1.818 & 0.364 & -0.727 & -0.364 & -0.727 \\ -0.667 & 0.364 & 1.788 & 0.000 & -0.727 & -0.364 \\ -0.364 & -0.727 & 0.000 & 1.455 & 0.000 & 0.000 \\ 0.000 & -0.364 & -0.727 & 0.000 & 1.455 & 0.000 \\ 0.000 & -0.727 & -0.364 & 0.000 & 0.000 & 1.455 \end{bmatrix}$$

The residual variance (σ^2_{e1}) for BW was assumed to be 20 kg^2 and the residual correlation (r_{12}) between BW and CD was assumed to be 0.459.

Therefore, from [10.20], b equals 0.1155. The matrix **G** assumed was:

$$\mathbf{G} = \begin{pmatrix} 0.7178 & 0.1131 \\ 0.1131 & 0.0466 \end{pmatrix}$$

Therefore, from [10.16]:

$$\mathbf{G}_c = \begin{pmatrix} 1 & 0 \\ -0.1155 & 1 \end{pmatrix}\begin{pmatrix} 0.7178 & 0.1131 \\ 0.1131 & 0.0466 \end{pmatrix}\begin{pmatrix} 1 & -0.1155 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.7178 & 0.0302 \\ 0.0302 & 0.0300 \end{pmatrix}$$

Thus the heritabilities for BW and CD are 0.14 and 0.18, respectively, with a genetic correlation of 0.62 between both traits.

The model [10.13] was used for the analysis of BW; thus $\boldsymbol{\beta}_1$ is the vector of solutions for origin of heifer, calving season and sex of calf and $\mathbf{u}_1$ is the vector of solution for sire effects. The same effects were fitted for CD, with $\boldsymbol{\tau}$ being the vector of solutions for the fixed effects and **v** for the sire effects. Let $\boldsymbol{\theta}$ be a vector with elements:

$$\boldsymbol{\beta}'_1 = (d_1, d_2, s_1, s_2, f_1, f_2)$$
$$\mathbf{u}'_1 = (\hat{u}_{11}, \hat{u}_{12}, \hat{u}_{13}, \hat{u}_{14}, \hat{u}_{15}, \hat{u}_{16})$$
$$\boldsymbol{\tau}' = (d'_1, d'_2, s'_1, s'_2, f'_1, f'_2)$$
$$\mathbf{v}' = (v_1, v_2, v_3, v_4, v_5, v_6)$$

where $d_i(d'_i)$, $s_i(s'_i)$ and $f_i(f'_i)$ are level i of the effects of heifer origin, calving season and sex of calf, respectively, for BW(CD), $\hat{u}_{1j}$ and v_j are the solutions for the sire j for BW and CD, respectively.

The matrix $\mathbf{X}_1$, which relates records for BW to the effects of heifer origin, calving season and sex of calf, can be set by principles already outlined in previous chapters. For the example data, all fixed effects affecting BW also affect CD; therefore **H** is an identity matrix and $\mathbf{X}_2 = \mathbf{X}_1$. Similarly, the matrix $\mathbf{Z}_1 = \mathbf{Z}_2$. The remaining matrix in [10.17] can be obtained through matrix multiplication and addition.

Equation [10.17] needs starting values for $\boldsymbol{\tau}$ and **v** to commence the iterative process. The starting values used were solutions ($\boldsymbol{\tau}^{(i)}$ and $\mathbf{v}^{(i)}$) from [10.17] with $\mathbf{W}^{[i-1]} = \mathbf{I}$, $\mathbf{q}^{[i-1]}$ = a vector of (0,1) variables (1, difficulty; 0, otherwise) and $\mathbf{v}^{[i-1]} = 0$. The solutions to [10.17] using these starting values are shown in Table 10.3, with equations for the second levels of calving season and sex of calf effects set to zero because of dependency in the systems of equations. Using these solutions, the calculations of $\mathbf{q}^{(i)}$ and $\mathbf{W}^{(i)}$, in the next round of iteration are illustrated for the first and last two animals in the example data.

First, $\boldsymbol{\mu}$ in [10.18] is calculated for these animals using [10.24].

For animals 1 and 2:

$$\mathbf{X}_2\boldsymbol{\tau} + \mathbf{Z}_2\mathbf{v} = (d'_1 + \hat{s}'_1 + f'_1 + \hat{v}_1) = 0.1873 + (-0.0874) + 0.2756 + (-0.1180)$$
$$= 0.2575$$

Therefore, from [10.24], using the mean of the birth weight, μ_1 is:

$$\mu_1 = 0.2575 + 0.1155(41 - 43.02) = 0.0242$$

and:

$$\mu_2 = 0.2575 + 0.1155(37.5 - 43.02) = -0.3800$$

For animals 46 and 47:

$$\mathbf{X}_2\boldsymbol{\tau} + \mathbf{Z}_2\mathbf{v} = (d'_2 + \hat{s}'_1 + f'_2 + \hat{v}_6) = 0.1484 + (-0.0874) + 0.0 + 0.0079$$
$$= 0.0690$$

Therefore, from [10.24]:

$$\mu_{46} = 0.0690 + 0.1155(43.5 - 43.02) = 0.1244$$

and:

$$\mu_{47} = 0.0690 + 0.1155(36.5 - 43.02) = -0.6841$$

Using [10.18], the elements of $\mathbf{q}$ for animals 1, 2, 46 and 47 are:

$$\mathbf{q}(1) = -\{0(-1)\phi(0.0242)/\Phi(0.0242) + (1-0)(\phi(0.0242)/(1-\Phi(0.0242))\}$$
$$= -\{0(-1)0.3988/0.5097 + 1(0.3988/0.4903)\} = -0.8134$$

$$\mathbf{q}(2) = -\{0(-1)\phi(-0.3800)/\Phi(-0.3800) + (1-0)(\phi(-0.3800)/(1-\Phi(-0.3800))\}$$
$$= -\{0(-1)0.3712/0.3520 + 1(0.3712/0.6480)\} = -0.5727$$

$$\mathbf{q}(46) = -\{0(-1)\phi(0.1244)/\Phi(0.1244) + (1-0)(\phi(0.1244)/(1-\Phi(0.1244))\}$$
$$= -\{0(-1)0.3959/0.5495 + 1(0.3959/0.4505)\} = -0.8787$$

$$\mathbf{q}(47) = -\{0(-1)\phi(-0.6841)/\Phi(-0.6841) + (1-0)(\phi(-0.6841)/(1-\Phi(-0.6841))\}$$
$$= -\{0(-1)0.3157/0.2470 + 1(0.3157/0.7530)\} = -0.4193$$

The diagonal elements of $\mathbf{W}$ for each of the four animals above can be calculated, using [10.19], as:

$$w(1,1) = 0.0242(-0.8134)\{0(-1)[\phi(0.0242)/\Phi(0.0242)]^2$$
$$+ (1-0)[(\phi(0.0242)/(1-\Phi(-0.0242))]^2\} = 0.6419$$

$$w(2,2) = -0.3800(-0.5727)\{0(-1)[\phi(-0.3800)/\Phi(-0.3800)]^2$$
$$+ (1-0)[(\phi(-0.3800)/(1-\Phi(-0.3800))]^2\} = 0.5458$$

$$w(46,46) = 0.1244(-0.8787)\{0(-1)[\phi(0.1244)/\Phi(0.1244)]^2$$
$$+ (1-0)[(\phi(0.1244)/(1-\Phi(0.1244))]^2\} = 0.6629$$

$$w(47,47) = -0.6841(-0.4193)\{0(-1)[\phi(-0.6841)/\Phi(-0.6841)]^2$$
$$+ (1-0)[(\phi(-0.6841)/(1-\Phi(-0.6841))]^2\} = 0.4626$$

The equations were solved iteratively and were said to have converged at the 13th round of iteration when $\Delta'\Delta/20 \leq 10^{-7}$, where $\Delta = \boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i-1)}$.

Solutions at convergence at the 13th round of iteration and at some intermediate rounds are shown in Table 10.3. Results from an analysis using a linear model fitting the same effects with the **G** matrix and residual variances of 20 kg^2 for BW, 1.036 for CD and residual covariance of 2.089 between the two traits are also presented.

The results indicate that the probability of a difficult calving is larger for a male calf than for a female calf. Similarly, there is a slightly higher probability for calving difficulty for calving in the first season.

Table 10.3. Solutions to Example 10.2 using equation [10.17].

		Iteration number					Linear
Trait[a]	Factor	0	1	4	8	13	model
BW	*Heifer origin*						
	1	41.6633	41.5471	41.6262	41.6182	41.6195	41.6175
	2	42.2530	42.1409	42.2178	42.2099	42.2112	42.2022
	Calving season						
	1	–1.2350	–1.2345	–1.2346	–1.2343	–1.2344	–1.2387
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sex of calf[b]						
	1	3.1589	3.1890	3.1687	3.1690	3.1690	3.1845
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sire						
	1	–0.4155	–0.2633	–0.3671	–0.3580	–0.3595	–0.3268
	2	0.1048	0.1687	0.1246	0.1311	0.1300	0.1171
	3	–0.3315	–0.2280	–0.3007	–0.2939	–0.2950	–0.2641
	4	0.1364	0.3365	0.2035	0.2139	0.2122	0.1886
	5	0.2730	0.3261	0.2893	0.2979	0.2965	0.2688
	6	0.1545	0.2270	0.1770	0.1821	0.1813	0.1690
CD	*Heifer origin*						
	1	0.1873	–1.0189	–1.4072	–1.3915	–1.3943	0.1349
	2	0.1484	–1.2813	–1.7342	–1.7472	–1.7452	0.0876
	Calving season						
	1	–0.0874	0.1871	0.1327	0.1415	0.1401	–0.0311
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sex of calf						
	1	0.2756	0.3218	0.8621	0.8369	0.8411	0.2410
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sire						
	1	–0.1180	0.0471	–0.0656	–0.0561	–0.0577	–0.0527
	2	0.0144	0.0705	0.0319	0.0379	0.0369	0.0285
	3	–0.0850	0.0185	–0.0546	–0.0477	–0.0488	–0.0427
	4	–0.0380	0.1698	0.0319	0.0424	0.0407	0.0350
	5	–0.0048	0.0362	0.0075	0.0163	0.0148	0.0195
	6	0.0079	0.0702	0.0270	0.0315	0.0308	0.0323

[a] BW, birth weight; CD, calving difficulty.

[b] 1, male; 2, female.

In general, sire rankings from the threshold and linear models were similar, except for sires 2 and 6 slightly changing rankings in the two models. The ranking of sires for calving difficulty based on the results from the threshold model could be based on $\hat{u}_2 = v + b_1\hat{u}_1$ using the information provided by BW. However, the interest might be on ranking sires in terms of probability of calving difficulty, under a given set of conditions. For instance, what is the probability that a heifer, sired by the *j*th bull born in region 2, calving a male calf in season 1 will experience a calving difficulty? This probability (V_{211j}) can be calculated as:

$$V_{211j} = \Phi[\hat{d}'_2 + \hat{s}'_1 + \hat{f}'_1 + \hat{v}_j + b_1(\hat{d}_2 + \hat{s}_1 + \hat{f}_1 - 43.02)] \qquad [10.25]$$

Using the above equation, this probability for sire 1 is:

$$V_{211j} = \Phi[-1.7452 + 0.1401 + 0.8411 + (-0.0577) + 0.1155(42.2112 + (-1.2344) + 3.1690 - 43.02)] = 0.245$$

Similar calculations gave probabilities of 0.275, 0.247, 0.276, 0.268 and 0.273 for sires 2, 3, 4, 5 and 6, respectively. In general, there might be interest in the probability of difficult calving associated with using the *j*th sire across all regions of origin by season of calving and sex of calf subclasses. Such a probability can be calculated as:

$$V_{\cdots j} = \sum_{ikl} \lambda_{ikl} V_{iklj} \qquad [10.26]$$

with V_{iklj} estimated as [10.25] and λ_{ikl} is an arbitrary weight such that $\Sigma_{ikl}\lambda_{ikl} = 1$. For the example data, λ can be set to be equal to $\frac{1}{8}$, as there are eight region–season–sex of calf subclasses. The probabilities obtained using [10.26] with $\lambda = \frac{1}{8}$ were 0.167, 0.188, 0.169, 0.189, 0.183 and 0.187 for sires 1, 2, 3, 4, 5 and 6, respectively.

The analysis of a binary trait with a quantitative trait has been discussed and illustrated in this section. However, if the category trait has several thresholds, then the method discussed in Section 10.1 would be used for the analysis of the categorical trait.

11 Estimation of Genetic Parameters

Robin Thompson

Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK and Roslin Institute, Roslin Biocentre, Midlothian EH25 9PS, UK

In order to carry out prediction of breeding values, estimates of variance components are normally needed. In this chapter the estimation of variance parameters is considered using univariate sire and animal models.

11.1 Univariate Sire Model

To motivate this work, the mixed effect sire model introduced in Chapter 3 is used.This model [3.15] has:

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

and:

$$\text{var}(\mathbf{s}) = \mathbf{A}\sigma_s^2$$

$$\text{var}(\mathbf{y}) = \mathbf{ZAZ}'\sigma_s^2 + \mathbf{R}$$

where $\mathbf{A}$ is the numerator relationship matrix for sires, $\sigma_s^2 = 0.25\sigma_a^2$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$. Interest is in estimating σ_s^2 and σ_e^2. The simple case with this sire model is with $\mathbf{X}$ being an $n \times 1$ matrix with elements 1, $\mathbf{b}$ having one element representing an overall effect and the q sires being unrelated, so that $\mathbf{A} = \mathbf{I}$.

An analysis of variance can be constructed by fitting: (i) a model with the overall effect $\mathbf{b}$; and (ii) a model with sire effects, these models giving residual sums of squares that can be put into an analysis of variance

of the following form:

Source	Degrees of freedom	Sums of squares
Overall	Rank (**X**) = 1	$\mathbf{y'X(X'X)^{-1}X'y} = F$
Sires	Rank (**Z**) – rank (**X**) = $q - 1$	$\mathbf{y'Z(Z'Z)^{-1}Z'y} - \mathbf{y'X(X'X)^{-1}X'y} = S$
Residual	n – rank (**Z**) = n – q	$\mathbf{y'y} - \mathbf{y'Z(Z'Z)^{-1}Z'y} = R$

Essentially, the effects **b** and **s** are thought of as fixed effects to construct an unweighted analysis. If estimates of σ_s^2 and σ_e^2 are required, then the sums of squares S and R can be equated to their expectation $E(R) = (n-q)\sigma_e^2$ and $E(S) = (q-1)\sigma_e^2 + \text{trace}(\mathbf{Z'SZ})\sigma_s^2$, where $\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$.

11.2 Numerical Example of Sire Model

Consider the data in Table 11.1 for the pre-weaning gain (WWG) of beef calves. The objective is to illustrate the estimation of variance components on a very small example so that the calculations can be expressed concisely.

The model to describe the observations is:

$$y_{ij} = o + s_j + e_i$$

where y_{ij} = the WWG of the ith calf, o = the overall effect, s_j = random effect of the jth sire ($j = 1, 2, 3$), and e_i = random error effect ($i = 1, 2, 3, 4$).

In matrix notation, the model is the same as described in equation [3.15], with $n = 4$, $p = 1$ and $q = 3$.

The matrix **X** in the mixed model equations (MME) relates records to the overall effects. For the example data set, its transpose is:

$$\mathbf{X'} = [1 \quad 1 \quad 1 \quad 1 \quad 1]$$

The matrix **Z** then relates records to sires. In this case, it is:

$$\mathbf{Z} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Table 11.1. Pre-weaning gain (kg) for four beef calves.

Calf	Sire	WWG (kg)
4	2	2.9
5	1	4.0
6	3	3.5
7	2	3.5

An analysis of variance can be constructed as:

Source	Degrees of freedom	Sums of squares (kg^2)
Overall	1	F = 48.3025
Sire	2	S = 0.6075
Residual	1	R = 0.1800

with:

$$\mathbf{y'X(X'X)^{-1}X'y} = \mathrm{F} = (2.9 + 4 + 3.5 + 3.5)^2/4 = 48.3025$$

$$\mathbf{y'Z(Z'Z)^{-1}Z'y} - \mathbf{y'X(X'X)^{-1}X'y} = \mathrm{S} = (4)^2/1 + (2.9 + 3.5)^2/2 + (3.5)^2/1 - 48.3085 = 0.6075$$

$$\mathbf{y'y} - \mathbf{y'Z(Z'Z)^{-1}Z'y} = \mathrm{R} = (2.9)^2 + (4)^2 + (3.5)^2 + (3.5)^2 - \mathrm{F} - \mathrm{S} = 0.18$$

In this case:

$$\mathbf{Z} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix} \quad \text{and} \quad \mathbf{Z'SZ} = \begin{bmatrix} 0.75 & -0.50 & -0.25 \\ -0.50 & 1.00 & -0.25 \\ -0.25 & -0.50 & 0.75 \end{bmatrix}$$

so that:

$$\mathrm{E(R)} = \sigma_e^2 = 0.18 \quad \text{and} \quad \mathrm{E(S)} = 2\sigma_e^2 + 2.5\sigma_s^2 = 0.6075$$

Then estimates of σ_e^2 and σ_s^2 are:

$$\sigma_e^2 = 0.18\ (\mathrm{kg}^2) \quad \text{and} \quad \sigma_s^2 = 0.027\ (\mathrm{kg}^2)$$

11.3 Extended Model

The model and analysis hold if the model is extended to allow **X** to represent an environmental effect with p levels. If sires are nested within levels of the environmental factor so that daughters of each sire are only

associated with one level of the environmental factor, then the above analysis could be used. If, however, as usually happens, daughters of a sire are associated with more than one level, a slightly more complicated analysis is appropriate:

Source	Degrees of freedom	Sums of squares
Fixed effects	Rank (**X**) = p	$\mathbf{y'X(X'X)^{-1}X'y}$ = F
Sires corrected for fixed effects	Rank (**Z′SZ**) = df_S	$\mathbf{y'SZ(Z'SZ)^{-1}Z'Sy}$ = S
Residual	n – rank (**X**) – rank (**Z′SZ**) = $n - p - df_S = df_R$	$\mathbf{y'y}$ – F – S = R

Now R and S have expectation:

$$\mathrm{E(R)} = df_R\,\sigma_e^2 \quad \text{and} \quad \mathrm{E(S)} = df_S\sigma_e^2 + \mathrm{trace}\,(\mathbf{Z'SZ})\sigma_s^2$$

The term involved in the trace (**Z′SZ**) can sometimes have a simple interpretation. If **X** represents a fixed effect matrix with p levels then the ith diagonal element of **Z′SZ** is $n_{i.} - \Sigma n_{ij}^2/n_{.j}$ (summation is from $j = 1$ to p), where n_{ij} is the number of daughters of sire i in fixed effect level j and $n_{i.} = \Sigma n_{ij}$ (summation is from $j = 1$ to p) and $n_{i.} = \Sigma n_{ij}$ (summation is from $i = 1$ to s). This number was called the effective number of daughters of sire i by Robertson and Rendel (1954) and measures the loss of information on a sire because his daughters are measured in different environmental classes. This method of analysis is called Henderson's method 3 (Henderson, 1953). These methods of analysis were very popular in that they related to sequential fitting of models and were relatively easy to compute. One problem was that the terms are generated under a fixed effect model with $\mathbf{V} = \mathbf{I}\sigma_e^2$ and then sums of squares are equated to their expectation under a different variance model. Only in special balanced cases will estimation based on R and S lead to efficient estimates of σ_s^2 and σ_e^2. In general, B is based on **Z′Sy** with variance matrix $\mathbf{Z'SZ}\sigma_e^2 + \mathbf{Z'SZAZ'SZ}\sigma_s^2$ and these can be transformed to df_S independent values **Q′Z′Sy** by using arguments similar to those used in Section 5.2 on the canonical transformation, where **Q** is a df_S n matrix and **Q′Z′SZQ = I**, and **Q′Z′SZAZ′SZQ = W**, where **W** is a diagonal matrix of size df_S with ith diagonal element w_i. The variance matrix of **Q′Z′Sy** is then $\mathbf{I}\sigma_e^2 + \mathbf{W}\sigma_s^2$. Then an analysis of variance can be constructed from squaring each of the df_S elements of **Q′Z′Sy** with the ith sum of squares u_i with expectation $\sigma_e^2 + w_i\sigma_s^2$ and R is the residual sum of squares, with expectation $\mathrm{E(R)} = df_R\,\sigma_e^2$. The individual u_i are distributed as chi-squared variables with variance $\mathrm{E}(u_i)^2$. A natural scheme is to fit a linear model in σ_s^2 and σ_e^2 to u_i and R. One can also use an iterative scheme with the weight dependent on the estimated parameters.

11.4 Numerical Example

For the example with data in Table 11.1, it was shown in Section 11.2 that:

$$\mathbf{Z'SZ} = \begin{bmatrix} 0.75 & -0.50 & -0.25 \\ -0.50 & 1.00 & -0.25 \\ -0.25 & -0.50 & 0.75 \end{bmatrix}$$

so that the sires have 0.75, 1.0 and 0.75 effective daughters, respectively.

It can be found that with $\mathbf{A} = \mathbf{I}$:

$$\mathbf{Z'SZAZ'SZ} = \begin{bmatrix} 0.875 & -0.750 & -0.125 \\ -0.750 & 1.500 & -0.750 \\ -0.125 & -0.750 & 0.875 \end{bmatrix}$$

The algorithm in Appendix E can be used to calculate the eigenvalues $\mathbf{Q}$ so that:

$$\mathbf{Q'Z'SZQ} = \mathbf{I} \quad \text{and} \quad \mathbf{Q'Z'SZAZ'SZQ} = \mathbf{W}$$

In this case:

$$\mathbf{Q} = \begin{bmatrix} -0.3333 & 0.6667 & -0.3333 \\ 0.7071 & 0.0000 & -0.7071 \end{bmatrix}$$

So $\mathbf{Q'Z'SZQ} = \mathbf{I}$ and $\mathbf{Q'Z'SZAZ'SZQ} = \mathbf{W}$ with:

$$\mathbf{W} = \begin{bmatrix} 1.5 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$$

The contrasts $\mathbf{Q'Z'Sy}$ are now:

$$\mathbf{Q'Z'S} = \begin{bmatrix} 0.5000 & -0.5000 & -0.5000 & 0.5000 \\ 0.0000 & 0.7071 & -0.7071 & 0.0000 \end{bmatrix}\mathbf{y}$$

The first contrast, $(y_1 - y_2 - y_3 + y_4)/2 = (2.9 - 4.0 - 3.5 + 3.5)/2 = -1.1/2 = -0.55$, is a scaled contrast comparing sire 2 with sire 1 and sire 3, and the second contrast, $(y_2 - y_3)/\sqrt{2} = (-4.0 + 3.5)/\sqrt{2} = (-0.5)/\sqrt{2}$, is a scaled contrast between sire 1 and sire 3.

An analysis of variance can be constructed:

Source	Degrees of freedom	Sums of squares (kg^2)	Expected mean squares (kg^2)
Overall	1	F = 48.3085	
Sire 2 compared with sires 1 and 3	1	$(-0.55)^2 = 0.3025$	$\sigma_e^2 + 1.5\sigma_s^2$
Sire 1 compared with sire 3	1	(–0.5)2/2 = 0.1250	$\sigma_e^2 + \sigma_s^2$
Residual	1	R = 0.1800	σ_e^2

Fitting a linear model in σ_e^2 and σ_s^2 to the three sums of squares, 0.3025, 0.1250 and 0.1800, gives estimates of $\sigma_e^2 = 0.413$ (kg^2) and $\sigma_s^2 = 0.079$ (kg^2). If a generalized linear model is fitted iteratively to the sum of squares with weights proportional to the variance of the sum of squares when the procedure converges, the estimate of σ_e^2 is 0.163 (kg^2) and σ_s^2 is 0.047 (kg^2). The estimated variances of these estimates (from the inverse of the generalized least-squares coefficient matrix) are 0.216 (kg^2) and 0.234 (kg^2).

11.5 Animal Model

It has been shown that estimates can be obtained from analysis of variance for some models. Now consider a more general model, the animal model introduced in Chapter 3. This linear model [3.1] is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

and the variance structure is defined, with:

$$\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}; \quad \text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2 = \mathbf{G} \quad \text{and} \quad \text{cov}(\mathbf{a}, \mathbf{e}) = \text{cov}(\mathbf{e}, \mathbf{a}) = 0$$

where **A** is the numerator relationship matrix, and there is interest in estimating σ_a^2 and σ_e^2. A popular method of estimation is by residual maximum likelihood (REML) (Patterson and Thompson, 1971). This is based on a log-likelihood of the form:

$$\text{L} \propto (\tfrac{1}{2})\{-(\mathbf{y} - \mathbf{Xb})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb}) - \text{logdet}(\mathbf{V}) - \text{logdet}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\}$$

where **b** is the generalized least-square solution (GLS) and satisfies:

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

There are three terms in L: the first is a weighted sum of squares of residuals, the second a term that depends on the variance matrix and a third that depends on the variance matrix of the fixed effects and can be thought of as a penalty because fixed effects are estimated. Mixed model equations (Chapter 3) play an important part in the analysis process.

For the particular model, these can be written as [3.4]:

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

with $\alpha = \sigma_e^2/\sigma_a^2$ or $(1 - h^2)/h^2$.

Extensive use of the prediction error variance matrix of **a** is made. In this case the prediction error variance matrix is PEV = $\text{var}(\mathbf{a} - \hat{\mathbf{a}}) = \mathbf{C}^{22}\sigma_e^2$ [3.14], where $\mathbf{C}^{22}$ is associated with the coefficient matrix of the mixed model equations.

Estimates of σ_a^2 and σ_e^2 are chosen to maximize L. It is useful to express relevant terms in this estimation process in terms of the projection matrix **P**:

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$$

Then:

$$\mathrm{L}\,\alpha(\tfrac{1}{2})\{-\mathbf{y}'\mathbf{P}\mathbf{y} - \mathrm{logdet}(\mathbf{V}) - \mathrm{logdet}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\} \qquad [11.1]$$

Estimation of a variance parameter $\theta_i (\theta_1 = \sigma_e^2, \theta_2 = \sigma_a^2)$ involves setting to zero the first derivatives:

$$\partial \mathrm{L}/\partial\theta_i = (\tfrac{1}{2})\{\mathbf{y}'\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}\mathbf{y} - \mathrm{trace}[\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)]\}$$

These equations could be thought of as equating a function of data (the first term in the expression) to its expectation.

Normally, finding a maximum requires an iterative scheme. One suggested by Patterson and Thompson (1971) was based on using the expected value of the second differential matrix. In this case these are:

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\theta_i\partial\theta_j) = -(\tfrac{1}{2})\mathrm{trace}[\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}(\partial\mathbf{V}/\partial\theta_j)]$$

Using the first and expected second differentials one can update $\boldsymbol{\theta}$ using terms that depend on the solution of the mixed model equations and prediction error variances. For the particular animal model that is being considered, then:

$$\partial \mathrm{L}/\partial\sigma_e^2 = (\tfrac{1}{2})\{(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a})'(\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a})/\sigma_e^4 - (n - p - q)/\sigma_e^2 - \mathrm{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]/\sigma_a^2\} \qquad [11.2]$$

$$\partial \mathrm{L}/\partial\sigma_a^2 = (\tfrac{1}{2})\{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/\sigma_a^4 + q/\sigma_a^2 - \mathrm{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2/\sigma_a^4\} \qquad [11.3]$$

and:

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\sigma_e^4) = -(\tfrac{1}{2})\{(n - p - q)/\sigma_e^4 + \mathrm{trace}[(\mathbf{C}^{22}\mathbf{A}^{-1})^2]/\sigma_a^4\}$$

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\sigma_a^4) = -(\tfrac{1}{2})\{\mathrm{trace}[\{\mathbf{I} - \mathbf{C}^{22}\mathbf{A}^{-1}(\sigma_e^2/\sigma_a^2)\}^2]/\sigma_a^4\}$$

$$\mathrm{E}(\partial \mathrm{L}^2/\partial\sigma_a^2\partial\sigma_e^2) = -(\tfrac{1}{2})\{\mathrm{trace}[\{\mathbf{I} - \mathbf{C}^{22}\mathbf{A}^{-1}(\sigma_e^2/\sigma_a^2)\}\{\mathbf{C}^{22}\mathbf{A}^{-1}\}]\sigma_a^4\}$$

Thinking of the variance parameters and the first differentials as vectors $\boldsymbol{\theta}$ and $\partial\mathbf{L}/\partial\boldsymbol{\theta}$ with *i*th ($i = 1, 2$) element θ_i and $\partial \mathrm{L}/\partial\theta_i$, respectively, and **Einf**, the expected information matrix, a matrix with *i,j*th element $-\mathrm{E}(\partial \mathrm{L}^2/\partial\theta_i\partial\theta_j)$, suggests an iterative scheme, with the new estimate $\boldsymbol{\theta}_n$ satisfying:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta} + \mathbf{Einf}^{-1}(\partial\mathbf{L}/\partial\boldsymbol{\theta}) \qquad [11.4]$$

There are two problems with this approach. First, the parameters might go negative and one would want estimates of variances to stay essentially positive. One popular way of avoiding this property is to note that at a maximum of the likelihood the first differentials are zero and to manipulate equations [11.1] and [11.2] in the form:

$$(n - p)\sigma_e^2 = (\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a})'(\mathbf{y}) \qquad [11.5]$$

$$q\sigma_a^2 = \mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + \mathrm{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2 \qquad [11.6]$$

so that it can be seen that σ_e^2 is estimated from a sum of squares of residuals and σ_a^2 is estimated from a weighted sum of squares of predicted values and their prediction error variance. This algorithm is an EM algorithm

(Dempster *et al.*, 1977) and successive iterates are positive. The algorithm can be written in the form of the updating formula if **Einf** is replaced by a matrix that depends on the information derived, as if one could directly observe the residuals and breeding values rather than predicting them. This algorithm can be slow to converge in animal breeding applications.

A second problem is that the expected second differentials are difficult to calculate. Sometimes it is recommended to use observed second differentials. These are of the form:

$$(\partial L^2/\partial\theta_i\partial\theta_j) = -\mathbf{y}'\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}(\partial\mathbf{V}/\partial\theta_j)\mathbf{P}\mathbf{y} + (\tfrac{1}{2})\text{trace}[\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}(\partial\mathbf{V}/\partial\theta_j)]$$

but again these terms involve the complicated trace terms. One suggestion (Gilmour *et al.*, 1995) is to use the average of the expected and observed information terms. These are of the form:

$$A(\partial L^2/\partial\theta_i\partial\theta_j) = -(\tfrac{1}{2})\{\mathbf{y}'\mathbf{P}(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}(\partial\mathbf{V}/\partial\theta_j)\mathbf{P}\mathbf{y}\}$$

These terms are similar to $\mathbf{y}'\mathbf{P}\mathbf{y}$ in that they could be thought of as a weighted sum of squares matrix with $\mathbf{y}$ replaced by two columns $(\partial\mathbf{V}/\partial\theta_i)\mathbf{P}\mathbf{y}\,(i = 1,2)$. In this particular case:

$$(\partial\mathbf{V}/\partial\sigma_e^2)\mathbf{P}\mathbf{y} = (\mathbf{y} - \mathbf{X}\mathbf{b} - \mathbf{Z}\mathbf{a})/\sigma_e^2$$

and:

$$(\partial\mathbf{V}/\partial\sigma_a^2)\mathbf{P}\mathbf{y} = \mathbf{Z}\mathbf{a}/\sigma_a^2$$

As in the formation of **Einf**, we can construct and base an iterative scheme using [11.3] and on **Ainf**, a matrix with elements $-A(\partial L^2/\partial\theta_i\partial\theta_j)$. Once the iterative scheme has converged, then the asymptotic variance matrix of $\boldsymbol{\theta}$ can be estimated from $\mathbf{Ainf}^{-1}$ or $\mathbf{Einf}^{-1}$. The animal model and estimation procedure introduced can easily be extended to deal with other models, just as prediction procedures can be developed for a variety of models. Software for estimating variance parameters using this average information algorithm is described by Jensen and Madsen (1997) and Gilmour *et al.* (2003).

11.6 Numerical Example

Consider the data in Table 11.2 for the pre-weaning gain (WWG) of beef calves. This is very similar to the data of Table 3.1, with the data changed to give positive variance estimates.

The model to describe the observations is:

$$y_{ijk} = p_i + a_j + e_{ijk}$$

where y_{ij} = the WWG of the jth calf of the ith sex, p_i = the effect of the ith sex, a_j = the random effect of the jth calf, and e_{ijk} = random error effect.

In matrix notation the model is the same as described in equation [3.1].

Again, the objective is to illustrate the estimation of variance components σ_e^2 and σ_a^2 on a very small example so that the calculations can be expressed concisely.

Table 11.2. Pre-weaning gain (kg) for five beef calves.

Calf	Sex	Sire	Dam	WWG (kg)
4	Male	1	–	2.6
5	Female	3	2	0.1
6	Female	1	2	1.0
7	Male	4	5	3.0
8	Male	3	6	1.0

In matrix notation the model is the same as described in equation [3.1], with $n = 5$, $p = 2$ and $q = 8$, with the design matrices as given in Section 3.2. Now $\mathbf{y}' = [2.6, 0.1, 1.0, 3.0, 1.0]$ and, using initial estimates of $\sigma_e^2 = 0.4$ and $\sigma_a^2 = 0.2$, solutions to mixed model equations [3.15] are:

Effects	Solutions
Sex[a]	
1	2.144
2	0.602
Animal	
1	0.117
2	–0.025
3	–0.222
4	–0.254
5	–0.135
6	0.032
7	0.219
8	–0.305

[a]1 = male, 2 = female.

Then:

$$(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})' = [0.2022 \quad -0.3661 \quad 0.3661 \quad 0.6374 \quad -0.8395]$$

$$\mathbf{C}^{22}\sigma_e^2 = \begin{bmatrix} 0.1884 & 0.0028 & 0.0131 & 0.0878 & 0.0180 & 0.0883 & 0.0554 & 0.0537 \\ 0.0028 & 0.1968 & -0.0041 & 0.0082 & 0.0949 & 0.0981 & 0.0479 & 0.0443 \\ 0.0131 & -0.0041 & 0.1826 & 0.0193 & 0.0805 & 0.0090 & 0.0504 & 0.0871 \\ 0.0878 & 0.0082 & 0.0193 & 0.1711 & 0.0188 & 0.0510 & 0.0971 & 0.0493 \\ 0.0180 & 0.0949 & 0.0805 & 0.0188 & 0.1712 & 0.0679 & 0.0879 & 0.0712 \\ 0.0883 & 0.0981 & 0.0090 & 0.0510 & 0.0679 & 0.1769 & 0.0609 & 0.0877 \\ 0.0554 & 0.0479 & 0.0504 & 0.0971 & 0.0879 & 0.0609 & 0.1767 & 0.0672 \\ 0.0537 & 0.0443 & 0.0871 & 0.0493 & 0.0712 & 0.0877 & 0.0672 & 0.1689 \end{bmatrix}$$

$\mathbf{y}'\mathbf{Py} = 4.8193$, $\log\det(\mathbf{V}) = -2.6729$ and $\log\det(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = 2.6241$, so L = –2.3852 from [11.1].

Then [11.2] and [11.3] give:

$$\partial L/\partial\sigma_e^2 = (0.5)\{(\mathbf{y}-\mathbf{Xb}-\mathbf{Za})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Za})/\sigma_e^4 - (n-p-q)/\sigma_e^2 - \text{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]/\sigma_a^2\}$$
$$\partial L/\partial\sigma_e^2 = (0.5)\{8.8753 - (-12.5000) - 18.0733\} = 1.6510$$
$$\partial L/\partial\sigma_a^2 = (0.5)\{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}/\sigma_a^4 - q/\sigma_a^2 + \text{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2/\sigma_a^4\}$$
$$\partial L/\partial\sigma_a^2 = (0.5)\{6.3461 - 40.0000 + 36.1466\} = 1.2464$$

and:

$$A(\partial L^2/\partial\sigma_e^4) = -(0.5)\{(\mathbf{y}-\mathbf{Xb}-\mathbf{Za})'\mathbf{P}(\mathbf{y}-\mathbf{Xb}-\mathbf{Za})/\sigma_e^4\}$$
$$A(\partial L^2/\partial\sigma_e^4) = -(0.5)16.5346 = -8.2673$$
$$A(\partial L^2/\partial\sigma_a^4) = -(0.5)\{\mathbf{a}'\mathbf{Z}'\mathbf{PZa}\}/\sigma_a^4$$
$$A(\partial L^2/\partial\sigma_a^4) = -(0.5)9.1163 = -4.5582$$
$$A(\partial L^2/\partial\sigma_a^2\partial\sigma_e^2) = -(0.5)\{\mathbf{a}'\mathbf{Z}'\mathbf{P}(\mathbf{y}-\mathbf{Xb}-\mathbf{Za})\}/(\sigma_e^2\sigma_a^4)$$
$$A(\partial L^2/\partial\sigma_a^2\partial\sigma_e^2) = -(0.5)11.3070 = -5.6535$$

and:

$$\mathbf{Ainf} = \begin{bmatrix} 8.2673 & 5.6535 \\ 5.6535 & 4.5582 \end{bmatrix} \quad \text{so } \mathbf{Ainf}^{-1} = \begin{bmatrix} 0.7967 & -0.9882 \\ -0.9882 & 1.4450 \end{bmatrix}$$

Using [11.4] and replacing **Einf** by **Ainf**:

$$\boldsymbol{\theta}_n = \boldsymbol{\theta} + \mathbf{Ainf}^{-1}(\partial\mathbf{L}/\partial\boldsymbol{\theta}) = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.7967 & -0.9882 \\ -0.9882 & 1.4450 \end{bmatrix}\begin{bmatrix} 1.6510 \\ 1.2464 \end{bmatrix}$$
$$= \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.0838 \\ 0.1695 \end{bmatrix}$$

so that new estimates of σ_e^2 and σ_a^2 are 0.4838 (kg^2), and 0.3695 (kg^2), respectively.

Table 11.3 gives six successive iterates and log-likelihood for these data. In the last iteration $\mathbf{Ainf}^{-1} = \begin{bmatrix} 2.4436 & -3.2532 \\ -3.2532 & 5.3481 \end{bmatrix}$, so that the estimate of σ_e^2 is 0.4835 with standard error $\sqrt{2.4436} = 1.563$ and the estimate of σ_a^2 is 0.5514 with standard error $\sqrt{5.3481} = 2.313$.

Table 11.3. Estimates of σ_e^2 and σ_a^2 and **L**.

Iterate	σ_e^2 (kg^2)	σ_a^2 (kg^2)	**L**
1	0.4000	0.2000	–2.3852
2	0.4838	0.3695	–2.2021
3	0.4910	0.5126	–2.1821
4	0.4839	0.5500	–2.1817
5	0.4835	0.5514	–2.1817
6	0.4835	0.5514	–2.1817

In contrast, if estimates of $\sigma_e^2 = 0.4$ and $\sigma_a^2 = 0.2$ are used in conjunction with [11.5] and [11.6], then:

$(n - p)\sigma_e^2 = (\mathbf{y} - \mathbf{Xb} - \mathbf{Za})'(\mathbf{y})$, so $3\sigma_e^2 = 1.9277$, so $\sigma_e^2 = 0.6426$ (kg^2) and $q\sigma_a^2 = \mathbf{a}'\mathbf{A}^{-1}\mathbf{a} + \text{trace}[\mathbf{C}^{22}\mathbf{A}^{-1}]\sigma_e^2$, so $8\sigma_a^2 = 0.2538 + 1.4458$, so $\sigma_a^2 = 0.2125$ (kg^2), with L = –2.3852. After 1000 iterations, the algorithm gives $\sigma_e^2 = 0.4842$ (kg^2) and $\sigma_a^2 = 0.5504$ (kg^2), with L = –2.1817, showing that this algorithm is slower to converge.

12 Application of Gibbs Sampling in Variance Component Estimation and Prediction of Breeding Value

12.1 Introduction

Gibbs sampling is a numerical integration method and is one of several Markov chain Monte Carlo (MCMC) methods. They involve drawing samples from specified distributions; hence they are called Monte Carlo and are referred to as Markov chain because each sample depends on the previous sample. Specifically, Gibbs sampling involves generating random drawings from marginal posterior distributions through iterative sampling from the conditional posterior distributions. For instance, given that $\mathbf{Q}' = (\mathbf{Q}_1, \mathbf{Q}_2)$ and $P(\mathbf{Q}_1, \mathbf{Q}_2)$ is the joint distribution of $\mathbf{Q}_1$ and $\mathbf{Q}_2$, Gibbs sampling involves sampling from the full conditional posterior distributions of $\mathbf{Q}_1$, $P(\mathbf{Q}_1 | \mathbf{Q}_2)$ and $\mathbf{Q}_2$, $P(\mathbf{Q}_2 | \mathbf{Q}_1)$.

Thus, given that the joint posterior distribution is known to proportionality, the conditional distributions can be generated. However, defining the joint density involves the use of Bayes' theorem. In general, given that the probability of two events occurring together, $P(B,Y)$, is:

$$P(B,Y) = P(B)P(Y|B) = P(Y)P(B|Y)$$

then:

$$P(B|Y) = P(B)P(Y|B)/P(Y) \qquad [12.1]$$

Equation [12.1] implies that inference about the variable B depends on the prior probability of its occurrence, $P(B)$. Given that observations on Y are available, this prior probability is then updated to obtain the posterior probability or density of B, $(P(B|Y))$. Equation [12.1] is commonly expressed as:

$$P(B|Y) \propto P(B)P(Y|B) \qquad [12.2]$$

as the denominator is not a function of B. Therefore the posterior density of B is proportional to the prior probability of B times the conditional distribution of Y given B. Assuming that B in [12.2] is replaced by $\mathbf{W}$, a vector of parameters, such that $\mathbf{W}' = (W_1, W_2, W_3)$ and that the joint posterior distribution is known to proportionality (equation [12.2]), the full conditional probabilities needed for the Gibbs sampler can be generated for each parameter as $P(W_1 \mid W_2, W_3, Y)$, $P(W_2 \mid W_1, W_3, Y)$ and $P(W_3 \mid W_1, W_2, Y)$. Assuming starting values $W_1^{[0]}, W_2^{[0]}$ and $W_3^{[0]}$, the implementation of the Gibbs sampler involves iterating the following loop:

1. Sample $W_1^{[i+1]}$ from $P(W_1 | W_2^{[i]}, W_3^{[i]}, Y)$
2. Sample $W_2^{[i+1]}$ from $P(W_2 | W_1^{[i+1]}, W_3^{[i]}, Y)$
3. Sample $W_3^{[i+1]}$ from $P(W_3 | W_2^{[i+1]}, W_3^{[i+1]}, Y)$

Usually, the initial samples are discarded (the so-called burn-in period). In summary, the application of the Gibbs sampler involves defining the prior distributions and the joint posterior density and generating the full conditional posterior distributions and sampling from the latter.

The Gibbs sampler was first implemented by Geman and Geman (1984). In animal breeding, Wang *et al.* (1993, 1994) used Gibbs sampling for variance component estimation in sire and animal models. It has been implemented for the study of covariance components in models with maternal effects (Jensen *et al.*, 1994), in threshold models (Sorensen *et al.,* 1995) and in random regression models (Jamrozik and Schaeffer, 1997). It has recently been employed for the purposes of variance component estmation and breeding value prediction in linear-threshold models (Heringstad *et al.*, 2002; Wang *et al.*, 2002). Detailed presentations of the Gibbs sampling within the general framework of Bayesian inference and its application for variance components estimation under several models have been published by Sorensen and Gianola (2002). In this chapter, the application of the Gibbs sampler for variance component estimation and prediction of breeding values with a univariate and multivariate animal models is presented and illustrated.

12.2. Univariate Animal Model

Consider the following univariate linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where terms are as defined in [3.1] but with $\mathbf{u} = \mathbf{a}$ in [3.1]. The conditional distribution which generates the data, $\mathbf{y}$, is:

$$\mathbf{y} | \mathbf{b}, \mathbf{u}, \sigma_e^2 \sim \mathrm{N}(\mathbf{Xb} + \mathbf{Zu} + \mathbf{R}\sigma_e^2) \qquad [12.3]$$

12.2.1 Prior distributions

Prior distributions of $\mathbf{b}$, $\mathbf{u}$, σ_u^2 and σ_e^2 are needed to complete the Bayesian specification of the model (Wang *et al.*, 1993). Usually a flat prior

distribution is assigned to **b**. Thus:

$$P(\mathbf{b}) \sim \text{constant} \qquad [12.4]$$

This represents an improper or 'flat' prior distribution, denoting lack of prior knowledge about this vector. However, if there is information a priori about the value of **b** in terms of upper or lower limits, this can be incorporated in defining the posterior distribution of **b**. Such a prior distribution will be termed proper prior distribution. Assuming an infinitesimal model, the distribution of **u** is multivariate normal and is:

$$\mathbf{u}|\mathbf{A}, \sigma_u^2 \sim \mathrm{N}(\mathrm{O}, \mathbf{A}\sigma_u^2) \qquad [12.5]$$

A scaled inverted chi-squared distribution (χ^2) is usually used as priors for the variance components (Wang *et al.*, 1993). Thus, for the residual variance:

$$P(\sigma_e^2|v_e, s_e^2) \propto (\sigma_e^2)^{-\left(\frac{v_e}{2}+1\right)} \exp\left(-\frac{v_e s_e^2}{2\sigma_e^2}\right) \qquad [12.6]$$

and the additive genetic variance:

$$P(\sigma_u^2|v_u, s_u^2) \propto (\sigma_u^2)^{-\left(\frac{v_u}{2}+1\right)} \exp\left(-\frac{v_u s_u^2}{2\sigma_u^2}\right) \qquad [12.7]$$

where v_e (v_u) is a 'degree of belief' parameter and s_e^2 (s_u^2) can be interpreted as a prior value of the appropriate variance component. Alternatively, prior uniform distribution could be assigned to the variance components such that:

$$P(\sigma_j^2) \propto \text{constant} \qquad [12.8]$$

where $\sigma_j^2 = \sigma_e^2$ or σ_u^2 and an upper limit might be assigned for σ_j^2 based on prior knowledge. Setting v_e or v_u to –2 and s_e^2 or s_u^2 to 0 in [12.6] or [12.7] gives [12.8].

12.2.2 Joint and full conditional distributions

The joint posterior distribution of the parameters (**b**, **u**, σ_e^2 or σ_u^2) is proportional to the product of the likelihood function and the joint prior distribution. Using [12.3] to [12.7], the joint posterior distribution can be written as:

$$\begin{aligned} &P(\mathbf{b}, \mathbf{u}, \sigma_u^2, \sigma_e^2|\mathbf{y}) \\ &\quad \propto (\sigma_e^2)^{-\left(\frac{n+v_e}{2}+1\right)} \exp\left[-\frac{(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})+v_e s_e^2}{2\sigma_e^2}\right] \\ &\qquad (\sigma_u^2)^{-\left(\frac{m+v_u}{2}+1\right)} \exp\left(\frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}+v_u s_u^2}{2\sigma_u^2}\right) \end{aligned} \qquad [12.9]$$

assuming n observations and m animals. Setting v_e or v_u and s_e^2 or s_u^2 to zero gives the joint posterior distributions for the uniform distribution in [12.8].

The full conditional posterior distribution of each parameter is obtained by regarding all other parameters in [12.9] as known. Thus, for **b**:

$$P(\mathbf{b}|\mathbf{u},\sigma_u^2,\sigma_e^2,\mathbf{y}) \propto \exp\left(-\frac{(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})}{2\sigma_e^2}\right) \qquad [12.10]$$

A corresponding distribution to the above is:

$$\mathbf{Xb}|\mathbf{u},\sigma_u^2,\sigma_e^2,\mathbf{y} \sim \mathrm{N}(\mathbf{y}-\mathbf{Zu},\mathbf{I}\sigma_e^2)$$

or:

$$\mathbf{X'Xb}|\mathbf{u},\sigma_u^2,\sigma_e^2,\mathbf{y} \sim \mathrm{N}(\mathbf{X'}(\mathbf{y}-\mathbf{Zu}),\mathbf{X'X}\sigma_e^2)$$

Therefore:

$$\mathbf{b}|\mathbf{u},\sigma_u^2,\sigma_e^2,\mathbf{y} \sim \mathrm{N}(\hat{\mathbf{b}},(\mathbf{X'X})^{-1}\sigma_e^2)$$

where:

$$\hat{\mathbf{b}} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{y}-\mathbf{Zu})$$

Thus, for the jth level of **b**:

$$b_j|\mathbf{b}_{-j},\mathbf{u},\sigma_u^2,\sigma_e^2,\mathbf{y} \sim \mathrm{N}(\hat{b}_j,(\mathbf{x}_j'\mathbf{x}_j)^{-1}\sigma_e^2) \qquad (12.11)$$

with $\hat{b}_j = (\mathbf{x}_j'\mathbf{x}_j)^{-1}\mathbf{x}_j'(\mathbf{y}_j - \mathbf{X}_{-j}\mathbf{b} - \mathbf{Zu})$, which is equivalent to [3.5], $\mathbf{x}_j$ is the jth row of **X** and $\mathbf{b}_{-j}$ is the vector **b** with level j deleted.

Similarly, the distribution for the jth random effect is:

$$\mathbf{u}_j|\mathbf{b},\mathbf{u}_{-j},\sigma_u^2,\sigma_e^2,\mathbf{y} \sim \mathrm{N}(\hat{\mathbf{u}}_j,(\mathbf{z}_j'\mathbf{z}_j + \mathbf{A}_{j,j}^{-1}\alpha)^{-1}\sigma_e^2) \qquad [12.12]$$

with:

$$\hat{\mathbf{u}}_j = (\mathbf{z}_j'\mathbf{z}_j + \mathbf{A}_{j,j}^{-1}\alpha)^{-1}\mathbf{z}_j'(y - \mathbf{Xb} - \mathbf{A}_{j,-j}^{-1}\alpha\mathbf{u}_{-j})$$

which is equivalent to [3.8].

The full conditional of distribution of the residual variance is derived from [12.9] by considering only terms that involve σ_e^2 and is in the scaled inverted χ^2 form (Wang *et al.*, 1993). Thus, for the residual variance:

$$P(\sigma_e^2|\mathbf{b},\mathbf{u},\sigma_u^2,\mathbf{y}) \propto (\sigma_e^2)^{-\left(\frac{n+v_e}{2}+1\right)} \exp\left(-\frac{\ddot{v}_e\ddot{s}_e^2}{2\sigma_e^2}\right)$$

where:

$$\ddot{v}_e = n + v_e \quad \text{and} \quad \ddot{s}_e^2 = ((\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})) + v_e s_e^2)/\ddot{v}_e$$

Hence:

$$\sigma_e^2|\mathbf{b},\mathbf{u},\sigma_u^2,\mathbf{y} \sim \ddot{v}_e\ddot{s}_e^2\chi_{\ddot{v}_e}^{-2} \qquad [12.13]$$

which involves sampling from a χ^2 distribution with scale parameter $(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu})'(\mathbf{y}-\mathbf{Xb}-\mathbf{Zu}) + v_e s_e^2$ and $\ddot{\mathbf{v}}_e$ degrees of freedom.

Similarly, the full conditional distribution of σ_u^2 is also in the form of an inverted chi-square. Thus:

$$P(\sigma_u^2|\mathbf{b},\mathbf{u},\sigma_e^2,\mathbf{y}) \propto (\sigma_u^2)^{-\left(\frac{m+v_u}{2}+1\right)} \exp\left(-\frac{\vec{v}_u \vec{s}_u^2}{2\sigma_u^2}\right)$$

where:

$$\vec{v}_u = m + v_u \quad \text{and} \quad \vec{s}_u^2 = ((\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}) + v_u s_u^2)/\vec{v}_u$$

Thus:

$$\sigma_u^2|\mathbf{b},\mathbf{u},\sigma_u^2,\mathbf{y} \sim \vec{v}_u \vec{s}_u^2 \chi_{\vec{v}_u}^{-2} \qquad [12.14]$$

which involves sampling from a χ^2 distribution with scale parameter $(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}) + v_u s_u^2$ and $\vec{\mathbf{v}}_u$ degrees of freedom.

The Gibbs sampling then consists of setting initial values for $\mathbf{b}$, $\mathbf{u}$, σ_u^2 and σ_e^2 and iteratively sampling successively from [12.11] to [12.14], using updated values of the parameters from the i round in the $i+1$ round. Assuming that k rounds of iteration were performed, then k is called the length of the chain. As mentioned earlier, the first j samples are usually discarded as the burn-in period. This is to ensure that samples saved are not influenced by the priors but are draws from the posterior distribution. The size of j is determined rather arbitrarily, but a graphical illustration could help.

Several strategies can be implemented in using the Gibbs sampler and these have an effect on the degree of correlation between the values sampled. Details of various strategies are discussed in detail by Sorensen and Gianola (2002) and therefore not presented here. One approach is to run a single long chain. A sample $(\mathbf{b},\mathbf{u},\sigma_u^2,\sigma_e^2)$ is saved at every dth iterate until a total t samples are saved and analysed. The larger d is, the lower the degree of autocorrelation between the samples. Another strategy, known as the multiple chain or short chain approach, involves carrying out several parallel t runs and saving the last nth sample from each run. Thus this approach produces $m = nt$ samples. The different chains will produce different realizations, even if the same starting values are used. However, if the parameters in the model are highly correlated, it might be useful to utilize different starting values in the different chains.

Determining convergence with the Gibbs sampler is not very straightforward but it is advisable, depending on the size of the problem, to run several chains and check convergence graphically.

12.2.3 Inferences from the Gibbs sampling output

The samples saved are usually analysed to estimate posterior means or variances of the posterior distribution. Detailed discussion of various estimation methods is covered in Sorensen and Gianola (2002) and not presented here. Given that $\mathbf{w}$ is a vector of size k, containing the saved

samples, then the posterior mean and variance can be computed, respectively, as:

$$\mu_f = \frac{\sum_{i=1}^{k} f(w_i)}{k} \qquad [12.15]$$

and:

$$\text{var}(\mu_f) = \frac{\sum_{i=1}^{k} (f(w_i) - \mu_f)^2}{k}$$

where $f(\mathbf{w})$ is a function of interest of the variables in **w**. For instance, in the linear animal model in Section 12.2 the function of interest would be the variance components (σ_u^2 and σ_e^2) and the vectors **b** and **u**.

The above estimates from the posterior distributions are associated with sampling variance (Monte Carlo variance). The larger the number of samples analysed, the smaller the sampling variance. It is usually useful to get an estimate of the sampling variance associated with the estimates from the posterior distributions. An empirical estimate could be obtained by running several independent runs and then computing the between-chain variance of the estimates obtained for each run. This is not computationally feasible in most practical situations and various methods are used to estimate this variance. A number of such estimators are fully discussed by Sorensen and Gianola (2002). A simple method that could be used involves calculating the batch effective chain size. Given a chain of size k, successive samples are grouped into b batches, each of size t. The average of the jth batch can be computed as:

$$\overline{u}_j = \frac{\sum_{i=1}^{t} f(w_i)}{t}$$

The batch estimator of the variance of μ in equation [12.15] is:

$$\text{var}_b(\mu) = \frac{\sum_{j=1}^{b} (\overline{u}_j - \mu)^2}{b(b-1)}$$

The batch effective chain size can be obtained as:

$$\psi_b = \frac{\sum_{i=1}^{k} [f(w_i) - \mu]^2}{(k-1)\,\text{var}_b(u)}$$

If samples are uncorrelated, then $\psi = k$. The difference between ψ and k gives an idea of the degree of the autocorrelation among the samples in the chain.

12.2.4 Numerical application

Example 12.1

Using the data in Example 3.1 and the variance components, the application of Gibbs sampling for estimation of variance components and the prediction of breeding values is illustrated. Uniform priors are assumed for the variance components such that $v_e = v_a = -2$ and $s_e^2 = s_u^2 = 0$. A flat prior is assumed for $\mathbf{b}$, and $\mathbf{u}$ is assumed to be normally distributed.

First, sample $\mathbf{b}_1^{[1]}$, where, the superscript in brackets denotes iteration number, using [12.11], with $\hat{\mathbf{b}}_1$ calculated using [3.5] and $(\mathbf{x}_j'\mathbf{x})^{-1}\sigma_e^2 = (3)^{-1}40 = 13.333$. From [3.5]:

$$\hat{\mathbf{b}}_1 = [(4.5 + 3.5 + 5.0) - (0 + 0 + 0)]/3 = 4.333$$

Assuming the random number (RN) generated from a normal distribution, N(0,1), is 0.1704, then $\mathbf{b}_1$ from [12.11] is:

$$\mathbf{b}_1^{[1]} = (4.333 + 0.1704\sqrt{(13.333)}) = 4.955$$

Then sample $\mathbf{b}_2$ using [12.11] with $(\mathbf{x}_j'\mathbf{x})^{-1}\sigma_e^2 = (2)^{-1}40 = 20$ and $\hat{\mathbf{b}}_2$ is:

$$\hat{\mathbf{b}}_2 = [(2.9 + 3.9) - (0 + 0)]/2 = 3.40$$

Assuming the RN from N(0,1) is –0.1294, then:

$$\mathbf{b}_2^{[1]} = 3.40 - 0.1294\sqrt{20} = 2.821$$

The vector of solution $\mathbf{u}_j$ for animal j is sampled using [12.12], with $\hat{\mathbf{u}}_j$ calculated using equation [3.8]. Thus, for animal 1:

$$\hat{\mathbf{u}}_1 = 0 \quad \text{and} \quad (\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}_{1,1}^{-1}\alpha)^{-1}\sigma_e^2 = (3.667)^{-1}40 = 10.908$$

The value of $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}_{1,1}^{-1}\alpha)^{-1}$ is taken from the diagonal element of the coefficient matrix of the mixed model equations (MME) for Example 3.1. Assuming the RN from N(0,1) is 0.2067:

$$\mathbf{u}_1^{[1]} = 0 + 0.2067\sqrt{10.98} = 0.683$$

For animal 2, $\hat{\mathbf{u}}_2$ from [3.8] = –0.171, and $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}_{1,1}^{-1}\alpha)^{-1}\sigma_e^2 = (4)^{-1}40 = 10$. Then, from [12.12], assuming RN from N(0,1) is –1.8025:

$$\mathbf{u}_2^{[1]} = -0.171 + (-1.8025\sqrt{10}) = -5.871$$

Similarly, given that $\hat{\mathbf{u}}_3$ from [3.8] = 1.468, $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}_{1,1}^{-1}\alpha)^{-1}\sigma_e^2 = (4)^{-1}40 = 10$ and RN = –0.5558, then:

$$\mathbf{u}_3^{[1]} = 1.468 - 0.5558\sqrt{10} = -0.290$$

For animal 4, $\hat{\mathbf{u}}_4 = 0.0976$ from [3.8], $(\mathbf{z}_1'\mathbf{z}_1 + \mathbf{A}_{1,1}^{-1}\alpha)^{-1}\sigma_e^2 = (4.667)^{-1}40 = 8.571$ and RN = –1.8654; then:

$$\mathbf{u}_4^{[1]} = 0.0976 - 1.8654\sqrt{8.571} = -5.364$$

Similar calculations using [12.12] gave estimates of $\mathbf{u}_5^{[1]}$, $\mathbf{u}_6^{[1]}$, $\mathbf{u}_7^{[1]}$ and $\mathbf{u}_8^{[1]}$ as –3.097, –2.577, –1.621 and –0.697, respectively.

The vector of residuals, $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{Xb} - \mathbf{Zu}$, is:

$$\begin{pmatrix} \hat{e}_4 \\ \hat{e}_5 \\ \hat{e}_6 \\ \hat{e}_7 \\ \hat{e}_8 \end{pmatrix} = \begin{pmatrix} 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5.0 \end{pmatrix} - \begin{pmatrix} 4.955 \\ 2.821 \\ 2.821 \\ 4.955 \\ 4.955 \end{pmatrix} - \begin{pmatrix} -5.364 \\ -3.097 \\ -2.577 \\ -1.621 \\ -0.697 \end{pmatrix} = \begin{pmatrix} 4.908 \\ 3.176 \\ 3.656 \\ 0.165 \\ 0.742 \end{pmatrix}$$

and $\hat{\mathbf{e}}'\hat{\mathbf{e}} = 48.118$. Sampling from the inverted χ^2 distribution with 3 degrees of freedom [12.13] gave an estimate of 39.870 for the residual variance.

Using [12.14], sampling for σ_u^2 is again from the inverted χ^2 distribution, with $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} = 93.11$ and degrees of freedom being 6. An estimate of 23.913 was obtained for σ_u^2. Note that it is easier to compute $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}$ using [2.3]. Thus $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u} = \mathbf{u}'(\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1}\mathbf{u} = \mathbf{m}'\mathbf{D}\mathbf{m}$, where $\mathbf{m} = \mathbf{T}^{-1}\mathbf{u}$, with $\mathbf{m}$ being a vector of Mendelian sampling for animals, calculated using [2.2].

The next round of iteration is then commenced, using the updated values computed for the parameters.

12.3 Multivariate Animal Model

In this section, the Gibbs sampling algorithm developed by Jensen *et al.* (1994) for models with maternal genetic effects is generalized for a multivariate situation. Given that animals are ordered within traits, the multivariate model for two traits could be written as:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix}$$

where terms are as defined in equation [5.1], with $\mathbf{u} = \mathbf{a}$. The conditional distribution of the complete data, given that animals are ordered within traits, is:

$$\left\langle \begin{matrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{matrix} \middle| \mathbf{b}_1, \mathbf{b}_2, \mathbf{u}_1, \mathbf{u}_2, \mathbf{R} \right\rangle \sim \mathbf{N}\left[\begin{matrix} \mathbf{X}_1\mathbf{b}_1 + \mathbf{Z}_1\mathbf{u}_1 \\ \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}_2\mathbf{u}_2 \end{matrix}, \mathbf{R} \otimes \mathbf{I} \right] \qquad [12.16]$$

It is assumed that:

$$\left\langle \begin{matrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{matrix} \middle| \mathbf{G}, \mathbf{A} \right\rangle \sim \mathbf{N}\left[\begin{matrix} \mathbf{0} \\ \mathbf{0} \end{matrix}, \mathbf{G} \otimes \mathbf{A} \right] \qquad [12.17]$$

where $\mathbf{G}$ is the genetic covariance matrix and $\mathbf{A}$ is the numerator relationship matrix.

12.3.1 Prior distributions

Assume that proper uniform distributions are defined for the fixed effects:

$P(\mathbf{b}_1) \propto$ constant; $P(\mathbf{b}_2) \propto$ constant

with:

$$\mathbf{b}_i(\min) \leq \mathbf{b}_i \leq \mathbf{b}_i(\max)$$

An inverted Wishart distribution (Jensen *et al.*, 1994) is used as prior distribution for the genetic and residual covariances. Thus the prior distribution for the residual covariance is:

$$P(\mathbf{R}|\mathbf{V}_e, v_e) \propto |\mathbf{R}|^{-\frac{1}{2}(v_e+p+1)} \exp[-\tfrac{1}{2}\operatorname{tr}(\mathbf{R}^{-1}\mathbf{V}_e^{-1})] \quad [12.18]$$

The above is a p-dimensional inverse Wishart distribution (IW_2), where p is the order of $\mathbf{R}$, $\mathbf{V}_e$ is a parameter of the prior distribution and v_e is the degree of freedom. If $\mathbf{V}_e = 0$ and $v_e = -(p+1)$, the above reduces to a uniform distribution. Similarly, for the genetic covariance, the following prior distribution is assumed:

$$P(\mathbf{G}|\mathbf{V}_u, v_u) \propto |\mathbf{G}|^{-\frac{1}{2}(v_u+p+1)} \exp[-\tfrac{1}{2}\operatorname{tr}(\mathbf{G}^{-1}\mathbf{V}_u^{-1})] \quad [12.19]$$

with terms $\mathbf{V}_u$ and v_u equivalent to $\mathbf{V}_e$ and v_e, respectively, in equation [12.18].

The joint posterior distribution, assuming n traits and using [12.16]–[12.19], is:

$$P(\mathbf{b}_1, \ldots, \mathbf{b}_n, \mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{R}, \mathbf{G}) \propto p(\mathbf{y}_1, \ldots, \mathbf{y}_n|\mathbf{b}_1, \ldots, \mathbf{b}_n, \mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{R})p(\mathbf{u}_1, \ldots, \mathbf{u}_n|\mathbf{G})p(\mathbf{G})p(\mathbf{R}) \quad [12.20]$$

12.3.2 Conditional distributions

Using the same principles as those for obtaining equations [12.11] and [12.12], the conditional distribution for level k of the ith trait is:

$$\mathbf{b}_{i,k}|\mathbf{b}_{i,-k}, \mathbf{b}_j, \mathbf{u}, \mathbf{R}_e, \mathbf{G}, \mathbf{y} \sim \mathbf{N}(\hat{\mathbf{b}}_{i,k}, (\mathbf{x}'_{i,k}\mathbf{r}^{ii}\mathbf{x}_{i,k})^{-1}); \quad j = 1, n \text{ and } j \neq i \quad [12.21]$$

with:

$$\hat{\mathbf{b}}_{i,k} = (\mathbf{x}'_{i,k}\mathbf{r}^{ii}\mathbf{x}_{i,k})^{-1}\mathbf{x}'_{i,k}(\mathbf{r}^{ii}\mathbf{y}_i + \mathbf{r}^{ij}\mathbf{y}_j) - \mathbf{r}^{ii}(\mathbf{x}'_{i,-k}\mathbf{b}_{i,-k} + \mathbf{z}_i\mathbf{u}_i) - \mathbf{r}^{ij}(\mathbf{x}_j\mathbf{b}_j + \mathbf{z}_j\mathbf{u}_j); \quad j = 1, n \text{ and } j \neq i$$

Similarly, for the random animal effect, the conditional distribution for animal k of the ith trait is:

$$\mathbf{u}_{i,k}|\mathbf{u}_{i,-k}, \mathbf{u}_j, \mathbf{b}, \mathbf{R}_e, \mathbf{G}, \mathbf{y} \sim \mathbf{N}(\hat{\mathbf{u}}_{i,k}, (\mathbf{r}^{ii}\mathbf{z}'_{i,k}\mathbf{z}_{i,k} + \mathbf{g}^{ii}\mathbf{A}^{-1}_{k,k})^{-1}), \quad j = 1, n \text{ and } j \neq i \quad [12.22]$$

with:

$$\hat{\mathbf{u}}_{i,k} = (\mathbf{z}'_{i,k}\mathbf{r}^{ii}\mathbf{z}_{i,k} + \mathbf{A}^{-1}_{k,k}\mathbf{g}^{ii})^{-1}\{\mathbf{z}'_{i,k}(\mathbf{r}^{ii}\mathbf{y}_i + \mathbf{r}^{ij}\mathbf{y}_j - \mathbf{r}^{ii}\mathbf{x}_i\mathbf{b}_i - \mathbf{r}^{ij}\mathbf{x}_j\mathbf{b}_j) - (\mathbf{z}'_{i,k}r^{ij}\mathbf{z}_{j,k} + \mathbf{A}^{-1}_{k,k}\mathbf{g}^{ij}\mathbf{u}_{j,k}) - \mathbf{A}^{-1}_{k,s}(\mathbf{g}^{ij}\mathbf{u}_{i,s} + \mathbf{g}^{ij}\mathbf{u}_{j,s})\}$$

where s represents the known parents of the kth animal.

However, instead of sampling for each level of fixed or random effects for one trait at a time, it is more efficient to implement block sampling for each level of fixed or random effect across all traits at once. The conditional distribution for level k of a fixed effect required for block sampling, assuming $n = 2$, is:

$$\left\langle \begin{matrix} \mathbf{b}_{1,k} \\ \mathbf{b}_{2,k} \end{matrix} \middle| \mathbf{b}_{-k}, \mathbf{u}, \mathbf{R}, \mathbf{G}, \mathbf{y} \right\rangle \sim \mathrm{N}\left[\begin{matrix} \hat{\mathbf{b}}_{1,k} \\ \hat{\mathbf{b}}_{2,k} \end{matrix}, (\mathbf{X}'_k \mathbf{R}^{-1} \mathbf{X}_k)^{-1} \right] \quad [12.23]$$

where:

$$\begin{pmatrix} \hat{\mathbf{b}}_{1,k} \\ \hat{\mathbf{b}}_{2,k} \end{pmatrix} = (\mathbf{X}'_k \mathbf{R}^{-1} \mathbf{X}_k)^{-1} (\mathbf{X}'_k \mathbf{R}^{-1} (\mathbf{y}_k - \mathbf{X}_{-k} \mathbf{b}_{-k} - \mathbf{Z}\hat{\mathbf{u}}))$$

which is equivalent to equation [5.4].

For the random animal effect, block sampling for animal k, assuming $n = 2$, the conditional distribution is:

$$\left\langle \begin{matrix} \mathbf{u}_{1,k} \\ \mathbf{u}_{2,k} \end{matrix} \middle| \mathbf{b}, \mathbf{u}_{j,-k}, \mathbf{R}, \mathbf{G}, \mathbf{y} \right\rangle \sim \mathrm{N}\left[\begin{matrix} \hat{\mathbf{u}}_{1,k} \\ \hat{\mathbf{u}}_{2,k} \end{matrix}, (\mathbf{Z}'_k \mathbf{R}^{-1} \mathbf{Z}_k + \mathbf{A}^{-1}_{k,k} \otimes \mathbf{G}^{-1})^{-1} \right] \quad [12.24]$$

where:

$$\begin{pmatrix} \hat{\mathbf{u}}_{1,k} \\ \hat{\mathbf{u}}_{2,k} \end{pmatrix} = (\mathbf{Z}'_k \mathbf{R}^{-1} \mathbf{Z}_k + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1})^{-1}$$
$$\{(\mathbf{Z}'_k \mathbf{R}^{-1} (\mathbf{y}_k - \mathbf{X}\mathbf{b}) - \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} (\hat{\mathbf{u}}_s + \hat{\mathbf{u}}_d)\}$$

where s and d are the sire and dam of the kth animal.

From [12.20], the full conditional distribution of the residual variance is:

$$P(\mathbf{R} | \mathbf{b}, \mathbf{u}, \mathbf{y}) \propto P(\mathbf{R}) P(\mathbf{y} | \mathbf{b}, \mathbf{u}, \mathbf{R})$$

Including the prior distribution, the above can be expressed (Jensen *et al.*, 1994) as:

$$P(\mathbf{R} | \mathbf{b}, \mathbf{u}, \mathbf{y}) \propto |\mathbf{R}|^{-\frac{1}{2}(v_e + p + 1 + m)} \exp[-\tfrac{1}{2} \mathrm{tr}\{\mathbf{R}^{-1} (\mathbf{S}^2_e + \mathbf{V}^{-1}_e)\}]$$

where m is the number of records and $\mathbf{S}^2_e$ is:

$$\mathbf{S}^2_e = \begin{pmatrix} \hat{\mathbf{e}}'_1 \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}'_1 \hat{\mathbf{e}}_2 \\ \hat{\mathbf{e}}'_2 \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}'_2 \hat{\mathbf{e}}_2 \end{pmatrix}$$

assuming that $n = 2$ and $\hat{\mathbf{e}} = \mathbf{y}_i - \mathbf{X}_i \mathbf{b}_i - \mathbf{Z}_i \mathbf{u}_i$, $i = 1, n$.

Thus:

$$\mathbf{R} | \mathbf{b}, \mathbf{u}, \mathbf{y} \sim \mathrm{IW}_2((\mathbf{S}^2_e + \mathbf{V}^{-1}_e)^{-1}, \mathbf{v}_e + m) \quad [12.25]$$

which is in the form of a p-dimensional inverted Wishart distribution, with $\mathbf{v}_e + m$ degrees of freedom and scale parameter $(\mathbf{S}^2_e + \mathbf{V}^{-1}_e)$.

Similarly, the conditional distribution for the additive genetic variance is:

$$P(\mathbf{G}|\mathbf{b},\mathbf{u},\mathbf{y}) \propto P(\mathbf{G})P(\mathbf{u}|\mathbf{G})$$

Including the prior distribution, the above can be expressed (Jensen *et al.*, 1994) as:

$$P(\mathbf{G}|\mathbf{b},\mathbf{u},\mathbf{y}) \propto |\mathbf{G}|^{-\frac{1}{2}(v_u+p+1+q)} \exp[-\tfrac{1}{2}\,\mathrm{tr}\{\mathbf{G}^{-1}(\mathbf{S}_u^2 + \mathbf{V}_u^{-1})\}]$$

where q is the number of animals and, assuming $n = 2$, $\mathbf{S}_u^2$ is:

$$\mathbf{S}_u^2 = \begin{pmatrix} \mathbf{u}_1'\mathbf{A}^{-1}\mathbf{u}_1 & \mathbf{u}_1'\mathbf{A}^{-1}\mathbf{u}_2 \\ \mathbf{u}_2'\mathbf{A}^{-1}\mathbf{u}_1 & \mathbf{u}_2'\mathbf{A}^{-1}\mathbf{u}_2 \end{pmatrix}$$

Thus:

$$\mathbf{G}\,|\mathbf{b},\mathbf{u},\mathbf{y} \sim \mathrm{IW}_2((\mathbf{S}_u^2 + \mathbf{V}_u^{-1})^{-1}, \mathbf{v}_u + q) \qquad [12.26]$$

which again is in the form of a p-dimensional inverted Wishart distribution with $\mathbf{v}_u + q$ degrees of freedom and scale parameter $(\mathbf{S}_u^2 + \mathbf{V}_u^{-1})$.

12.3.3 Numerical illustration

Example 12.2

Using the data in Example 5.1 and the variance components, the application of Gibbs sampling for estimation of variance components and the prediction of breeding values are illustrated. Uniform priors are assumed for the variance components such that $\mathbf{v}_e = \mathbf{v}_u = -3$ and $\mathbf{V}_e = \mathbf{V}_u = 0$. A flat prior is assumed for **b**, and **u** is assumed to be normally distributed.

Process data and accumulate right-hand side (rhs) and diagonals (Diag) for level j of sex of calf effects as:

$$\mathrm{rhs}_{1j} = \mathrm{rhs}_{1j} + \mathbf{R}^{11}(\mathbf{y}_1 - \hat{\mathbf{u}}_{1i}) + \mathbf{R}^{12}(\mathbf{y}_2 - \hat{\mathbf{u}}_{2i})$$
$$\mathrm{rhs}_{2j} = \mathrm{rhs}_{2j} + \mathbf{R}^{21}(\mathbf{y}_2 - \hat{\mathbf{u}}_{1i}) + \mathbf{R}^{22}(\mathbf{y}_2 - \hat{\mathbf{u}}_{2i})$$
$$\mathbf{Diag}_j = \mathbf{Diag}_j + \mathbf{R}$$

When all data have been read, calculate solutions for level j of sex effect as:

$$\begin{pmatrix} \hat{b}_{1j} \\ \hat{b}_{2j} \end{pmatrix} = \mathbf{Diag}_j^{-1} \begin{pmatrix} \mathrm{rhs}_{1j} \\ \mathrm{rhs}_{2j} \end{pmatrix}$$

Sample $\mathbf{b}_j$ in [12.23] as:

$$\mathbf{b}_j = \begin{pmatrix} \hat{b}_{1j} \\ \hat{b}_{2j} \end{pmatrix} + \{\mathrm{CHOL}(\mathbf{Diag}_j^{-1})\}\mathbf{h}$$

where **h** is the vector of normal deviates from a population of mean zero and variance 1 and CHOL is the Cholesky decomposition of the inverse of the matrix **Diag**.

Next, process data and accumulate right-hand side (rhs) and diagonals (Diag) for animal *i* as:

$$\text{rhs}_{1i} = \text{rhs}_{1i} + \mathbf{R}^{11}(\mathbf{y}_1 - \hat{\mathbf{b}}_{1j}) + \mathbf{R}^{12}(y_2 - \hat{\mathbf{b}}_{2j})$$
$$\text{rhs}_{2i} = \text{rhs}_{2i} + \mathbf{R}^{21}(\mathbf{y}_2 - \hat{\mathbf{b}}_{1j}) + \mathbf{R}^{22}(y_2 - \mathbf{b}_{2j})$$
$$\mathbf{Diag}_i = \mathbf{Diag}_i + \mathbf{R}$$

When all data have been read, calculate solutions for animal *i* as:

$$\begin{pmatrix} \hat{u}_{1i} \\ \hat{u}_{2i} \end{pmatrix} = \mathbf{Diag}_i^{-1} \begin{pmatrix} \text{rhs}_{1i} \\ \text{rhs}_{2i} \end{pmatrix}$$

Sample $\mathbf{u}_i$ in [12.24] as:

$$\mathbf{u}_i = \begin{pmatrix} \hat{u}_{1i} \\ \hat{u}_{2i} \end{pmatrix} + \{\text{CHOL}(\mathbf{Diag}_i^{-1})\}\mathbf{h}$$

All data are then processed to obtain residual effects as:

$$\hat{\mathbf{e}} = \begin{pmatrix} \hat{\mathbf{e}}_1 \\ \hat{\mathbf{e}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 - \mathbf{X}_1\hat{\mathbf{b}}_1 - \mathbf{Z}_1\hat{\mathbf{u}}_1 \\ \mathbf{y}_2 - \mathbf{X}_2\hat{\mathbf{b}}_2 - \mathbf{Z}_2\hat{\mathbf{u}}_2 \end{pmatrix}$$

and calculate residual sums of squares, $\mathbf{S}_e^2 = \hat{\mathbf{e}}\hat{\mathbf{e}}'$. Then compute $\mathbf{T} = (\mathbf{S}_e^2 + \mathbf{V}_e^{-1})^{-1}$. Cholesky decomposition of **T** is carried out to obtain **LL′**, where **L** is a lower triangular matrix. Sampling from a Wishart distribution, with **L** as the input matrix and $\mathbf{v}_e + m$ degrees of freedom [12.25], generates a new sample value of **R**.

Similarly, to compute a new sample value of **G** using [12.26], first compute $\mathbf{T}^{-1} = (\mathbf{S}_u^2 + \mathbf{V}_u^{-1})^{-1}$. Decompose **T** to obtain **LL′** and sample from a Wishart distribution with **L** as the input matrix and $\mathbf{v}_u + q$ degrees of freedom. Another cycle of sampling is then initiated until the desired length of chain is achieved. Post-processing of results can be carried out, as discussed in Section 12.2.3.

13 Solving Linear Equations

Different methods can be utilized to solve the mixed model equations (MME) covered in the previous chapters. These various methods could broadly be divided into three main categories:

1. Direct inversion (Section 13.1).
2. Iteration on the MME (Section 13.2).
3. Iteration on the data (Section 13.3).

The manner in which the MME are set up depends on the method to be utilized in solving these equations. As shown later in Section 13.3 below, the third method, for instance, does not involve setting up the MME directly.

13.1 Direct Inversion

The solutions to the MME in the various examples given so far in this book have been based on this method. It involves setting up the MME and inverting the coefficient matrix. Solutions are obtained by multiplying the right-hand side (RHS) by the inverse of the coefficient matrix. Thus **b**, the vector of solution, is calculated as:

$$\hat{\mathbf{b}} = \mathbf{C}^{-1}\mathbf{y}$$

where **C** is the coefficient matrix and **y** is the RHS. Since the coefficient matrix is symmetric, only the upper triangular portion is usually set up and inverted. The major limitation of this approach is that it can only be applied to small data sets in view of the memory requirements and computational difficulties of inverting large matrices.

13.2 Iteration on the Mixed Model Equations

This involves setting up the MME and iterating on these equations until convergence is achieved at a predetermined criterion. The iterative procedures are based on the general theory for solving simultaneous equations. For instance, given two simultaneous equations with unknown parameters, b_1 and b_2, the first equation can be solved for b_1 in terms of b_2. This value of b_1 can then be substituted in the second equation to solve for b_2. The value of b_2 is then substituted in the first equation to calculate b_1. This is the principle upon which the iterative procedures are based. In the iterative procedure, the above process is continued until the solutions for the **b** terms are more or less the same in each round of iteration and the equations are said to have converged. There are various iterative procedures that can be utilized.

13.2.1 Jacobi iteration

One of the simplest methods is Jacobi iteration or total step iteration.

Consider the following set of simultaneous equations:

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

These equations can also be written as:

$$\begin{aligned} c_{11}b_1 + c_{12}b_2 + c_{13}b_3 &= y_1 \\ c_{21}b_1 + c_{22}b_2 + c_{23}b_3 &= y_2 \\ c_{31}b_1 + c_{32}b_2 + c_{33}b_3 &= y_3 \end{aligned}$$

or as:

$$\mathbf{Cb} = \mathbf{y} \qquad [13.1]$$

The system of equations is rearranged so that the first is solved for b_1, the second for b_2 and the third for b_3. Thus:

$$\begin{aligned} b_1^{r+1} &= (1/c_{11})(y_1 - c_{12}b_2^r - c_{13}b_3^r) \\ b_2^{r+1} &= (1/c_{22})(y_2 - c_{21}b_1^r - c_{23}b_3^r) \\ b_3^{r+1} &= (1/c_{33})(y_3 - c_{31}b_1^r - c_{32}b_2^r) \end{aligned} \qquad [13.2]$$

The superscript r refers to the number of the round of iteration. In the first round of iteration, r equals 1 and b_1 to b_3 could be set to zero or an assumed set of values which are used to solve the equations to obtain a new set of solutions (**b** terms). The process is continued until two successive sets of solutions are within previously defined allowable deviations and the equations are said to converge. One commonly used convergence criterion is the sum of squares of differences between the current and previous solutions divided by the sum of squares of the current solution.

Once this is lower than a predetermined value, for instance 10^{-9}, the equations are considered to have converged.

From the set of equations above, the solution for b_i was obtained by dividing the adjusted right-hand side by the diagonal (a_{ii}). It is therefore mandatory that the diagonal element, often called the pivot element, is not zero. If a zero pivot element is encountered during the iterative process, the row containing the zero should be exchanged with a row below it in which the element in that column is not zero. To avoid the problem of encountering a zero pivot element and generally to improve the efficiency of the iterative process, it is sometimes recommended that the system of equations should be ordered such that the coefficient of b_1 of the greatest magnitude occurs in the first equation, the coefficient of b_2 of the greatest magnitude in the remaining equations occurs in the second equation, etc.

The iterative procedure described above is usually termed Jacobi iteration, in which all new solutions in the current (r) round of iteration are obtained using solutions only from the previous ($r-1$) round of iteration. The Jacobi iterative procedure is inefficient in handling systems of equations which are not constrained (that is, with no restrictions placed on the solutions for the levels of an effect) and convergence is not guaranteed (Maron, 1987; Misztal and Gianola, 1988). When random animal effect is involved in the system of equations with relationships included, it is usually necessary to use a relaxation factor of below 1.0; otherwise equations may not converge (Groeneveld, 1990). The relaxation factor refers to a constant, estimated on the basis of the linear changes in the solutions during the iteration process and applied to speed up the solutions towards convergence. When iterating on the data (Section 13.3), the Jacobi iterative procedure involves reading only one data file, even with several effects in the model. With large data sets this has the advantage of reducing memory requirement and processing time compared with the Gauss–Seidel iterative procedure (see Section 13.2.2).

The Jacobi iterative procedure can be briefly summarized as follows. Following Ducrocq (1992), equation [13.1] can be written as:

$$[\mathbf{M} + (\mathbf{C} - \mathbf{M})]\mathbf{b} = \mathbf{y}$$

if $\mathbf{M}$ is the diagonal matrix containing the diagonal elements of $\mathbf{C}$; then the algorithm for Jacobi iteration is:

$$\mathbf{b}^{(r+1)} = \mathbf{M}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{b}^{(r)}) + \mathbf{b}^{(r)} \qquad [13.3]$$

When a relaxation factor is applied, the above equation becomes:

$$\mathbf{b}^{(r+1)} = w[\mathbf{M}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{b}^{(r)})] + \mathbf{b}^{(r)}$$

Another variation of Jacobi iteration, called second-order Jacobi, is usually employed in the analysis of large data sets and it can increase the rate of convergence. The iterative procedure for second-order Jacobi is:

$$\mathbf{b}^{(r+1)} = \mathbf{M}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{b}^{(r)} + \mathbf{b}^{(r)} + w(\mathbf{b}^{(r)} - \mathbf{b}^{(r-1)}))$$

Example 13.1

Using the coefficient matrix and the right-hand side for Example 3.1, Jacobi iteration [13.2] is carried out using only the non-zero element of the coefficient matrix. Solutions for sex effect (**b** vector) and random animal effect (**u** vector) are shown below with the round of iteration. The convergence criterion (CONV) was the sum of squares of differences between the current and previous solutions divided by the sum of squares of the current solution.

	Rounds of iteration									
Effects	0[a]	1	2	3	4	16	17	18	19	20
$\hat{b}_1$	4.333	4.333	4.381	4.370	4.368	4.358	4.358	4.358	4.358	4.358
$\hat{b}_2$	3.400	3.400	3.433	3.365	3.414	3.404	3.404	3.404	3.404	3.404
$\hat{u}_1$	0.000	0.267	0.164	0.185	0.131	0.099	0.099	0.099	0.099	0.099
$\hat{u}_2$	0.000	0.000	–0.073	–0.003	–0.039	–0.018	–0.018	–0.018	–0.018	–0.018
$\hat{u}_3$	0.000	–0.033	–0.080	–0.049	–0.070	–0.041	–0.041	–0.041	–0.041	–0.041
$\hat{u}_4$	0.167	–0.138	–0.007	–0.035	0.000	–0.008	–0.008	–0.008	–0.008	–0.008
$\hat{u}_5$	–0.500	–0.411	–0.248	–0.265	–0.204	–0.185	–0.185	–0.185	–0.185	–0.185
$\hat{u}_6$	0.500	0.345	0.318	0.237	0.236	0.178	0.178	0.178	0.177	0.177
$\hat{u}_7$	–0.833	–0.406	–0.390	–0.301	–0.295	–0.249	–0.249	–0.249	–0.249	–0.249
$\hat{u}_8$	0.667	0.400	0.286	0.232	0.207	0.183	0.183	0.183	0.183	0.183
CONV	1.000	2.3^{-2}	3.9^{-3}	1.4^{-3}	5.9^{-4}	4.2^{-8}	1.6^{-8}	1.0^{-8}	4.1^{-9}	3.0^{-9}

[a] Starting values.

The starting solutions for sex effect were the mean yield for each sex subclass and, for animals with records, starting solutions were the deviation of their yields from the mean yield of their respective sex subclass and zero for ancestors. The final solutions obtained after the 20th round of iteration were exactly the same as those obtained in Section 3.1 by direct inversion of the coefficient matrix. The solutions for sex effect were obtained using [13.2]. Thus, in the first round of iteration the solution for males was:

$$\hat{b}_1 = \frac{1}{c_{11}}\left(\sum_{k=1}^{m} y_k - (1)\hat{u}_4 - (1)\hat{u}_7 - (1)\hat{u}_8\right)$$

where c_{ii} is the diagonal element of the coefficient matrix for level i of sex effect and m is the number of records for males.

$$b_1 = 1/3(13.0 - 0.167 - (-0.833) - 0.667) = 4.333$$

However, using [13.2] to obtain animal solutions caused the system of equations to diverge. A relaxation factor (w) of 0.8 was therefore employed and solutions for animal j were computed as:

$$\hat{u}_j^r = w\left[\left(\frac{1}{c_{ll}}\right)\left(y_j - c_{li}\hat{b}_i^{r-1} - \sum_k c_{lt}\hat{u}_k^{r-1}\right) - \hat{u}_j^{r-1}\right] + \hat{u}_j^{r-1}$$

where $l = j + n$, $t = k + n$, with $n = 2$; the total number of levels of fixed effect, c_{lt} and c_{li}, for instance, are the elements of the coefficient matrix between animals j and k, and animal j and level i of sex effect, respectively. Thus, in the first round of iteration, solutions for animals 1 and 8 are calculated as:

$$\begin{aligned}\hat{u}_1^1 &= w[\{1/c_{33}(y_1 - (1)\hat{u}_2^0 - (-1.333)\hat{u}_4^0 - (-2)\hat{u}_6^0)\} - \hat{u}_1^0] + \hat{u}_1^0 \\ &= w[\{1/3.667(0 - 0 - (-0.223 - (-1))\} - 0] + 0 \\ &= 0.8(0.334 - 0) + 0 = 0.267\end{aligned}$$

and:

$$\begin{aligned}\hat{u}_8^1 &= w[\{1/c_{1010}(y_8 - (1)b_1^0 - (-2)\hat{u}_3^1 - (-2)\hat{u}_6^1)\} - \hat{u}_8^0] + \hat{u}_8^0 \\ &= w[\{1/5(5 - 4.333 - 0 - (-1)\} - 0.667] + 0.667 \\ &= 0.8(0.333 - 0.667) + 0.667 = 0.400\end{aligned}$$

13.2.2 Gauss–Seidel iteration

Another iterative procedure commonly used is Gauss–Seidel iteration. This is similar to Jacobi iteration except that most current solutions are calculated from the most recent available solution rather than the solution from the previous round of iteration. Using the same set of simultaneous equations as in equation [13.1], solutions for b_1, b_2 and b_3 in the first round of iteration become:

$$\begin{aligned}b_1^{r+1} &= (1/c_{11})(y_1 - c_{12}b_2^r - c_{13}b_3^r) \\ b_2^{r+1} &= (1/c_{22})(y_2 - c_{21}b_1^{r+1} - c_{23}b_3^r) \\ b_3^{r+1} &= (1/c_{33})(y_3 - c_{31}b_1^{r+1} - c_{32}b_2^{r+1})\end{aligned} \qquad [13.4]$$

Thus the solution for b_2 in the $r + 1$ round of iteration is calculated using the most recent solution for b_1 (b_1^{r+1}) instead of the previous solution (b_1^r), and the current solution for b_3 is calculated from the current solutions for b_1 (b_1^{r+1}) and b_2 (b_2^{r+1}). If, in equation [13.3], **L** is strictly the lower triangular matrix of **C** and **D** is a diagonal matrix of **C**, then equation [13.3] becomes the Gauss–Seidel iteration when $\mathbf{M} = \mathbf{L} + \mathbf{D}$. The convergence criteria could equally be defined as discussed in Section 13.2.1. Generally, equations are guaranteed to converge with the Gauss–Seidel iterative procedure. However, when iterating on the data, this iterative procedure involves reading one data file for each effect in the model. With large data sets, the setting up of data files for each effect could result in a large memory requirement and the reading of several files in each round of iteration could increase processing time.

Example 13.2

Using the same coefficient matrix, right-hand side and starting values as in Example 13.1 above, the Gauss–Seidel iteration [13.4] is carried out for the same number of iterations as in Jacobi's method and the results are shown below. The convergence criterion is as defined in Example 13.1.

	Rounds of iteration									
Effects	0	1	2	3	4	16	17	18	19	20
b_1	4.333	4.333	4.400	4.372	4.364	4.359	4.359	4.359	4.359	4.359
b_2	3.400	3.400	3.392	3.403	3.407	3.405	3.405	3.405	3.405	3.405
$\hat{u}_1$	0.000	0.333	0.194	0.149	0.115	0.098	0.098	0.098	0.098	0.098
$\hat{u}_2$	0.000	–0.083	–0.035	–0.006	–0.008	–0.019	–0.019	–0.019	–0.019	–0.019
$\hat{u}_3$	0.000	–0.021	–0.136	–0.109	–0.076	–0.041	–0.041	–0.041	–0.041	–0.041
$\hat{u}_4$	0.167	–0.119	0.001	0.004	–0.003	–0.009	–0.009	–0.009	–0.009	–0.009
$\hat{u}_5$	–0.500	–0.376	–0.261	–0.218	–0.199	–0.186	–0.186	–0.186	–0.186	–0.186
$\hat{u}_6$	0.500	0.392	0.254	0.204	0.185	0.177	0.177	0.177	0.177	0.177
$\hat{u}_7$	–0.833	–0.364	–0.284	–0.260	–0.253	–0.250	–0.250	–0.250	–0.250	–0.250
$\hat{u}_8$	0.667	0.282	0.167	0.164	0.171	0.182	0.183	0.183	0.183	0.183
CONV	1.000	1.9^{-2}	3.4^{-3}	3.1^{-4}	1.0^{-4}	7.0^{-10}	4.0^{-10}	2.0^{-10}	1.0^{-10}	8.0^{-11}

CONV, convergence criterion.

The solutions obtained are the same as those obtained from Jacobi iteration and by direct inversion of the coefficient matrix in Example 3.1. In addition, the equations converged faster than when using Jacobi iteration and no relaxation factor was applied.

Iterating on the MME equations could be carried out as described above, once the equations have been set up, using only the stored non-zero elements of the coefficient matrix. In practice, it may be necessary to store the non-zero elements and their rows and columns on disk for large data sets because of the memory requirement, and these are read in each round of iteration.

13.3 Iterating on the Data

This is the most commonly used methodology in national genetic evaluations, which usually involve millions of records. Schaeffer and Kennedy first presented this method in 1986. It does not involve setting up the coefficient matrix directly, but it involves setting up equations for each level of effects in the model as the data and pedigree files are read and solved using either Gauss–Seidel or Jacobi iteration or a combination of both or a variation of any of the iterative procedures, such as second-order Jacobi. Presented below are the basic equations for the solutions of various effects under several models and these form the basis of the iterative process for each of the models.

The equation for the solution of level i for a fixed effect in the model in a univariate animal situation is equation [3.5], which is derived from the MME and can be generalized as:

$$\hat{b}_i = \frac{\sum_{k=1}^{n_i} y_{ki} - \sum_{j=1}^{m} \hat{w}_{ij}}{n_i} \qquad [13.5]$$

where y_{ki} is the kth record in level i, m is the total number of levels of other effects within subclass i of the fixed effect and $\hat{w}_{ij}$ is the solution for the jth level, and n_i is the number of records in fixed effect subclass i. However, when there are many fixed effects in the model, the above formula may be used to obtain solutions for the major fixed effect with many levels, such as herd–year–season, while the vector of solutions (**f**) for other minor fixed effects with few levels may be calculated as:

$$\mathbf{f} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \hat{\mathbf{w}} - \mathbf{b}) \qquad [13.6]$$

where **y** is the vector of observations, $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse of the coefficient matrix for the minor fixed effects, and $\hat{\mathbf{w}}$ and $\hat{\mathbf{b}}$ are vectors of solutions for effects as defined in [13.5]. The matrix $\mathbf{X}'\mathbf{X}$ could be set up in the first round of iteration and stored in the memory for use in subsequent rounds of iterations.

The solution ($\hat{u}$) for the level j (animal j) of the random animal effect in the univariate animal model is calculated using equation [3.8], which can be rewritten (replacing n_3 by k) as:

$$\hat{u}_j = \left[n_1\alpha(\hat{u}_s + \hat{u}_d) + n_2 yd + \sum_o \{k_o\alpha(\hat{u}_o - 0.5(\hat{u}_{mo}))\} \right] \Big/ \text{diag}_j \qquad [13.7]$$

with:

$$\text{diag}_j = 2(n_1)\alpha + n_2 + \sum_o \{(k_o/2)\alpha\}$$

where $\hat{u}_s$, $\hat{u}_d$ and $\hat{u}_o$ are solutions or estimated breeding values for the sire, dam and oth progeny of animal j, respectively, $\hat{u}_{mo}$ is the solution of the mate of animal j with respect to progeny o, yd is yield deviation, that is, yield of animal j corrected for all other effects in the model, $n_1 = 1$ or 2/3 if both or one parent of animal j is known, n_2 is the number of records, k_o is 1 or 2/3 if the other parent of progeny o (mate of animal j) is known or not known and $\alpha = \sigma_e^2/\sigma_a^2$.

In the multivariate animal model situation with equal design and random animal effect as the only random effect in addition to residual effects, the solutions for the levels of fixed effect and animal effects are obtained using equations [5.4] and [5.8], respectively, which are derived from the mixed model equations [5.3].

For maternal animal model equations, the solutions for fixed effects could be calculated using equations [6.3]. The equations for animal and genetic maternal effects are based on equation [6.4], given earlier. From equation [6.4], the solution ($\hat{u}$) for direct effect for animal i is:

$$\begin{aligned}\hat{u}_i = \Big[& n_1\alpha_1(\hat{u}_s + \hat{u}_d) + n_1\alpha_2(\hat{m}_s + \hat{m}_d) - n_4\alpha_2(\hat{m}_i) - (k_o/2)\alpha_2(\hat{m}_i) \\ & + n_2(y_i - b_j - \hat{m}_d - \hat{p}e_d) + \sum_o \{k_o\alpha_1(\hat{u}_o - 0.5(\hat{u}_{mo}))\} \\ & + \sum_o k_o\alpha_2(\hat{m}_o - 0.5(\hat{m}_{mo}))\} \Big] / \text{diag}_i \end{aligned} \qquad [13.8]$$

with:

$$\text{diag}_i = 2(n_1)\alpha_1 + n_2 + \sum_o \{(k_o/2)\alpha_1\}$$

where $\hat{m}_i$, $\hat{m}_s$, $\hat{m}_d$, $\hat{m}_o$ and $\hat{m}_{mo}$ are solutions for genetic maternal effects for animal i, sire, dam and oth progeny of animal i, respectively, y_i is the yield for animal i, $\hat{b}_j$ is the solution for fixed effect j, $\hat{pe}_d$ is the permanent environmental effect for the dam of animal i, n_1, n_2 and k_o are as defined above, $n_4 = 2(n_1)$ and α terms are as defined in equation [6.4].

The solution ($\hat{m}$) for genetic maternal effect for animal i from [6.4] is:

$$\begin{aligned} m_i = \Big[& n_1\alpha_2(\hat{u}_s + \hat{u}_d) + n_1\alpha_3(\hat{m}_s + \hat{m}_d) - n_4\alpha_2(\hat{u}_i) - (k_o/2)\alpha_2(\hat{u}_i) \\ & + n_2(y_i - \hat{b}_j - \hat{u}_i - \hat{m}_d - \hat{pe}_d) + \sum_o \{k_o\alpha_2(\hat{u}_o - 0.5(\hat{u}_{mo}))\} \\ & + \sum_o \{k_o\alpha_3(\hat{m}_o - 0.5(\hat{m}_{mate}))\} \Big] / \text{diag}_i \end{aligned} \qquad [13.9]$$

with:

$$\text{diag}_i = 2(n_1)\alpha_3 + n_2 + \sum_o \{k_o/2)\alpha_3\}$$

Solutions for permanent environmental effect are obtained using equation [6.5].

The computational procedure for a reduced animal model was presented by Schaeffer and Wilton (1987) using a bivariate analysis. The procedure is similar to the animal model described above except that records for non-parents are written twice, one record for each parent. Consequently the residual variance of non-parental records (r_2) is multiplied by 2, that is:

$$r_2 = 2(\sigma_e^2 + d(\sigma_a^2)) = 2(1 + d\alpha^{-1})\sigma_e^2$$

where $d = \frac{1}{2}$ or $\frac{3}{4}$ if both or one parent is known and the contribution of non-parents' records to the diagonal of their parents is 0.5 instead of 0.25 (see Example 6.2).

The equations for solutions for levels of fixed and random effects are similar to those defined earlier. From [13.5], if the residual variance for parental records is defined as r_1, the contribution of parental records to the RHS for level i of a major fixed effect is:

$$\text{RHS}_i = \sum_{k=1}^{n_i} (r_1^{-1}(y_{ik} - \hat{w}_{kj})) \qquad [13.10]$$

where n_i is the number of parental records in level i of fixed effect and $\hat{w}_{kj}$ is the solution for the jth level of other effects in the model affecting record k. The contribution of non-parental records to the RHS is included as:

$$\text{RHS}_i = \text{RHS}_i + \sum_{k=1}^{m_l} (r_{2k}^{-1}(y_{ki} - 0.5(\hat{u}_s + \hat{u}_d) - \hat{w}_{kj})) \qquad [13.11]$$

where m_i is the number of non-parental records in level i of fixed effect, $\hat{u}_s$ and $\hat{u}_d$ are solutions for the sire and dam of the non-parent with record k, r_{2k}^{-1} is the inverse of the residual variance for the non-parental record k and $\hat{w}_{kj}$ is the solution for level j of other effects in the model apart from random animal effects affecting record k. Then:

$$\hat{b}_i = \frac{\text{RHS}_i}{\sum_{k=1}^{n_i} r_1^{-1} + \sum_{j=1}^{m_i} r_2^{-1}}$$

The equation for the breeding value of the jth animal, which is a parent with its own yield record, a non-parental record from progeny i and information from another progeny (o), which is itself a parent, is:

$$\hat{u}_j = \Big[n_1\alpha(\hat{u}_s + \hat{u}_d) + n_2 r_1^{-1}(yd_j) + n_3 r_2^{-1}(yd_i - (0.5)\hat{u}_{mi}) + \sum_o \{k_o\alpha(\hat{u}_o - 0.5(\hat{u}_{mo}))\}\Big] / \text{diag}_j \qquad [13.12]$$

with:

$$\text{diag}_j = 2(n_1)\alpha + n_2 r_1^{-1} + (0.5)n_3 r_2^{-1} + \sum_o \{(k_o/2)\alpha\}$$

where yd_j and yd_i are yield deviations for animal j and progeny i, which is a non-parent, $\hat{u}_{mi}$ is the breeding value for the mate of animal j with respect to the ith progeny (non-parent), n_2 is the number of observations (records) on animal j, n_3 is the number of non-parental records, r_1^{-1} and r_2^{-1} are as defined earlier and all other terms are as defined in equation [13.7]. Note that contributions from the oth progeny in the above equation refer to those progeny of animal j which are themselves parents and that non-parental records are adjusted for half the breeding value of the mate of animal j. If animal j has no non-parental records from its progeny, equation [13.12] is the same as [13.7].

The principles of evaluation based on iterating on the data are illustrated below using a univariate animal model and a reduced animal model with maternal effects.

13.3.1 Animal model without groups

Example 13.3

Using the same data as in Example 3.1 (Table 3.1) on the weaning weight of beef calves, parameters and model, the principles of predicting breeding values and estimating solutions for fixed effects iterating on the data are illustrated using Gauss–Seidel iteration.

Data arrangement

Gauss–Seidel iteration requires the data files to be sorted by the effect to be solved for. The pedigree file is needed when solving for animal solutions.

The pedigree file is created and ordered in such a manner that contributions to the diagonal and right-hand side of an animal from the pedigree, due to the number of parents known (see type 1 record below) and from progeny accounting for whether the mate is known (type 2 record), can be accumulated while processing the animal. Thus, initially, a pedigree file is created, consisting of two types of records:

1. Type 1 record for all animals in the data comprising the animal identity, record type and sire and dam identities.
2. Type 2 record for each parent in the data comprising the parent identity, record type, identities for progeny and other parent (mate) if known. The type 2 records are used for adjusting the contribution of the progeny to each parent for the mate's breeding value when solving for animal solutions.

The pedigree file is sorted by animal and record type. The sorted pedigree file for the example data is given below.

Animal	Code	Sire or progeny	Dam or mate
1	1	0	0
1	2	4	0
1	2	6	2
2	1	0	0
2	2	5	3
2	2	6	1
3	1	0	0
3	2	5	2
3	2	8	6
4	1	1	0
4	2	7	5
5	1	3	2
5	2	7	4
6	1	1	2
6	2	8	3
7	1	4	5
8	1	3	6

Secondly, a data file is set up consisting of animal identity, fixed effects, covariates and traits. If there is a major fixed effect with many levels, two data files need to be set up, one sorted by the major fixed effects, such as herd or herd–year–season (file A), to be used when solving for the major fixed effect, and the other sorted by animal identity (file B), to be used to solve for animal solutions. Assuming sex effect to be the

major fixed effect in the example data, the data sorted by sex are as follows:

Calf	Sex[a]	Weaning weight
4	1	4.5
7	1	3.5
8	1	5.0
5	2	2.9
6	2	3.9

[a]1 = male, 2 = female.

ITERATION STAGE

Let $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ be vectors of solutions for sex and animal effects. Starting values for sex and animal effects are assumed to be the same as in Example 13.1.

SOLVING FOR FIXED EFFECTS. In each round of iteration, file A is read one level of sex effect at a time with adjusted right-hand sides (ARHS) and diagonals (DIAG) accumulated for the *i*th level as:

$$\text{ARHS}_i = \text{ARHS}_i + y_{ik} - \hat{u}_k$$
$$\text{DIAG}_i = \text{DIAG}_i + 1$$

At the end of the *i*th level, the solution for the level is computed as:

$$\hat{b}_i = \text{ARHS}_i / \text{DIAG}_i$$

The above step essentially involves adjusting the yields for animal effects using previous solutions and calculating solutions for each level of sex effect. For example, the solution for level one of the sex effect in the first round of iteration is:

$$\hat{b}_1 = [(4.5 - 0.167) + (3.5 - (-0.833)) + (5.0 - 0.667)]/3 = 4.333$$

After calculating solutions for fixed effect in the current round of iteration, file B and the pedigree file are processed to compute animal solutions.

SOLVING FOR ANIMAL SOLUTIONS. Diagonals (DIAG) and adjusted right-hand sides (ARHS) are accumulated as data for each animal are read from the pedigree file or from both pedigree file and file B for animals with records. When processing type 1 records in the pedigree file for the *k*th animal, the contribution to the DIAG and ARHS according to the number of parents known is as follows:

Number of parents known		
None	One (sire(*s*))	Both
$ARHS_k = 0$ $DIAG_k = \alpha$	$ARHS_k = (\frac{2}{3})\,\alpha\,(\hat{u}_s)$ $DIAG_k = (\frac{4}{3})\,\alpha$	$ARHS_k = \alpha\,(\hat{u}_s + \hat{u}_d)$ $DIAG_k = 2\alpha$

where $\hat{u}_s$ and $\hat{u}_d$ are current solutions for the sire and dam, respectively.

When processing type 2 records in the pedigree file for the *k*th animal, the contribution to the DIAG and ARHS according to whether the mate of animal *k* is known or not is as follows:

Mate unknown	Mate known
$ARHS_k = ARHS_k + (\frac{2}{3})\,\alpha(\hat{u}_o)$ $DIAG_k = DIAG_k + (\frac{1}{3})\,\alpha$	$ARHS_k = ARHS_k + \alpha(\hat{u}_o - 0.5\hat{u}_m)$ $DIAG_k = DIAG_k + (\frac{1}{2})\,\alpha$

where $\hat{u}_o$ and $\hat{u}_m$ are current solutions for the progeny and mate, respectively, of the *k*th animal. If the *k*th animal has a yield record:

$$ARHS_k = ARHS_k + y_{ik} - \hat{b}_i$$
$$DIAG_k = DIAG_k + 1$$

where $\hat{b}_i$ are current solutions for level *i* of sex effect.

When all pedigree and yield records for the *k*th animal have been processed, the solution for the animal is computed as:

$$\hat{u}_k = ARHS_k / DIAG_k$$

For the example data, the solution for animal 5 in the first round of iteration is computed as follows.

Contribution to diagonal from pedigree is:

$$DIAG_5 = (2 + 0.5)\alpha = 5.00$$

Accounting for yield record, diagonal becomes:

$$DIAG_5 = 5.00 + 1 = 6.00$$

Contribution to right-hand side from yield is:

$$ARHS_5 = 2.9 - 3.40 = -0.5$$

Contribution to RHS from parents and progeny (pedigree) is:

$$\begin{aligned} ARHS_5 &= ARHS_5 + \alpha(\hat{u}_2 + \hat{u}_3) + \alpha(\hat{u}_7 - 0.5(\hat{u}_4)) \\ &= -0.5 + 2(-0.083 + (-0.021)) + 2(-0.833 - 0.5(-0.119)) \\ &= -2.255 \end{aligned}$$

and:

$\hat{u}_5 = -2.255/6.00 = -0.376$

When all animals have been processed, then the current round of iteration is completed. However, the iteration process is continued for sex and animal effects until convergence is achieved. The convergence criterion can be defined as in Section 13.2. In this example, solutions were said to have converged when the sum of squares of differences between the current and previous solutions divided by the sum of squares of the current solution was less than 10^{-7}. The solutions for all effects in the first round of iteration and at convergence at the 20th iteration are as follows:

	Solutions	
Effects	At round 1	At convergence
Sex[a]		
1	4.333	4.359
2	3.400	3.404
Animal		
1	0.333	0.098
2	–0.083	–0.019
3	–0.021	–0.041
4	–0.119	–0.009
5	–0.376	–0.186
6	0.392	0.177
7	–0.364	–0.249
8	0.282	0.183

[a]1 = male, 2 = female.

These solutions are the same as those obtained by direct inversion of the coefficient matrix in Section 3.1 or iterating on the coefficient matrix in Section 13.1. However, as stated earlier, the advantage of this method is that the MME are not set up; therefore memory requirement is minimal and can be applied to large data sets.

13.3.2 Animal model with groups

Example 13.4

With unknown parents assigned to phantom groups, the procedure is very similar to that described in Section 13.3.1 above with no groups in the model except in the manner in which the pedigree file is set up and animal solutions are computed. Using the same data, parameters and model as in Example 3.4, the methodology is illustrated below.

Data preparation

The pedigree file is set up as described in Section 13.3.1 with ancestors with unknown parentage assigned to groups. The assignment of unknown parents for the example pedigree has been described in Section 3.5. However, there is also an additional column for each animal indicating the number of unknown parents for each animal.

The pedigree with unknown parents assigned to groups and the additional column indicating the number of unknown parents is as follows:

Calf	Sire	Dam	Number of unknown parents
1	9	10	2
2	9	10	2
3	9	10	2
4	1	10	1
5	3	2	0
6	1	2	0
7	4	5	0
8	3	6	0

and the ordered pedigree for the analysis is:

Animal	Code	Sire or progeny	Dam or mate	Number of unknown parents
1	1	9	10	2
1	2	4	10	1
1	2	6	2	0
2	1	9	10	2
2	2	5	3	0
2	2	6	1	0
3	1	9	10	2
3	2	5	2	0
3	2	8	6	0
4	1	1	10	1
4	2	7	5	0
5	1	3	2	0
5	2	7	4	0
6	1	1	2	0
6	2	8	3	0
7	1	4	5	0
8	1	3	6	0

(*Continued*)

Animal	Code	Sire or progeny	Dam or mate	Number of unknown parents
9	2	1	10	2
9	2	2	10	2
9	2	3	10	2
10	2	1	9	2
10	2	2	9	2
10	2	3	9	2
10	2	4	1	1

The arrangement of yield data is the same as in Section 13.3.1 in the animal model analysis without groups.

ITERATIVE STAGE

SOLVING FOR FIXED EFFECTS. This is exactly as described for the animal model without groups in Section 13.3.1, with yield records adjusted for other effects in the model and solutions for fixed effects computed.

SOLVING FOR ANIMAL SOLUTIONS. Solutions for animals are computed one at a time as both pedigree and data file sorted by animals are read, as described for the animal model without groups. Therefore, only the differences in terms of the way diagonals and adjusted right-hand sides are accumulated are outlined.

For the kth animal in the pedigree file, calculate:

$$w_k = \alpha(4/(2 + \text{no. of unknown parents}))$$

For the type 1 record in the pedigree file for the kth animal:

$$\text{ARHS}_k = \text{ARHS}_k + (\hat{u}_s + \hat{u}_d)0.5w_k$$

$$\text{DIAG}_k = \text{DIAG}_k + w_k$$

For the type 2 record in the pedigree file for the kth animal:

$$\text{ARHS}_k = \text{ARHS}_k + (\hat{u}_o - 0.5\hat{u}_m)0.5w_k$$

Accumulation of adjusted right sides from the data file is as specified in Section 13.3.1 in the model without groups.

The solution for the kth animal is computed as $\text{ARHS}_k/\text{DIAG}_k$ when all records for the animal in the pedigree and data file have been read. The solutions in the first round of iteration and at convergence without and with constraint on group solutions, as in Example 3.4, are as follows:

Effects	Solutions		
	At round 1	At convergence	At convergence[a]
Sex[b]			
1	4.333	4.509	5.474
2	3.400	3.364	4.327
Animal			
1	0.333	0.182	–0.780
2	–0.083	0.026	–0.937
3	–0.021	–0.014	–0.977
4	–0.119	–0.319	–1.287
5	–0.376	–0.150	–1.113
6	0.392	0.221	–0.741
7	–0.364	–0.389	–1.355
8	0.282	0.181	–0.782
9	0.153	0.949	0.000
10	–0.176	–0.820	–1.795

[a] With solutions for groups constrained to those in Example 3.4.
[b] 1 = male, 2 = female.

When the solutions for groups are constrained as those in Example 3.4, this method gives the same solutions obtained in Example 3.4. However, when there is no constraint on group solutions, the ranking of animals is the same and linear differences between levels of effects are very similar to those obtained when group solutions are constrained.

13.3.3 Reduced animal model with maternal effects

The principles of genetic evaluation iterating on the data with a reduced animal model with maternal effects are illustrated using the same data set, parameters and model as in Example 6.1. The genetic parameters were:

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{m}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & 0 & 0\\ g_{21} & g_{22} & 0 & 0\\ 0 & 0 & \sigma^2_{pe} & 0\\ 0 & 0 & 0 & \sigma^2_{e}\end{bmatrix} = \begin{bmatrix} 150 & -40 & 0 & 0\\ -40 & 90 & 0 & 0\\ 0 & 0 & 40 & 0\\ 0 & 0 & 0 & 350\end{bmatrix}$$

and:

$$\mathbf{G}^{-1} = \begin{bmatrix} g^{11} & g^{12}\\ g^{21} & g^{22}\end{bmatrix} = \begin{bmatrix} 0.00756 & 0.00336\\ 0.00336 & 0.01261\end{bmatrix}$$

The inverse of the residual variance for parental records is $1/\sigma_e^2 = r_{pa}^{-1} = 0.002857$ and for non-parental records is $1/(\sigma_e^2 + dg_{11}) = r_{np}^{-1}$, where $d = 3/4$ or 1/2 when one or both parents are known and the inverse of the variance due to permanent environmental effect = $1/\sigma_{pe}^2 = 0.025$.

Data arrangement

The pedigree file is set up as described in Section 13.3.1 but only for animals that are parents. The pedigree file for the example data is:

Animal	Code	Sire or progeny	Dam or mate
1	1	0	0
1	2	5	2
1	2	9	6
2	1	0	0
2	2	5	1
2	2	6	3
3	1	0	0
3	2	6	2
3	2	8	5
4	1	0	0
4	2	7	6
5	1	1	2
5	2	8	3
6	1	3	2
6	2	7	4
6	2	9	1
7	1	4	6
8	1	3	5
9	1	1	6

A data file is set up consisting of a code to identify parents and non-parents. For non-parents, one record is set up for each parent, comprising the parent, a code indicating it is a non-parent, the animal that has the yield record, the other parent (mate), the sire and dam of the animal with the yield record, fixed effects, covariates (if any) and traits. A single record is set up for parents, comprising the animal, a code indicating that it is a parent, the animal again, a field set to zero corresponding to the column for the other parent in non-parents' records, the sire and dam of the animal, fixed effects, covariates (if any) and traits. The data file may be sorted in three sequences if there is a major fixed effect in the model: sorted by major fixed effect, such as herd–year–season (HYS) (file A),

sorted by animal (file B), and sorted by dam code (file C). For the example, file A is:

Parent/ animal	Code[a]	Animal	Mate	Sire	Dam	Herd	Sex[b]	Birth weight
5	0	5	0	1	2	1	1	35.0
6	0	6	0	3	2	1	2	20.0
7	0	7	0	4	6	1	2	25.0
8	0	8	0	3	5	1	1	40.0
9	0	9	0	1	6	2	1	42.0
3	1	10	2	3	2	2	2	22.0
2	1	10	3	3	2	2	2	22.0
3	1	11	7	3	7	2	2	35.0
7	1	11	3	3	7	2	2	35.0
8	1	12	7	8	7	3	2	34.0
7	1	12	8	8	7	3	2	34.0
9	1	13	2	9	2	3	1	20.0
2	1	13	9	9	2	3	1	20.0
3	1	14	6	3	6	3	2	40.0
6	1	14	3	3	6	3	2	40.0

[a]0 = Parental record, 1 = non-parental record.
[b]1 = male, 2 = female.

Iteration stage

The solution vectors for herd ($\hat{\mathbf{hd}}$), sex ($\hat{\mathbf{b}}$), direct animal effect ($\hat{\mathbf{u}}$), genetic maternal effect ($\hat{\mathbf{m}}$) and permanent environmental effect ($\hat{\mathbf{pe}}$) are initially set to zero.

SOLVING FOR FIXED EFFECTS. Data file A is read at each round of iteration one herd at a time with adjusted right-hand sides (ARHS) and diagonals (DIAG) accumulated for the *i*th herd as:

$$\text{ARHS}_i = \text{ARHS}_i + r_{pd}^{-1}(y_{ijklt} - \hat{b}_j - \hat{u}_k - \hat{m}_l - \hat{p}e_t)$$

for parental records [13.10]:

$$\text{ARHS}_i = \text{ARHS}_i + r_{np}^{-1}(y_{ijklt} - \hat{b}_j - 0.5(\hat{u}_s + \hat{u}_d) - \hat{m}_l - \hat{p}e_t)$$

for non-parent records [13.11]:

$$\text{DIAG}_i = \text{DIAG}_i + r_n^{-1}$$

where r_n^{-1} is the inverse of the residual variance of the *n*th record being read.

At the end of records for the *i*th herd, the solution is computed as:

$$\hat{h}d_i = \text{ARHS}_i/\text{DIAG}_i$$

In the first round of iteration, the solution for the first herd is:

$$\begin{aligned}\hat{h}d_1 &= [r_{pd}^{-1}(y_1 - \hat{b}_1 - \hat{u}_5 - \hat{m}_2 - \hat{p}e_2) + (y_2 - \hat{b}_2 - \hat{u}_6 - \hat{m}_2 - \hat{p}e_2) \\ &\quad + (y_3 - \hat{b}_2 - \hat{u}_7 - \hat{m}_6 - \hat{p}e_6) + (y_4 - \hat{b}_1 - \hat{u}_8 - \hat{m}_5 - \hat{p}e_5)]/4(r_{pd}^{-1}) \\ &= [r_{pd}^{-1}((35-0-0-0-0) + (20-0-0-0-0) \\ &\quad + (25-0-0-0-0) + (40-0-0-0-0)]/4(r_{pd}^{-1}) \\ &= 0.3432/0.01144 = 30.00\end{aligned}$$

While reading data file A, adjusted right-hand sides, consisting of yield adjusted for previous animal, maternal and permanent environmental solutions, are accumulated for each level of sex effect. Thus, for the *j*th level of sex effect:

$$\text{ARHS}_j = \text{ARHS}_j + r_{pd}^{-1}(y_{ijklt} - \hat{u}_k - \hat{m}_l - \hat{p}e_t)$$

for parent records:

$$\text{ARHS}_j = \text{ARHS}_j + r_{np}^{-1}(y_{ijklt} - 0.5(\hat{u}_s + \hat{u}_d) - \hat{m}_l - \hat{p}e_t)$$

and for non-parent records:

$$\text{DIAG}_j = \text{DIAG}_j + r_n^{-1}$$

After reading file A, the solution for the *j* sex class is computed as:

$$\begin{aligned}\text{ARHS}_j &= \text{ARHS}_j - nr_{ij}^{-1}\hat{h}d_i \\ \hat{b}_j &= \text{ARHS}_j/\text{DIAG}_j\end{aligned}$$

where $\hat{h}d_i$ is the current solution of herd *i* and nr_{ij}^{-1} is the sum of the inverse of the residual variance for records of the *j*th level of sex effect in herd *i*. The latter is accumulated while reading file A. For the example data, solutions for level 1 of sex effect in the first round of iteration are:

$$\begin{aligned}\hat{b}_1 &= \text{ARHS}_1 - 2r_{pd}^{-1}(\hat{h}d_1) - r_{pd}^{-1}(\hat{h}d_2) - 2r_{np}^{-1}(\hat{h}d_3)/[3r_{pd}^{-1} + 2r_{np}^{-1}] \\ &= 0.38134 - 2r_{pd}^{-1}(30.0) - r_{pd}^{-1}(33.638) - 2r_{np}^{-1}(31.333)/0.01092 \\ &= 3.679\end{aligned}$$

After obtaining solutions for fixed effects in the current round of iteration, the solutions for animals are solved for.

SOLVING FOR ANIMAL SOLUTIONS . As described in Section 13.3.1, animal solutions are computed one at a time as the pedigree file and file B are read. Briefly, for a type 1 record in the pedigree file for the *k*th animal, contributions to DIAG and ARHS according to the number of parents known [13.8] are:

Number of parents known		
None	One (sire(*s*))	Both
$\text{ARHS}_k = 0$ $\text{DIAG}_k = g^{11}$	$\text{ARHS}_k = \frac{2}{3}g^{11}(\hat{u}_s)$ $\text{DIAG}_k = \frac{4}{3}g^{11}$	$\text{ARHS}_k = g^{11}(\hat{u}_s + \hat{u}_d)$ $\text{DIAG}_k = 2g^{11}$

where $\hat{u}_s$ and $\hat{u}_d$ are current solutions for direct effects for the sire and dam of the animal k.

The adjusted right-hand side is augmented by contributions from the maternal effect as a result of the genetic correlation between animal and maternal effects. These contributions are from the sire, dam and the kth animal (see equation [13.9]), and these are:

Number of parents known		
None	One (sire(s))	Both
–	$\text{ARHS}_k = \text{ARHS}_k + (\hat{m}_s)\frac{2}{3}g^{12}$	$\text{ARHS}_k = \text{ARHS}_k + (\hat{m}_s + \hat{m}_d)g^{12}$
$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)g^{12}$	$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)\frac{4}{3}g^{12}$	$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)2g^{12}$

where $\hat{m}_s$, $\hat{m}_d$ and $\hat{m}_k$ are current maternal solutions for the sire and dam of animal k, respectively.

In processing a type 2 record in the pedigree file for the kth animal, contributions to DIAG and ARHS according to whether the mate of k is known are:

Mate is unknown	Mate is known
$\text{ARHS}_k = \text{ARHS}_k + \frac{2}{3}g^{11}(\hat{u}_o)$	$\text{ARHS}_k = \text{ARHS}_k + (\hat{u}_o - 0.5\hat{u}_{ma})g^{11}$
$\text{DIAG}_k = \text{DIAG}_k + \frac{1}{3}g^{11}$	$\text{DIAG}_k = \text{DIAG}_k + \frac{1}{2}g^{11}$

where $\hat{u}_o$ and $\hat{u}_{ma}$ are current solutions for direct effects for the progeny and mate of the animal k.

Accounting for contributions from the maternal effect to ARHS:

Mate is unknown	Mate is known
$\text{ARHS}_k = \text{ARHS}_k + \frac{2}{3}g^{12}(\hat{m}_o)$	$\text{ARHS}_k = \text{ARHS}_k + (\hat{m}_o - 0.5\hat{m}_{ma})g^{12}$
$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)\frac{1}{3}g^{12}$	$\text{ARHS}_k = \text{ARHS}_k - (\hat{m}_k)\frac{1}{2}g^{12}$

where $\hat{m}_o$ and $\hat{m}_{ma}$ are current maternal solutions for the progeny and mate of the animal k.

If the animal has a yield record:

$$\text{DIAG}_k = \text{DIAG}_k + r_n^{-1} \quad \text{if it is a parent}$$

or:

$$\text{DIAG}_k = \text{DIAG}_k + (r_n^{-1})\,0.5 \quad \text{if it is a non-parent}$$

The diagonals of non-parents are multiplied by 0.5 instead of 0.25 because records of non-parents have been written twice (see Section 13.3).

Contributions to the right-hand side are accumulated as:

$$\text{ARHS}_k = \text{ARHS}_k + r_{pa}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - \hat{m}_l - \hat{p}e_t)$$

for parent records and:

$$\text{ARHS}_k = \text{ARHS}_k + r_{np}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - 0.5(\hat{u}_{ma}) - \hat{m}_l - \hat{p}e_t)$$

for non-parent records. In the equations above $\hat{h}d_i$, $\hat{b}_j$, $\hat{m}_l$, $\hat{p}e_t$ and $\hat{u}_{ma}$ are current solutions for herd *i*, *j*th level for sex effect, *l*th maternal effect level, *t*th level of permanent environment effect and animal solution for the other parent (mate), respectively. The solution for animal *k* is computed as usual when all records for the animal in the pedigree and data file have been read as:

$$\hat{u}_k = \text{ARHS}_k / \text{DIAG}_k$$

The solution for animal 2 in the example data in the first round of iteration is as follows.

The contribution to the diagonal from pedigree is:

$$\text{DIAG}_2 = (1 + \tfrac{1}{2} + \tfrac{1}{2})\, 0.00756 = 0.01512$$

The contribution to the diagonal from yield is:

$$\text{DIAG}_2 = \text{DIAG}_2 + 2(0.00059) = 0.01512 + 0.00118 = 0.0163$$

The contributions to ARHS from the pedigree (both parents are known) and from its progeny are zero in the first round of iteration. The contribution to ARHS from yield record is:

$$\begin{aligned}\text{ARHS}_2 &= r_{np}^{-1}(y_{10} - h\hat{d}_2 - \hat{b}_2 - \hat{u}_3 - \hat{m}_3 - \hat{p}e_3) \\ &\quad + r_{np}^{-1}(y_{13} - hd_3 - b_1 - \hat{u}_9 - m_2 - pe_2) \\ \text{ARHS}_2 &= r_{np}^{-1}(22 - (-2.567) - 33.600 - 0 - 0 - 0) \\ &\quad + r_{np}^{-1}(20 - 3.679 - 31.333) = -0.02818\end{aligned}$$

Therefore:

$$\hat{u}_2 = -0.02818/0.0163 = -1.729$$

After processing all animals in the pedigree and data file in the current round of iteration, equations for maternal effects are set and solved as described below.

SOLVING FOR MATERNAL EFFECT. Solutions for maternal effects are computed using both the pedigree file and the data file sorted by dam. Records for the *l*th animal are read in from the pedigree file and from file C if it is a dam that has progeny with a yield record, while accumulating diagonals (DIAG) and adjusted right-hand sides (ARHS). For the type 1 record in the pedigree file for animal *l*, contributions to ARHS and DIAG according to the

number of parents known are as follows:

Number of parents known		
None	One (dam(d))	Both
$ARHS_l = 0$ $DIAG_l = g^{22}$	$ARHS_l = \frac{2}{3} g^{22}(\hat{m}_d)$ $DIAG_l = \frac{4}{3} g^{22}$	$ARHS_l = g^{22}(\hat{m}_s + \hat{m}_d)$ $DIAG_l = 2g^{22}$

Taking into account contributions from animal effects to the ARHS due to the genetic correlation gives:

Number of parents known		
None	One (dam(d))	Both
–	$ARHS_l = ARHS_l + (\hat{u}_d)\frac{2}{3} g^{12}$	$ARHS_l = ARHS_l + (\hat{u}_s + \hat{u}_d)g^{12}$
$ARHS_l = ARHS_l - (\hat{u}_l)g^{12}$	$ARHS_l = ARHS_l - (\hat{u}_l)\frac{4}{3} g^{12}$	$ARHS_l = ARHS_l - (\hat{u}_l)2g^{12}$

For the type 2 record in the pedigree file for animal l, contributions to the ARHS and DIAG according to whether the mate of animal l is known or not are:

Mate is unknown	Mate is known
$ARHS_l = ARHS_l + (\frac{2}{3})g^{22}(\hat{m}_o)$ $DIAG_l = DIAG_l + (\frac{1}{3})g^{22}$	$ARHS_l = ARHS_l + g^{22}(\hat{m}_o - 0.5m_{\hat{ma}})$ $DIAG_l = DIAG_l + (\frac{1}{2})g^{22}$

Taking into account contributions from animal effect (see equation [13.9]) gives:

Mate is unknown	Mate is known
$ARHS_l = ARHS_l + \frac{2}{3} g^{12}(\hat{u}_0)$ $ARHS_l = ARHS_l - (\hat{u}_l)(\frac{1}{3})g^{12}$	$ARHS_l = ARHS_l + (\hat{u}_o - 0.5\hat{u}_{ma})g^{12}$ $ARHS_l = ARHS_l - (\hat{u}_l)(\frac{1}{2})g^{12}$

For the l animal which is a dam with progeny having yield records, diagonals and adjusted right-hand sides from pedigree are augmented with information from yield as:

$$DIAG = DIAG + r_n^{-1}$$

and:

$$\text{ARHS}_l = \text{ARHS}_l + r_{pa}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - \hat{u}_k - \hat{p}e_t)$$

for parent records and:

$$\text{ARHS}_l = \text{ARHS}_l + r_{np}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - 0.5(\hat{u}_s + \hat{u}_d) - \hat{p}e_t)$$

for non-parent records.

After processing all records from pedigree and yield records for the *l*th animal, the solution for the maternal effect is computed as:

$$m_l = \text{ARHS}_l / \text{DIAG}_l$$

The calculation of the solution for animal 5 in the first round of iteration is as follows:

The contribution from type 1 record in the pedigree is:

$$\begin{aligned}\text{ARHS}_5 &= (\hat{m}_1 + \hat{m}_2)g^{22} + (\hat{u}_1 + \hat{u}_2)g^{12} - (\hat{u}_5 2g^{12}) \\ &= (0.0217 + (-1.7027))\ 0.01261 + (0 + (-1.7294))\ 0.00336 \\ &\quad -((-0.5831)(2)0.0336) = -0.02309\end{aligned}$$

$$\text{DIAG}_5 = (2)0.01261 = 0.02522$$

The contribution from type 2 record in the pedigree is:

$$\begin{aligned}\text{ARHS}_5 &= \text{ARHS}_5 + (\hat{m}_8 - \tfrac{1}{2}\hat{m}_3)g^{22} + (\hat{u}_8 - \tfrac{1}{2}\hat{u}_3)g^{12} - (\hat{u}_5\ \tfrac{1}{2}g^{12}) \\ &= -0.02309 + (0 - \tfrac{1}{2}(0.4587))\ 0.01261 \\ &\quad + (1.4382 - \tfrac{1}{2}(0.8960))\ 0.00336 - ((-0.5831)(\tfrac{1}{2})\ 0.00336) \\ &= -0.021675\end{aligned}$$

$$\text{DIAG}_5 = \text{DIAG}_5 + \tfrac{1}{2}g^{22} = 0.02522 + 0.0063 = 0.03153$$

The contribution from yield of progeny (animal 8) to dam 5 is:

$$\begin{aligned}\text{ARHS}_5 &= \text{ARHS}_5 + r_{pa}^{-1}(y_8 - \hat{h}d_1 - \hat{b}f_1 - \hat{u}_8 - \hat{p}e_5) \\ &= -0.021675 + r_{pa}^{-1}(40 - 30.00 - 3.679 - (1.4382) - 0) \\ &= -0.007724\end{aligned}$$

$$\text{DIAG}_5 = \text{DIAG}_5 + r_{pa}^{-1} = 0.03153 + 0.002857 = 0.034387$$

and the solution is:

$$\hat{m}_5 = -0.007724/0.034387 = -0.225$$

Solutions for permanent environmental effects are solved for after processing all animals for maternal effects in the current round of iteration.

SOLVING FOR PERMANENT ENVIRONMENTAL (*pe*) EFFECTS. Only the data file sorted by dams is required to obtain solutions for *pe* effects.

The records for the tth dam are read from file C while adjusted right-hand sides and diagonals are accumulated as:

$$\text{ARHS}_t = \text{ARHS}_t + r_{pa}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - \hat{u}_k - \hat{m}_l)$$

for parent records and:

$$\text{ARHS}_t = \text{ARHS}_t + r_{np}^{-1}(y_{ijklt} - \hat{h}d_i - \hat{b}_j - 0.5\,(\hat{u}_s + \hat{u}_d) - \hat{m}_l)$$

for non-parent records.

$$\text{DIAG}_t = \text{DIAG}_t + r_n^{-1}$$

At the end of records for the tth dam, solutions are computed as:

$$\hat{pe}_t = \text{ARHS}_t / (\text{DIAG}_t + 1/\sigma_{pe}^2)$$

The solution for permanent environmental effect for animal 5 in the first round of iteration is:

$$\begin{aligned}\text{ARHS}_5 &= r_{pa}^{-1}(y_8 - \hat{h}d_1 - \hat{b}_1 - \hat{u}_8 - \hat{m}_5) \\ &= r_{pa}^{-1}(40.0 - 3.679 - 30.0 - 1.4822 - (-0.2246)) \\ &= 0.01459\end{aligned}$$

$$\text{DIAG}_5 = r_{pa}^{-1} + 0.025 = 0.02786$$

and:

$$pe_5 = 0.01459/0.02786 = 0.524$$

Further iterations are carried out until convergence is achieved. The convergence criteria defined in Section 13.2.1 could also be used. The solutions for the first round of iteration and at convergence are shown below:

	Solutions	
Effects	At round 1	At convergence
Herd		
1	30.000	30.563
2	33.600	33.950
3	31.333	31.997
Sex[a]		
1	3.679	3.977
2	–2.657	–2.872
Animal		
1	0.000	0.564
2	–1.729	–1.246
3	0.896	1.166
4	0.000	–0.484
5	–0.583	0.630

(*Continued*)

	Solutions	
Effects	At round 1	At convergence
6	–0.554	–0.859
7	–0.020	–1.156
8	1.438	1.918
9	–0.396	–0.553
Maternal		
1	0.022	0.261
2	–1.703	–1.582
3	0.459	0.735
4	0.046	0.586
5	–0.225	–0.507
6	0.425	0.841
7	0.788	1.299
8	–0.224	–0.158
9	0.255	0.659
Permanent environment		
2	–1.386	–1.701
5	0.524	0.415
6	0.931	0.825
7	0.527	0.461

[a]1 = male, 2 = female.

These solutions are exactly the same as those obtained in Section 6.2 by directly inverting the coefficient matrix.

BACK-SOLVING FOR NON-PARENTS

The solutions for direct animal and maternal effects for non-parents are calculated after convergence has been achieved, as described in Section 6.2. The solutions for non-parents for this example have been calculated in Section 6.2.

13.4 Preconditioned Conjugate Gradient Algorithm

Berger *et al*. (1989) investigated the use of a plain or Jacobi conjugate gradient (CG) iterative scheme for solving MME for the prediction of sire breeding values. They indicated that plain CG was superior to a number of other iterative schemes, including Gauss–Seidel. Strandén and Lidauer (1999) implemented the use of the preconditioned conjugate gradient (PCG) in genetic evaluation models for the routine evaluation of dairy cattle with very large data. In the PCG method, the linear system of equations

(equation [13.1], for instance) is made simpler by solving an equivalent system of equations:

$$\mathbf{M}^{-1}\mathbf{Cb} = \mathbf{M}^{-1}\mathbf{r}$$

where **M** is a symmetric, positive definite, preconditioner matrix that approximates **C** and **r** is the right-hand side. In the plain CG method, the preconditioner **M** is an identity matrix.

The implementation of the PCG method requires storing four vectors of size equal to the number of unknowns in the MME: a vector of residuals (**e**), a search-direction vector (**d**), a solution vector (**b**) and a work vector (**v**). The PCG method can be implemented with less memory by storing the solution vector on disk and reading it in during the iteration. The pseudo-code for the PCG method (Lidauer *et al.*, 1999) is outlined below, assuming that starting values are:

$$\mathbf{b}^{(0)} = 0, \quad \mathbf{e}^{(0)} = \mathbf{r} - \mathbf{Cb} = \mathbf{r}, \quad \mathbf{d}^{(0)} = \mathbf{M}^{-1}\mathbf{e}^{(0)} = \mathbf{M}^{-1}\mathbf{r}^{(0)}$$

For $k = 1,2, \ldots, n$:

$$\mathbf{v} = \mathbf{Cd}^{(k-1)}$$

$$\omega = \mathbf{e}'^{(k-1)}\mathbf{M}^{-1}\mathbf{e}^{(k-1)}/(\mathbf{d}'^{(k-1)}\mathbf{v})$$

$$\mathbf{b}^{(k)} = \mathbf{b}^{(k-1)} + \omega\mathbf{d}^{(k-1)}$$

$$\mathbf{e}^{(k)} = \mathbf{e}^{(k-1)} - \omega\mathbf{v}$$

$$\mathbf{v} = \mathbf{M}^{-1}\mathbf{e}^{(k)}$$

$$\beta = \mathbf{e}'^{(k)}\mathbf{v}/(\mathbf{e}'^{(k-1)}\mathbf{M}^{-1}\mathbf{e}^{(k-1)})$$

$$\mathbf{d}^{(k)} = \mathbf{v} + \beta\mathbf{d}^{(k-1)}$$

If not converged, continue iteration until converged. End where **ω** and **β** are step sizes in the PCG method.

13.4.1 Computation strategy

The major task in the PCG algorithm above is calculating **Cd**, where **C** is the coefficient matrix of the MME. The vector **d** is the search direction vector and every iteration of the PCG minimizes the distance between the current and the true solutions in the search direction. Strandén and Lidauer (1999) presented an efficient computation strategy for computing **Cd** for a multivariate model. Assuming, for instance, that data are ordered by animals, the MME for the multivariate model (equation [5.2]), can be written as:

$$\begin{bmatrix} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{R}_i^{-1} \mathbf{x}_i' & \sum_{i=1}^{N} \mathbf{x}_i \mathbf{R}_i^{-1} \mathbf{z}_i' \\ \sum_{i=1}^{N} \mathbf{z}_i \mathbf{R}_i^{-1} \mathbf{x}_i' & \sum_{i=1}^{N} \mathbf{z}_i \mathbf{R}_i^{-1} \mathbf{z}' + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{R}_i^{-1} \mathbf{y}_i \\ \sum_{i=1}^{N} \mathbf{z}_i \mathbf{R}_i^{-1} \mathbf{y}_i \end{bmatrix}$$

where N is the number of animals with records, $\mathbf{x}'_i$ and $\mathbf{z}'_i$ are matrices having rows equal to the number of traits observed on animal i. Denote $\mathbf{w}' = [\mathbf{x}'_i \mathbf{z}'_i]$ and $\mathbf{V}$ as:

$$\mathbf{V} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{A} \otimes \mathbf{G} \end{bmatrix}$$

Computing $\mathbf{Cd}$ then implies calculating:

$$\sum_{i=1}^{N} \mathbf{w}_i \mathbf{R}_i^{-1} \mathbf{w}'_i \mathbf{d} + \mathbf{V}^{-1}\mathbf{d} = \sum_{i=1}^{N} \mathbf{v}_i + \mathbf{v}_d \qquad [13.13]$$

If solving the MME with iteration on the data for a univariate model without any regression effects, this calculation can be achieved by accumulating, for each individual i, the product $\mathbf{v}_i = \mathbf{T}_i\mathbf{d}$, where the coefficients in $\mathbf{T}_i = \mathbf{w}_i \mathbf{R}_i^{-1} \mathbf{w}'_i$ can be deduced without performing any of the products, as $\mathbf{w}_i$ contains zeros and ones only and $\mathbf{R}_i^{-1}$ is a scalar or $\mathbf{R}_i^{-1}$ is factored out (equation [3.4]). For a multivariate model, the principles for computing $\mathbf{T}_i$ are essentially the same but with scalar contributions replaced by matrix $\mathbf{R}_i$. Strandén and Lidauer (1999) suggested the following three-step method for calculating the product $\mathbf{w}_i \mathbf{R}_i^{-1}\mathbf{d}$:

$$\mathbf{s}_i \leftarrow \mathbf{w}'_i\mathbf{d}; \quad \mathbf{s}_i^* \leftarrow \mathbf{R}_i^{-1}\mathbf{s}_i; \quad \mathbf{v}_i \leftarrow \mathbf{w}_i\mathbf{s}_i^*$$

where vectors $\mathbf{s}_i$ and $\mathbf{s}_i^*$ are of size equal to the number of traits observed on individual i (l_i). They demonstrated that this three-step approach substantially reduced the number of floating point operations (multiplications) compared with a multivariate accumulation technique, as used by Groeneveld and Kovac (1990). For instance, given that q_i is the number of effects over traits observed for individual i, the number of floating point operations were 720 with $l_i = 3$ and $q_i = 15$ using the multivariate accumulation technique, compared with 78 with the three-point approach. They also suggested that $\mathbf{v}_d = \mathbf{V}^{-1}\mathbf{d}$ in equation [13.13] can be evaluated in a two-step approach:

$$\mathbf{x} \leftarrow (\mathbf{I} \otimes \mathbf{A}^{-1})\mathbf{d}; \quad \mathbf{v}_d \leftarrow (\mathbf{G}^{-1} \otimes \mathbf{I})\mathbf{x}$$

13.4.2 Numerical application

The application of PCG to solve MME is illustrated using data for Example 3.1 in Chapter 3 for a univariate model and iterating on the data.

Computing starting values

Initially, the pedigree is read and diagonal elements of $\mathbf{A}^{-1}$ multiplied by α are accumulated for animals, where the variance ratio α is 2, as in Example 3.1. This is straightforward and has not been illustrated, but elements for animals 1 to 8 stored in a vector $\mathbf{h}$ are:

$$\mathbf{h}' = [3.667 \quad 4.0 \quad 4.0 \quad 3.667 \quad 5.0 \quad 5.0 \quad 4.0 \quad 4.0]$$

Secondly, read through the data as shown in Table 3.1 and accumulate right-hand side (**r**) for all effects, diagonals for the levels of sex of calf effect and add contribution of information from data to diagonals from $\mathbf{A}^{-1}\alpha$ for animals. Assuming that diagonals for all effects are stored as diagonal elements of **M**, such that the first two elements are for the two levels of sex of calf effect and the remaining elements for animals 1 to 8, then **r** and **M** are:

$$\mathbf{r}' = [13.0 \;\; 6.8 \;\; 0.0 \;\; 0.0 \;\; 0.0 \;\; 4.5 \;\; 2.9 \;\; 3.9 \;\; 3.5 \;\; 5.0]$$

and:

$$\mathbf{M} = \text{diag}[3.0 \;\; 2.0 \;\; 3.667 \;\; 4.0 \;\; 4.0 \;\; 4.667 \;\; 6.0 \;\; 6.0 \;\; 5.0 \;\; 5.0]$$

The starting values for PCG can now be calculated. Thus:

$$\mathbf{b}^{(0)} = 0, \quad \mathbf{e}^{(0)} = \mathbf{r} - \mathbf{C}\mathbf{b}^{(0)} = \mathbf{y} \quad \text{and} \quad \mathbf{d}^{(0)} = \mathbf{M}^{-1}\mathbf{r}$$

Thus:

$$\mathbf{d}^{(0)\prime} = [4.333 \;\; 3.4 \;\; 0.0 \;\; 0.0 \;\; 0.0 \;\; 0.964 \;\; 0.483 \;\; 0.650 \;\; 0.70 \;\; 1.0]$$

Iterative stage

Reading through the data and performing the following calculatons in each round of iteration start the PCG iterative process. Calculations are shown for the first round of iteration.

The vector $\mathbf{v} = \mathbf{C}\mathbf{d}$ is accumulated as data are read. For the ith level of fixed effect:

$$\mathbf{v}(i) = \mathbf{v}(i) + 1(\mathbf{d}(i)) + 1(\mathbf{d}(\text{anim}_k))$$

where anim_k refers to the animal k associated with the record. Thus for the level 1 of sex of calf effect:

$$\mathbf{v}(1) = 3(4.333) + \mathbf{d}(\text{anim}_4) + \mathbf{d}(\text{anim}_7) + \mathbf{d}(\text{anim}_8) = 15.663$$

As each record is read, calculate:

z = 4/(2 + number of unknown parents for animal with record)
$xx = -0.5(z)\alpha$ if either parent is known, otherwise $xx = 0$
$xm = 0.25(z)\alpha$ if both parents are known, otherwise $xm = 0$

If only one parent, p, of animal k is known, then accumulate:

$$\mathbf{v}(\text{anim}_k) = \mathbf{v}(\text{anim}_k) + 1(\mathbf{d}(i)) + \mathbf{M}_{k,k}(\mathbf{d}(\text{anim}_k)) + xx(\mathbf{d}(anim_p)) \quad [13.14]$$

where $\mathbf{d}(i)$ refers to the ith element of the ith level of the fixed effect and $\mathbf{M}_{k,k}$ the diagonal element of **M** for animal k.

Accumulate the contribution to the known parent, p, of k at the same time:

$$\mathbf{v}(\text{anim}_p) = \mathbf{v}(\text{anim}_p) + xx(\mathbf{d}(\text{anim}_k))$$

If both parents p and j of animal k are known, then accumulate for animal k as:

$$\begin{aligned}\mathbf{v}(\text{anim}_k) = \mathbf{v}(\text{anim}_k) &+ 1(\mathbf{d}(i)) + \mathbf{M}_{k,k}(\mathbf{d}(\text{anim}_k)) \\ &+ xx(\mathbf{d}(\text{anim}_p) + \mathbf{d}(\text{anim}_j))\end{aligned} \quad [13.15]$$

Accumulate for both parents as:

$$\mathbf{v}(\text{anim}_p) = \mathbf{v}(\text{anim}_p) + xx(\mathbf{d}(\text{anim}_k))$$
$$\mathbf{v}(\text{anim}_p) = \mathbf{v}(\text{anim}_p) + xm(\mathbf{d}(\text{anim}_j))$$
$$\mathbf{v}(\text{anim}_j) = \mathbf{v}(\text{anim}_j) + xx(\mathbf{d}(\text{anim}_k))$$
$$\mathbf{v}(\text{anim}_j) = \mathbf{v}(\text{anim}_j) + xm(\mathbf{d}(\text{anim}_p))$$

After processing all animals with records, the contribution for animals in the pedigree without records is accumulated. The equations for accumulating contributions for these animals is the same as shown above except that the coefficient for $\mathbf{d}(i)$ in equations [13.14] and [13.15] is zero instead of one, indicating no contribution from records.

For example, for animal 4 with only the sire known:

$$\mathbf{v}(4) = \mathbf{v}(4) + \mathbf{d}(1) + \mathbf{M}_{4,4}(\mathbf{d}(\text{anim}_4)) + (-2/3)\alpha(\mathbf{d}(\text{anim}_1)) = 8.833$$

Add contribution from progeny when processing the record for animal 7:

$$\mathbf{v}(4) = 8.833 + -1.0\alpha(\mathbf{d}(\text{anim}_7)) + 0.25\alpha(\mathbf{d}(\text{anim}_5)) = 7.917$$

The vector $\mathbf{v}$ for all effects is:

$$\mathbf{v}' = [15.664 \;\; 7.933 \;\; -2.586 \;\; -2.267 \;\; -2.317 \;\; 7.917 \;\; 5.864 \;\; 5.300 \;\; 4.938 \;\; 8.033]$$

Next ω is computed using matrix multiplication and scalar division as:

$$\omega = 95.1793/120.225 = 0.7915$$

The solution vector is then computed as $\mathbf{b}^{(1)} = \mathbf{b}^{(0)} + \omega\mathbf{d}^{(0)}$. The vector $\mathbf{b}^{(1)}$ is:

$$\mathbf{b}'^{(1)} = [3.430 \;\; 2.691 \;\; 0.0 \;\; 0.0 \;\; 0.0 \;\; 0.763 \;\; 0.383 \;\; 0.514 \;\; 0.554 \;\; 0.791]$$

The updated vector of residuals $\mathbf{e}^{(1)}$ is computed as $\mathbf{e}^{(0)} - \omega\mathbf{v}$. For the example data, $\mathbf{e}^{(1)}$ is:

$$\mathbf{e}'^{(1)} = [0.602 \;\; 0.521 \;\; 2.047 \;\; 1.794 \;\; 1.834 \;\; -1.766 \;\; -1.741 \;\; -0.295 \;\; -0.408 \;\; -1.358]$$

The vector $\mathbf{v}$ is then computed as $\mathbf{M}^{-1}\mathbf{e}^{(1)}$. For the example data, $\mathbf{v}$ is:

$$\mathbf{v}' = [0.201 \;\; 0.260 \;\; 0.558 \;\; 0.449 \;\; 0.458 \;\; -0.378 \;\; -0.290 \;\; -0.049 \;\; -0.082 \;\; -0.272]$$

Next, compute the scalar β. The denominator of β is equal to the numerator of ω and this has already be computed. Using the example data:

$$\beta = 4.634/95.179 = 0.0487$$

Finally, $\mathbf{d}^{(1)}$, the search-direction vector for the next iteration is computed as $\mathbf{v} + \beta\mathbf{d}^{(0)}$. This vector for the example data is:

$$\mathbf{d}'^{(1)} = [0.412 \;\; 0.426 \;\; 0.558 \;\; 0.449 \;\; 0.458 \;\; -0.331 \;\; -0.267 \;\; -0.017 \;\; -0.048 \;\; -0.223]$$

The next cycle of iteration is continued until the system of equations converges. Convergence can be monitored using either the criteria defined

in Example 13.1 or the relative difference between the right-hand and left-hand sides:

$$c_d^{(r)} = \frac{\left\|\mathbf{y} - \mathbf{C}\mathbf{b}^{(r+1)}\right\|}{\left\|\mathbf{y}\right\|}$$

where:

$$\left\|\mathbf{X}\right\| = \left(\sum_i x_i^2\right)^{1/2}$$

Using the convergence criteria used in Example 13.1, the iteration was stopped at the tenth iteration when equations converged to 8.3^{-07}. Some intermediary and final solutions are shown in the following table:

	Iteration number				
Effects	1	3	5	7	10
Sex of calf[a]					
1	3.430	3.835	4.280	4.367	4.359
2	2.691	3.122	3.154	3.377	3.404
Animal					
1	0.000	0.475	0.170	0.092	0.098
2	0.000	0.224	0.116	0.012	–0.019
3	0.000	0.272	0.058	–0.056	–0.041
4	0.763	0.390	0.032	–0.029	–0.009
5	0.383	0.249	–0.072	–0.155	–0.186
6	0.514	0.547	0.435	0.194	0.177
7	0.554	0.193	–0.178	–0.231	–0.249
8	0.791	0.537	0.334	0.171	0.183

[a]1 = male, 2 = female.

The equations converged at the tenth round of iteration compared with 20 iterations on the data in Example 13.3.

Appendix A: Introductory Matrix Algebra

The basic elements of matrix algebra necessary for the understanding of the principles for the prediction of breeding values are briefly covered in this appendix. Little or no previous knowledge of matrix algebra is assumed. For a detailed study of matrix algebra, the reader should see Searle (1982).

A.1 Matrix: a Definition

A matrix is a rectangular array of numbers set in rows and columns. These elements are called the elements of a matrix. The matrix **B**, for instance, consisting of two rows and three columns, may be represented as:

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

or:

$$\mathbf{B} = \begin{bmatrix} 2 & 4 & 5 \\ 6 & 8 & 9 \end{bmatrix}$$

The element b_{ij} is called the ij element of the matrix, the first subscript referring to the row the element is in and the second to the column. The order of a matrix is the number of rows and columns. Thus a matrix of r rows and c columns has order $r \times c$ (read as r by c). The matrix **B** above is of the order 2×3 and can be written as $\mathbf{B}_{2\times 3}$.

A matrix consisting of a single row of elements is called a row vector. A row vector consisting of three elements may be represented as:

$$\mathbf{c} = [2 \quad 6 \quad -4]$$

Only one subscript is needed to specify the position of an element in a row vector. Thus the ith element in the row vector **c** above refers to the element in the ith column. For instance, $\mathbf{c}_3 = -4$.

Similarly, a matrix consisting of a single column is called a column vector. Again, only one subscript is needed to specify the position of an

element, which refers to the row the element is in, since there is only one column. A column vector **d** with four elements can be shown as below:

$$\mathbf{d} = \begin{bmatrix} -20 \\ 60 \\ 8 \\ 2 \end{bmatrix}$$

A scalar is a matrix with one row and one column.

A.2 Special Matrices

A.2.1 Square matrix

A matrix with an equal number of rows and columns is referred to as a square matrix. Shown below is a square matrix **G** of order 3×3.

$$\mathbf{G} = \begin{bmatrix} 2 & 1 & 6 \\ 4 & 2 & 7 \\ 0 & 4 & 8 \end{bmatrix}$$

The *ij* elements in a square matrix with *i* equal to *j* are called the diagonal elements. Other elements of the square matrix are called off-diagonal or non-diagonal elements. Thus the diagonal elements in the **G** matrix above are 2, 2 and 8.

A.2.2 Diagonal matrix

A square matrix having zero for all of its off-diagonal elements is referred to as a diagonal matrix. For example, a diagonal matrix **B** can be shown as below:

$$\mathbf{B} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 18 \end{bmatrix}$$

When all the diagonal elements of a diagonal matrix are one, it is referred to as an identity matrix. Given below is an identity matrix, **I**:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A.2.3 Triangular matrix

A square matrix with all elements above the diagonal being zero is called a lower triangular matrix. When all the elements below the diagonal

are zeros, it is referred to as an upper triangular matrix. For instance, the matrices **D**, a lower triangular matrix and **E**, an upper triangular matrix, can be illustrated as below:

$$\mathbf{D} = \begin{bmatrix} 4 & 0 & 0 \\ 1 & 3 & 0 \\ -2 & 7 & 9 \end{bmatrix}; \quad \mathbf{E} = \begin{bmatrix} 3 & 9 & 1 \\ 0 & 4 & 8 \\ 0 & 0 & 6 \end{bmatrix}$$

The transpose (see A.3.1) of an upper triangular matrix is a lower triangular matrix and vice versa.

A.2.4 Symmetric matrix

A symmetric matrix is a square matrix with the elements above the diagonal equal to the corresponding elements below the diagonal, that is, element *ij* is equal to element *ji*. The matrix **A** below is an example of a symmetric matrix:

$$\mathbf{A} = \begin{bmatrix} 2 & -4 & 0 \\ -4 & 6 & 3 \\ 0 & 3 & 7 \end{bmatrix}$$

In the above matrix, note that:

$$a_{12} = a_{21} = -4$$

$$a_{13} = a_{31} = 0$$

$$a_{23} = a_{32} = 3$$

A.3 Basic Matrix Operations

A.3.1 Transpose of a matrix

The transpose of a matrix **A** is usually written as **A′** or $\mathbf{A}^{\mathrm{T}}$, is the matrix whose *ji* elements are the *ij* elements of the original matrix, that is, $a'_{ji} = a_{ij}$. In other words, the columns of **A′** are the rows of **A** and the rows of **A′** the columns of **A**. For instance, the matrix **A** and its transpose **A′** are illustrated below:

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \\ 4 & 0 \end{bmatrix}; \quad \mathbf{A}' = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 1 & 0 \end{bmatrix}$$

Note that **A** is not equal to **A′** but the transpose of a symmetric matrix is equal to the symmetric matrix. Also (**AB**)′ = **B′A′**, where **AB** refers to the product (see A.3.3) of **A** and **B**.

A.3.2 Matrix addition and subtraction

Two matrices can be added together only if they have the same number of rows and columns, that is, they are of the same order and they are said to be conformable for addition. Given that **W** is the sum of the matrices **X** and **Y**, then $w_{ij} = x_{ij} + y_{ij}$. For example, if **X** and **Y**, both of order 2×2, are as illustrated below,

$$\mathbf{X} = \begin{bmatrix} 40 & 10 \\ 39 & -25 \end{bmatrix}; \quad \mathbf{Y} = \begin{bmatrix} -2 & 20 \\ 4 & 40 \end{bmatrix}$$

Then the matrix **W**, the sum of **X** and **Y**, is:

$$\mathbf{W} = \begin{bmatrix} 40+(-2) & 10+20 \\ 39+4 & -25+40 \end{bmatrix} = \begin{bmatrix} 38 & 30 \\ 43 & 15 \end{bmatrix}$$

Matrix subtraction follows the same principles used for matrix addition. If $\mathbf{B} = \mathbf{X} - \mathbf{Y}$, then $b_{ij} = x_{ij} - y_{ij}$. Thus the matrix **B** obtained by subtracting **Y** from **X** above is:

$$\mathbf{B} = \mathbf{X} - \mathbf{Y} = \begin{bmatrix} 40-(-2) & 10-20 \\ 39-4 & -25-40 \end{bmatrix} = \begin{bmatrix} 42 & -10 \\ 35 & -65 \end{bmatrix}$$

A.3.3 Matrix multiplication

Two matrices can be multiplied only if the number of columns in the first matrix equals the number of rows in the second. The order of the product matrix is equal to the number of rows of the first matrix by the number of columns in the second. Given that **C** = **AB**, then:

$$\mathbf{C} = c_{ij} = \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{z} a_{ik} b_{kj}$$

where m = number of columns in **B**, n = number of rows in **A** and z = number of rows in **B**. Let:

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & -1 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 2 & 5 \\ 4 & 3 \\ 6 & 1 \end{bmatrix}$$

Then **C** can be obtained as:

$C_{11} = 1(2) + 4(4) + -1(6) = 12$ (row 1 of **A** multiplied by column 1 of **B**)
$C_{21} = 2(2) + 5(4) + 0(6) = 24$ (row 2 of **A** multiplied by column 1 of **B**)
$C_{31} = 3(2) + 6(4) + 1(6) = 36$ (row 3 of **A** multiplied by column 1 of **B**)
$C_{12} = 1(5) + 4(3) + -1(1) = 16$ (row 1 of **A** multiplied by column 2 of **B**)
$C_{22} = 2(5) + 5(3) + 0(1) = 25$ (row 2 of **A** multiplied by column 2 of **B**)
$C_{32} = 3(5) + 6(3) + 1(1) = 34$ (row 3 of **A** multiplied by column 2 of **B**)

$$\mathbf{C} = \begin{bmatrix} 12 & 16 \\ 24 & 25 \\ 36 & 34 \end{bmatrix}$$

Note that **C** has order 3×2, where 3 equals the number of rows of **A** and 2 the number of columns in **B**. Also note that **AB** is not equal to **BA**, but **IA** = **AI** = **A** where **I** is an identity matrix. If **M** is the product of a scalar g and a matrix **B**, then $\mathbf{M} = b_{ij}g$; that is, each element of **M** equals the corresponding element in **B** multiplied by g.

A.3.4 Direct product of matrices

Given a matrix **G** of order n by m and **A** of order t by s, the direct product is:

$$\mathbf{G} \otimes \mathbf{A} = \begin{bmatrix} g_{11}\mathbf{A} & g_{12}\mathbf{A} \\ g_{21}\mathbf{A} & g_{22}\mathbf{A} \end{bmatrix}$$

The direct product is also known as the Kronecker product and is of the order nt by ms. For instance, assuming that:

$$\mathbf{G} = \begin{bmatrix} 10 & 5 \\ 5 & 20 \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 4 \\ 2 & 4 & 1 \end{bmatrix}$$

the Kronecker product is:

$$\mathbf{G} \otimes \mathbf{A} = \begin{bmatrix} 10 & 0 & 20 & 5 & 0 & 10 \\ 0 & 10 & 40 & 0 & 5 & 20 \\ 20 & 40 & 10 & 10 & 20 & 5 \\ 5 & 0 & 10 & 20 & 0 & 40 \\ 0 & 5 & 20 & 0 & 20 & 80 \\ 10 & 20 & 5 & 20 & 80 & 20 \end{bmatrix}$$

The Kronecker product is useful in multiple trait evaluations.

A.3.5 Matrix inversion

An inverse matrix is one which, when multiplied by the original matrix, gives an identity matrix as the product. The inverse of a matrix **A** is usually denoted as $\mathbf{A}^{-1}$ and, from the above definition, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, where **I** is an identity matrix. Only square matrices can be inverted and for a diagonal matrix the inverse is calculated simply as the reciprocal of the diagonal elements. For instance, the diagonal matrix **B** and its inverse are:

$$\mathbf{B} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 18 \end{bmatrix} \quad \text{and} \quad \mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{18} \end{bmatrix}$$

For a 2×2 matrix, the inverse is easy to calculate and is illustrated below. Let:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

First, calculate the determinant, which is the difference between the product of the two diagonal elements and the two off-diagonal elements ($a_{11}a_{22} - a_{12}a_{21}$). Secondly, the inverse is obtained by reversing the diagonal elements, multiplying the off-diagonal elements by –1 and dividing all elements by the determinant. Thus:

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

For instance, given that:

$$\mathbf{A} = \begin{bmatrix} 8 & 4 \\ 6 & 4 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{1}{(8)(4) - (6)(4)} \begin{bmatrix} 4 & -4 \\ -6 & 8 \end{bmatrix} = \begin{bmatrix} 0.50 & -0.50 \\ -0.75 & 1.00 \end{bmatrix}$$

Note that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1}$, as stated earlier. Calculating the inverse of a matrix becomes more cumbersome as the order increases and inverses are usually obtained using computer programs. The methodology has not been covered in this text. It is obvious from the above that an inverse of a non-diagonal matrix cannot be calculated if the determinant is equal to zero. A square matrix with a determinant equal to zero is said to be singular and does not have an inverse. A matrix with a non-zero determinant is said to be non-singular.

Note that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. The inverses of matrices may be required when solving linear equations. Thus, given the following linear equation:

$$\mathbf{Ab} = \mathbf{y}$$

premultiplying both sides by $\mathbf{A}^{-1}$ gives the vector of solutions **b** as:

$$\mathbf{b} = \mathbf{A}^{-1}\mathbf{y}$$

A.3.6 Rank of a matrix

The rank of a matrix is the number of linearly independent rows or columns. A square matrix with the rank equal to the number of rows or columns is said to be of full rank. In some matrices, some of the rows or columns are linear combination of other rows or columns, therefore the rank is less than the number of rows or columns. Such a matrix is not of full rank. Consider the following set of equations:

$$\begin{aligned} 3x_1 + 2x_2 + 1x_3 &= y_1 \\ 4x_1 + 3x_2 + 0x_3 &= y_2 \\ 7x_1 + 5x_2 + 1x_3 &= y_3 \end{aligned}$$

The third equation is the sum of the first and second equations; therefore the vector of solutions, $\mathbf{x}(\mathbf{x}' = [x_1\ x_2\ x_3])$, cannot be estimated due to the lack of information. In other words, if the system of equations were expressed in matrix notation as:

$$\mathbf{Dx} = \mathbf{y}$$

that is, as:

$$\begin{bmatrix} 3 & 2 & 1 \\ 4 & 3 & 0 \\ 7 & 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

a unique inverse does not exist for **D** because of the dependency in the rows. Only two rows are linearly independent in **D** and it is said to be of rank 2, usually written as $r(\mathbf{D}) = 2$. When a square matrix is not of full rank, the determinant is zero and hence a unique inverse does not exist.

A.3.7 Generalized inverses

While an inverse does not exist for a singular matrix, a generalized inverse can, however, be calculated. A generalized inverse for a matrix **D** is usually denoted as $\mathbf{D}^-$ and satisfies the expression:

$$\mathbf{D}\mathbf{D}^-\mathbf{D} = \mathbf{D}$$

Generalized inverses are not unique and may be obtained in several ways. One of the simple ways to calculate a generalized inverse of a matrix, say **D** in Section A.3.6, is to initially obtain a matrix **B** of full rank as a subset of **D**. Set all elements of **D** to zero. Calculate the inverse of **B** and replace the elements of **D** with corresponding elements of **B** and the result is $\mathbf{D}^-$. For instance, for the matrix **D** above, the matrix **B**, a full-rank subset of **D**, is:

$$\mathbf{B} = \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B}^{-1} = \begin{bmatrix} 3 & -2 \\ -4 & 3 \end{bmatrix}$$

Replacing elements of **D** with the corresponding elements of **B** after all elements of **D** have been set to zero gives $\mathbf{D}^-$ as:

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

A.3.8 Eigenvalues and eigenvectors

Eigenvalues are also referred to as characteristic or latent roots and are useful in simplifying multivariate evaluations when transforming data. The sum of the eigenvalues of a square matrix equals its trace (sum of the diagonal elements of a square matrix) and their product equals its determinant (Searle, 1982). For symmetric matrices, the rank equals the number of non-zero eigenvalues.

For a square matrix **B**, the eigenvalues are obtained by solving:

$$|\mathbf{B} - d\mathbf{I}| = 0$$

where the vertical lines denote finding the determinant.

With the condition specified in the above equation, **B** can be represented as:

$$\mathbf{BL} = \mathbf{LD}$$

$$\mathbf{B} = \mathbf{LDL}^{-1} \qquad [a.1]$$

where **D** is a diagonal matrix containing the eigenvalues of **B**, and **L** is a matrix of corresponding eigenvectors. The eigenvector (k) is found by solving:

$$(\mathbf{B} - d_k\mathbf{I})l_k = 0$$

where d_k is the corresponding eigenvalue.

For symmetric matrices **L** is orthogonal (that is, $\mathbf{L}^{-1} = \mathbf{L}'$; $\mathbf{LL}' = \mathbf{I} = \mathbf{L}'\mathbf{L}$); therefore, given that **B** is symmetric, [a.1] can be expressed as:

$$\mathbf{B} = \mathbf{LDL}'$$

Usually eigenvalues and eigenvectors are calculated by means of computer programs.

Appendix B: Fast Algorithms for Calculating Inbreeding Based on the L Matrix

In this appendix, two algorithms based on the **L** matrix for calculating inbreeding are discussed.

B.1 Meuwissen and Luo Algorithm

The algorithm given by Quaas (1976) involves the calculation of one column of **L** at a time. The algorithm requires $n(n+1)/2$ operations and computational time is proportional to n^2, where n is the size of the data set. It suffers from the disadvantage of not being readily adapted for updating when a new batch of animals is available without restoring a previously stored **L**. Meuwissen and Luo (1992) presented a faster algorithm which involves computing the elements of **L** row by row.

The fact that each row of **L** is calculated independently of other rows makes it suitable for updating. The row i of **L** for animal i gives the fraction of genes the animal derives from its ancestors. If s_i and d_i are the sire and dam of animal i, then $l_{is_i} = l_{id_i} = 0.5$. The ith row of **L** can be calculated by proceeding through a list of i's ancestors from the youngest to the oldest and updating continually as $l_{is_j} = l_{is_j} + 0.5l_{ij}$ and $l_{id_j} = l_{id_j} + 0.5l_{ij}$, where j is an ancestor of i. The fraction of genes derived from an ancestor is:

$$l_{ij} = \sum_{k \in \mathbf{P}_j} 0.5 l_{ik}$$

where $\mathbf{P}_j$ is a set of identities of the progeny of j. However $l_{ij} = 0$ only when k is not an ancestor of i or k is not equal to i. Thus, if **AN** is the set of identities of the number of ancestors of i, then:

$$l_{ij} = \sum_{k \in \mathbf{AN} \cap \mathbf{P}_j} 0.5 l_{ik}$$

that is, the summation of $0.5l_{ik}$ is over those k animals that are both ancestors of i and progeny of j. This forms the basis of the algorithm given below for the calculation of the row i of **L**, one row at a time. As each row of **L** is

calculated, its contribution to the diagonal elements of the relationship matrix (a_{ii}) is accumulated. Initially, set row i of **L** and a_{ii} to zero. The list of ancestors whose contributions to a_{ii} are yet to be included are added to the vector **AN** (if not already there) as each row of **L** is being processed.

The algorithm is:

$F_0 = -1$

For $i = 1, N$ (all rows of **L**):

$\mathbf{AN}_i = i$

$l_{ii} = 1$

$\mathbf{D}_{ii} = [0.5 - 0.25(F_{s_i} + F_{d_i})]$, if both parents are known; otherwise use appropriate formula (see Chapter 2, Section 2.2).

Do while $\mathbf{AN}_i$ is not empty.

$j = \max(\mathbf{AN}_i)$ (j = youngest animal in $\mathbf{AN}_i$)

If s_j is known, add s_j to $\mathbf{AN}_i$:

$l_{is_j} = l_{is_j} + 0.5 l_{ij}$

If d_j is known, add d_j to $\mathbf{AN}_i$:

$l_{id_j} = l_{id_j} + 0.5 l_{ij}$

$a_{ii} = a_{ii} + l_{ij}^2 + \mathbf{D}_{jj}$

Delete j from $\mathbf{AN}_i$.

End while:

$F_i = a_{ii} - 1$

B.1.1 Illustration of the algorithm

Using the pedigree in Table 2.1, the algorithm is illustrated for animals 1 and 5.

For animal 1:

$a_{11} = 0$

$\mathbf{AN}_1 = 1, \quad l_{11} = 1$

Since both parents are unknown:

$\mathbf{D}_{11} = 1$

Processing animals in $\mathbf{AN}_1$:

$j = \max(\mathbf{AN}_1) = 1$

Both parents of j are unknown:

$a_{11} = a_{11} + l_{11}^2 \mathbf{D}_{11} = (1^2)1 = 1$

Delete animal 1 from $\mathbf{AN}_1$; $\mathbf{AN}_1$ is now empty.

$\mathbf{F}_1 = 1 - 1 = 0$

For animal 5:

$a_{55} = 0$

$\mathbf{AN}_5 = 5, \quad l_{55} = 1$

$\mathbf{D}_{55} = 0.5$, since neither parent is inbred.

Processing animals in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 5$

Add sire and dam of 5 (animals 4 and 3) to AN_5:

$l_{54} = l_{54} + 0.5l_{55} = 0.5$

$l_{53} = l_{53} + 0.5l_{55} = 0.5$

$a_{55} = a_{55} + l_{55}^2\mathbf{D}_{55} = 1^2(0.5) = 0.5$

Delete animal 5 from $\mathbf{AN}_5$; animals 4 and 3 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 4$

Add sire of 4 (animal 1) to $\mathbf{AN}_5$:

$l_{51} = l_{51} + 0.5l_{54} = 0.25$

$a_{55} = a_{55} + l_{54}^2\mathbf{D}_{44} = 0.5 + (0.5)^2(0.75) = 0.6875$

Delete animal 4 from $\mathbf{AN}_5$; animals 3 and 1 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 3$

Since animal 1, the sire of j, is already in $\mathbf{AN}_5$, add only dam of 3 (animal 2) to $\mathbf{AN}_5$:

$l_{51} = l_{51} + 0.5l_{53} = 0.25 + (0.5)0.5 = 0.5$

$l_{52} = l_{52} + 0.5l_{53} = 0 + (0.5)0.5 = 0.25$

$a_{55} = a_{55} + l_{53}^2\mathbf{D}_{33} = 0.6875 + (0.5)^2 0.5 = 0.8125$

Delete animal 3 from $\mathbf{AN}_5$; animals 1 and 2 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 2$

Both parents are unknown:

$a_{55} = a_{55} + l_{52}^2\mathbf{D}_{22} = 0.8125 + (0.25)^2 1 = 0.875$

Delete animal 2 from $\mathbf{AN}_5$; animal 1 left in $\mathbf{AN}_5$.
Next animal in $\mathbf{AN}_5$:

$j = \max(\mathbf{AN}_5) = 1$

Both parents are unknown:

$a_{55} = a_{55} + l_{51}^2\mathbf{D}_{11} = 0.875 + (0.5)^2 1 = 1.125$

Delete 1 from $\mathbf{AN}_5$; $\mathbf{AN}_5$ is empty.

$\mathbf{F}_5 = 1.125 - 1 = 0.125$

which is the same inbreeding coefficient as that obtained for animal 5 in Section 2.1.

B.2 Modified Meuwissen and Luo Algorithm

The approach of Meuwissen and Luo given above was modified by Quaas (1995) to improve its efficiency. The disadvantage of the above method is that, while calculating a row of **L** at a time (Henderson, 1976), it is, however, accumulating diagonal elements of **A**, as in Quaas (1976), and this necessitates tracing the entire pedigree for *i*, but what is really needed is only the common ancestors. Thus a more efficient approach is to accumulate $a_{s_i d_i}$ as $\Sigma_k l_{s_i k} l_{d_i k} \mathbf{D}_{kk}$ *(Henderson, 1976) and calculate* F_i as $0.5a_{s_i d_i} = \Sigma_k l_{s_i k} l_{d_i k}$ *(0.5*$\mathbf{D}_{kk}$). Instead of computing the *i*th row of **L**, only the non-zero elements in the rows for the sire and dam of *i* are calculated. Quaas (1995) suggested setting up a separate ancestor list ($\mathbf{AS}_{si}$) for s_i and another ($\mathbf{AD}_{d_i}$) for d_i; then $\mathbf{F}_i = 0.5a_{s_i d_i} = \sum_k \varepsilon_{s_i} U_{d_i} l_{s_i k} l_{d_i k} (0.5\mathbf{D}_{kk})$.

Similarly to the approach of Meuwissen and Luo (1992), the two lists can be set up simultaneously while processing the *i*th animal by continually adding the parents of the next youngest animal in either list to the appropriate list. If the next youngest in each list is the same animal, say *k*, then it has a common ancestor and $\mathbf{F}_i$ is updated as $\mathbf{F}_i = \mathbf{F}_i + l_{s_i k} l_{d_i k} (0.5\mathbf{D}_{kk})$. When ancestors of one of the parents have been processed, the procedure can be stopped, and it is not necessary to search both lists completely. The algorithm for this methodology is as follows:

$\mathbf{F}_0 = -1$

For $i = 1, N$:

$\mathbf{F}_i = 0$

If s_i is known, add s_i to $\mathbf{AS}_{s_i}$, $l_{s_i s_i} = 1$.
If d_i is known, add d_i to $\mathbf{AD}_{d_i}$, $l_{d_i d_i} = 1$.
Do while $\mathbf{AS}_{s_i}$ not empty and $\mathbf{AD}_{d_i}$ not empty.

$j = \max(\mathbf{AS}_{s_i}), \quad k = \max(\mathbf{AD}_{d_i})$

If $j > k$, then (next youngest ancestor j is in $\mathbf{AS}_{s_i}$)
If s_j is known, add s_j to $\mathbf{AS}_{s_i}: l_{s_i s_j} = l_{s_i sj} + 0.5 l_{s_i j}$
If d_j is known, add d_j to $\mathbf{AS}_{s_i}: l_{s_i d_j} = l_{s_i d_j} + 0.5 l_{s_i j}$
Delete j from $\mathbf{AS}_{s_i}$.
If $k > j$, then (next youngest k is in $\mathbf{AD}_{d_i}$)
If s_k is known, add s_k to $\mathbf{AD}_{d_i}: l_{d_i s_k} = l_{d_i s_k} + 0.5 l_{d_i k}$
If d_k is known, add d_k to $\mathbf{AD}_{d_i}: l_{d_i d_k} = l_{d_i d_k} + 0.5 l_{d_i k}$
Delete k from $\mathbf{AD}_{d_i}$.
Else (next youngest ancestor $j = k$ is a common ancestor)
If s_j is known, add s_j to $\mathbf{AS}_{s_i}: l_{s_i s_j} = l_{s_i s_j} + 0.5 l_{s_i j}$
add s_j to $\mathbf{AD}_{d_i}: l_{d_i s_j} = l_{d_i s_j} + 0.5 l_{d_i j}$
If d_j is known, add d_j to $\mathbf{AS}_{s_i}: l_{s_i d_j} = l_{s_i d_j} + 0.5 l_{s_i j}$
add d_j to $\mathbf{AD}_{d_i}: l_{d_i d_j} = l_{d_i d_j} + 0.5 l_{d_i j}$

$F_i = F_i + l_{s_i j} l_{d_i j} 0.5(\mathbf{D}_{jj})$

Delete j from $\mathbf{AS}_{s_i}$ and $\mathbf{AD}_{d_i}$
End if
End while
End do

B.2.1 Illustration of the algorithm

Using the pedigree in Table 2.1, the algorithm is illustrated for animal 5, which is inbred.

For animal 5:

$\mathbf{F}_5 = 0$

Both parents known, $s = 4$ and $d = 3$.

Add 4 to $\mathbf{AD}_4$: $l_{44} = 0.5$

Add 3 to $\mathbf{AD}_3$: $l_{33} = 0.5$

Processing animals in $\mathbf{AS}_4$ and $\mathbf{AD}_3$:

$j = 4, \quad k = 3$

$j > k$ therefore.

Add sire of 4, animal 1, to $\mathbf{AS}_4$: $l_{41} = l_{41} + 0.5l_{44} = 0.5$

Delete animal 4 from $\mathbf{AS}_4$.

Next animals in $\mathbf{AS}_4$ and $\mathbf{AD}_3$:

$j = 1, \quad k = 3$

$k > j$ therefore.

Add sire of 3, animal 1, to $\mathbf{AD}_3$: $l_{31} = l_{31} + 0.5l_{33} = 0.5$

Add dam of 3, animal 2, to $\mathbf{AD}_3$: $l_{32} = l_{32} + 0.5l_{33} = 0.5$

Delete 3 from $\mathbf{AD}_3$.

Next animals in $\mathbf{AS}_4$ and $\mathbf{AD}_{3:}$

$j = 1, \quad k = 2$

$k > j$

Both parents of 2 are unknown.

Delete 2 from $\mathbf{AD}_3$.

Next animals in $\mathbf{AS}_4$ and $\mathbf{AD}_3$:

$j = 1, \quad k = 1$

$j = k$

Both parents are unknown.

$$F_5 = F_5 + l_{41}l_{31}0.5(\mathbf{D}_{11}) = 0.5(0.5)(0.5)(1) = 0.125$$

which is the same inbreeding coefficient as that obtained from the algorithm in B.1.

Appendix C

C.1 Outline of the Derivation of the Best Linear Unbiased Prediction (BLUP)

Consider the following linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e} \qquad [c.1]$$

where the expectations are:

$$E(\mathbf{y}) = \mathbf{Xb}; \quad E(\mathbf{a}) = E(\mathbf{e}) = 0$$

and:

$$\text{var}(\mathbf{a}) = \mathbf{A}\sigma_a^2 = \mathbf{G}, \quad \text{var}(\mathbf{e}) = \mathbf{R} \quad \text{and} \quad \text{cov}(\mathbf{a}, \mathbf{e}) = \text{cov}(\mathbf{e}, \mathbf{a}) = 0$$

Then, as shown in Section 3.1:

$$\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}, \quad \text{cov}(\mathbf{y}, \mathbf{a}) = \mathbf{ZG} \quad \text{and} \quad \text{cov}(\mathbf{y}, \mathbf{e}) = \mathbf{R}$$

The prediction problem involves both **b** and **a**. Suppose we want to predict a linear function of **b** and **a**, say $\mathbf{k}'\mathbf{b} + \mathbf{a}$, using a linear function of **y**, say $\mathbf{L}'\mathbf{y}$, and $\mathbf{k}'\mathbf{b}$ is estimable. The predictor $\mathbf{L}'\mathbf{y}$ is chosen such that:

$$E(\mathbf{L}'\mathbf{y}) = E(\mathbf{k}'\mathbf{b} + \mathbf{a})$$

that is, it is unbiased and the prediction error variance (PEV) is minimized (Henderson, 1973). Now PEV (Henderson, 1984) is:

$$\begin{aligned} \text{PEV} &= \text{var}(\mathbf{L}'\mathbf{y} - \mathbf{k}'\mathbf{b} + \mathbf{a}) \\ &= \text{var}(\mathbf{L}'\mathbf{y} - \mathbf{a}) \\ &= \mathbf{L}'\,\text{var}(\mathbf{y})\mathbf{L} + \text{var}(\mathbf{a}) - \mathbf{L}'\,\text{cov}(\mathbf{y}, \mathbf{a}) - \text{cov}(\mathbf{a}, \mathbf{y})\mathbf{L} \\ &= \mathbf{L}'\mathbf{VL} + \mathbf{G} - \mathbf{L}'\mathbf{ZG} - \mathbf{ZG}'\mathbf{L} \end{aligned} \qquad [c.2]$$

Minimizing PEV subject to $\mathbf{E}(\mathbf{L}'\mathbf{y}) = \mathbf{E}(\mathbf{k}'\mathbf{b} + \mathbf{a})$ and solving (see Henderson, 1973, 1984, for details of derivation) gives:

$$\mathbf{L}'\mathbf{y} = \mathbf{k}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})$$

Let $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}\mathbf{V}^{-1}\mathbf{y}$, the generalized least square solution for $\mathbf{b}$, then the predictor can be written as:

$$\mathbf{L}'\mathbf{y} = \mathbf{k}'\hat{\mathbf{b}} + \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \qquad [c.3]$$

which is the BLUP of $\mathbf{k}'\mathbf{b} + \mathbf{a}$.

Note that, if $\mathbf{k}'\mathbf{b} = 0$, then:

$$\mathbf{L}'\mathbf{y} = \text{BLUP}(\mathbf{a}) = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \qquad [c.4]$$

which is equivalent to the selection index. Thus BLUP is the selection index with the generalized least-square solution (GLS) of $\mathbf{b}$ substituted for $\mathbf{b}$.

C.2 Proof that $\hat{b}$ and $\hat{a}$ from Mixed Model Equations are the GLS of b and BLUP of a, respectively

In computation terms, the use of equation [c.3] to obtain BLUP of $\mathbf{k}'\mathbf{b} + \mathbf{a}$ is not feasible because the inverse of $\mathbf{V}$ is required. Henderson (1950) formulated the mixed model equations (MME) which are suitable for calculating solutions for $\mathbf{b}$ and $\mathbf{a}$, and showed later that $\mathbf{k}'\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$, where $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ are solutions from the MME, are the best linear unbiased estimator (BLUE) of $\mathbf{k}'\mathbf{b}$ and BLUP of $\mathbf{a}$, respectively.

The usual mixed model equations for [c.1] are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}' \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad [c.5]$$

The proof that $\hat{\mathbf{b}}$ from the MME is the GLS of $\mathbf{b}$ and therefore $\mathbf{k}'\hat{\mathbf{b}}$ is the BLUE of $\mathbf{k}'\mathbf{b}$ was given by Henderson *et al.* (1959). From the second row of [c.5]:

$$(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{a}} = \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$$

$$\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \qquad [c.6]$$

From the first row of [c.5]:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\mathbf{b}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{a}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$$

Substituting solution for $\hat{\mathbf{a}}$ into the above equation gives:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\mathbf{b}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y}$$

where $\mathbf{W} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}$

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\mathbf{b}} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}(\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}$$

$$\mathbf{X}'(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1})\mathbf{y}$$

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

with $\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}$

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad [c.7]$$

It can be shown that:

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\mathbf{W}\mathbf{Z}'\mathbf{R}^{-1}$$

By premultiplying the right-hand side by **V** and obtaining an identity matrix (Henderson *et al.*, 1959):

$$\begin{aligned}\mathbf{V}[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{ZWZ'R}^{-1}] &= (\mathbf{R} + \mathbf{ZGZ'})(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{ZWZ'R}^{-1}) \\ &= \mathbf{I} + \mathbf{ZGZ'R}^{-1} - \mathbf{ZWZ'R}^{-1} - \mathbf{ZGZ'R}^{-1}\mathbf{ZWZ'R}^{-1} \\ &= \mathbf{I} + \mathbf{ZGZ'R}^{-1} - \mathbf{Z}(\mathbf{I} + \mathbf{GZ'RZ})\mathbf{WZ'R}^{-1} \\ &= \mathbf{I} + \mathbf{ZGZ'R}^{-1} - \mathbf{ZG}(\mathbf{G}^{-1} + \mathbf{Z'RZ})\mathbf{WZ'R}^{-1} \\ &= \mathbf{I} + \mathbf{ZGZ'R}^{-1} - \mathbf{ZG}(\mathbf{W}^{-1})\mathbf{WZ'R}^{-1} \\ &= \mathbf{I} + \mathbf{ZGZ'R}^{-1} - \mathbf{ZGZ'R}^{-1} \\ &= \mathbf{I}\end{aligned}$$

Thus the solution for **b** from the MME is equal to the generalized least solution for **b** in equation [c.3].

The proof that $\hat{\mathbf{a}}$ from the MME is equal to $\mathbf{GZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$ in equation [c.3] was given by Henderson (1963). Replace $\mathbf{V}^{-1}$ in $\mathbf{GZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$ by $\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{ZWZ'R}^{-1}$. Thus:

$$\begin{aligned}\mathbf{GZ'V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) &= \mathbf{GZ'}(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{ZWZ'R}^{-1})(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= \mathbf{G}(\mathbf{Z'R}^{-1} - \mathbf{Z'R}^{-1}\mathbf{ZWZ'R}^{-1})(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= \mathbf{G}(\mathbf{I} - \mathbf{Z'R}^{-1}\mathbf{ZW})\mathbf{Z'R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= \mathbf{G}(\mathbf{W}^{-1} - \mathbf{Z'R}^{-1}\mathbf{Z})\mathbf{WZ'R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= \mathbf{G}((\mathbf{Z'RZ} + \mathbf{G}^{-1}) - \mathbf{Z'R}^{-1}\mathbf{Z})\mathbf{WZ'R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= \mathbf{G}(\mathbf{Z'R}^{-1}\mathbf{Z} + \mathbf{I} - \mathbf{GZ'R}^{-1}\mathbf{Z})\mathbf{WZ'R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= (\mathbf{I})\mathbf{WZ'R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \\ &= \mathbf{WZ'R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \hat{\mathbf{a}} \qquad \text{(see equation [c.6])}\end{aligned}$$

Thus the BLUP of $\mathbf{k'b} + \mathbf{a} = \mathbf{k'}\hat{\mathbf{b}} + \hat{\mathbf{a}}$, where $\hat{\mathbf{b}}$ and $\hat{\mathbf{a}}$ are solutions to the MME.

C.3 Deriving the Equation for Progeny Contribution (PC)

Considering an individual *i* that has one record with both sire (*s*) and dam (*d*) known, the mixed model equations for the three animals can be written (assuming the sire and dam are ancestors with unknown parents) as:

$$\begin{bmatrix} u_{ss}\alpha & u_{sd}\alpha & u_{si}\alpha \\ u_{ds}\alpha & u_{dd}\alpha & u_{di}\alpha \\ u_{is}\alpha & u_{id}\alpha & 1+u_{ii}\alpha \end{bmatrix} \begin{bmatrix} \hat{a}_s \\ \hat{a}_d \\ \hat{a}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1'y \end{bmatrix} \qquad \text{[c.8]}$$

Where the u terms are elements of $\mathbf{A}^{-1}$.
From equation [c.8], the equation for solution of the sire is:

$$u_{ss}\alpha\hat{a}_s = 0 - u_{sd}\alpha\hat{a}_d - u_{si}\alpha\hat{a}_i$$
$$u_{ss}\alpha\hat{a}_s = \text{PC}$$

with:

$$\text{PC} = 0 - u_{sd}\alpha\hat{a}_d - u_{si}\alpha\hat{a}_i$$

When the mate is known:

$$\text{PC} = 0 - \tfrac{1}{2}\alpha\hat{a}_d + (1)\alpha\hat{a}_i$$

$$\text{PC} = \alpha(\hat{a}_i - \tfrac{1}{2}\hat{a}_d) = 0.5\alpha(2\hat{a}_i - \hat{a}_d)$$

In general, assuming sire s has k progeny:

$$\text{PC}_s = 0.5\alpha \sum_k u_{prog}(2\hat{a}_i - \hat{a}_m) \Big/ \sum_k u_{prog}$$

where u_{prog} is 1 when the mate of s is known or $\frac{2}{3}$ when the mate is not known.

Appendix D: Methods for Obtaining Approximate Reliability for Genetic Evaluations

D.1 Computing Approximate Reliabilities for an Animal Model

Presented below is a method published by Meyer (1989) for obtaining approximate values of repeatability or reliability for genetic evaluations from an animal model and was used to estimate reliabilities in the national dairy evaluation system in Canada in the 1990s. The reliability for each animal is derived from the corresponding diagonal element in the mixed model equations (MME), adjusting for selected off-diagonal coefficients. For instance, the section of the coefficient matrix (**C**) pertaining to an animal i with parents f and m and with a record in a subclass h of a major fixed effect such as herd–year–season (HYS) could be represented as:

$$\begin{bmatrix} c_{ii} & -\alpha & -\alpha & 1 \\ -\alpha & c_{ff} & 0.5\alpha & 0 \\ -\alpha & 0.5\alpha & c_{mm} & 0 \\ 1 & 0 & 0 & n_h \end{bmatrix}$$

where n_h is the number of records in subclass h of the major fixed effect and $\alpha = \sigma_e^2/\sigma_a^2$. If this were the complete coefficient matrix for this animal, $\mathbf{C}^{-1}$ and hence true reliability could be obtained using partition matrix results. Thus the coefficient c^{ii} can be calculated as the reciprocal of the ith diagonal element of **C** after absorbing all other rows and columns.

For animal i:

$$c^{ii} = (c_{ii} - 1/n_h - \alpha^2(c_{ff} + c_{mm} - \alpha)/(c_{ff}c_{mm} - \tfrac{1}{4}\alpha^2))^{-1}$$

and for parent f:

$$c^{ff} = (c_{ff} - Q - (\tfrac{1}{2}\alpha - Q)^2/(c_{mm} - Q))^{-1}$$

with:

$$Q = \alpha^2(c_{ii} - 1/n_h)^{-1}$$

Exchange m for f for parent m.

However, if there are other off-diagonals for animal *i*, the above equations will yield approximations of the diagonal elements of **C** and hence reliability. Based on the above principle of forming and inverting the submatrix of the MME for each animal, Meyer outlined three steps for calculating approximate reliabilities, which was similar to the true reliabilities from her simulation study. These steps are as follows:

1. Diagonal elements (D) of animals with records are adjusted for the effect of major fixed effects, such as HYS. Thus:

$$D_{1i} = D_{0i} - 1/n_h$$

and for animals without records:

$$D_{1i} = D_{0i}$$

where D_{0i} is the diagonal element for animal i in the MME and in general, its composition depending on the amount of information available on the animal, is:

$$D_{0i} = x_i + n_i\alpha + n_{1i}\alpha/3 + n_{2i}\alpha/2$$

where $x_i = 1$ if the animal has a record and otherwise it is zero, n_i equals 1 or $\frac{4}{3}$ or 2 if none or one or both parents are known, n_{1i} and n_{2i} are number of progeny with one or both parents known, respectively.

2. Diagonal elements for parents (f and m) are adjusted for the fact that the information on their progeny is limited.

For each progeny i with only one parent known, adjust the diagonal element of the parent f as:

$$D_{2f} = D_{1f} - \alpha^2(\tfrac{4}{9}D_{1i}^{-1})$$

and if both parents are known adjust the diagonal of parent f as:

$$D_{2f} = D_{1f} - \alpha^2 D_{1i}^{-1}$$

Replace subscript f with m for the other parent. For animals that are not parents:

$$D_{2i} = D_{1i}$$

3. Adjustment of progeny diagonals for information on parents.

This involves initially unadjusting the diagonals of the parents for the contribution of the ith progeny in question by reversing step 2 before adjusting progeny diagonals for parental information. If only one parent f is known, the diagonal is unadjusted initially as:

$$D_{2f}^{*} = D_{2f} + \alpha^2(\tfrac{4}{9}D_{2i}^{-1})$$

and if both parents are known as:

$$D_{2f}^{*} = D_{2f} - \alpha^2 D_{2i}^{-1}$$

for parent f. Exchange m for f in the above equation to calculate for parent m.

Adjustment of progeny i diagonal then is:

$$D_{3i} = D_{2i} - \alpha^2(\tfrac{4}{9}D_{2f}^{*-1})$$

if only parent f is known and:

$$D_{3i} = D_{2i} - \alpha^2((D^*_{2f} + D^*_{2m} - \alpha)/(D^*_{2f}D^*_{2m} - \tfrac{1}{4}\alpha^2))$$

when both parents f and m are known.

For animals with unknown parents:

$$D_{3i} = D_{2i}$$

Reliability for progeny i is calculated as:

$$r^2 = \text{const.}(1 - \alpha D^{-1}_{3i})$$

where const. is a constant of between 0.90 to 0.95 from Meyer simulation studies, which gave the best estimate of r^2.

D.2 Computing Approximate Reliabilities for Random Regression Models

Meyer and Tier (2003) extended the method in Appendix D.1 to estimate reliabilities for multivariate and random regression models. They outlined several steps.

1. Determine value of observation for an animal

Compute the diagonal block ($\mathbf{D}_i$) for animal i in the MME, based on the information from the data, as:

$$\mathbf{D}_i = \mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{Z}_i$$

However, to account for the limited subclass sizes of contemporary group effect, such as herd–test–day in dairy cattle, $\mathbf{D}_i$ can better be calculated as:

$$\mathbf{D}_i = \mathbf{Z}'_i(\mathbf{R}^{-1}_i - \mathbf{R}^{-1}_i(\mathbf{S}^{-1}_i)\mathbf{R}^{-1}_i)\mathbf{Z}_i$$

where $\mathbf{Z}_i$ and $\mathbf{R}^{-1}_i$ are submatrices of $\mathbf{Z}$ and $\mathbf{R}^{-1}$ for animal i and $\mathbf{S}_i$ is the block of the coefficient matrix pertaining to the contemporary groups of which animal i is a member. Then the permanent environmental (**pe**) effects are also absorbed into the block corresponding to animal genetic effects:

$$\mathbf{D}^*_i = \mathbf{D}_i - \mathbf{Z}'_i\mathbf{R}^{-1}_i\mathbf{Q}_i(\mathbf{Q}'_i\mathbf{R}^{-1}_i\mathbf{Q}_i + \mathbf{P}^{-1})\mathbf{Q}'_i\mathbf{R}^{-1}_i\mathbf{Z}_i$$

where $\mathbf{Q}_i$ is a submatrix of the matrix $\mathbf{Q}$ defined in Section 7.2. Limited subclass effects of **pe** can be accounted for by using weights $w_m = (n_m - k)/n_m \leq 1$, for the mth record, with n_m the size of the subclass to which the record belongs and k the number of 'repeated' records it has in that subclass. Then $\mathbf{R}_i$ in the above equation is replaced with $\mathbf{R}^*_i = \text{Diag}(w_m\alpha^2_e)$.

2. Value of records on descendants

In this second step, the contributions from progeny and other descendants are accumulated for each animal, processing the pedigree from the

youngest to the oldest. Let $\mathbf{E}_i$ be the block of contributions for animal i that has n_i progeny. Then:

$$\mathbf{E}_i = \tfrac{1}{3}\mathbf{G}^{-1} - \tfrac{4}{9}\mathbf{G}^{-1}\left(\mathbf{D}_i^* + \sum_{k=1}^{n_i}\mathbf{E}_k + \tfrac{4}{3}\mathbf{G}^{-1}\right)^{-1}\mathbf{G}^{-1}$$

This block is accumulated for both sire and dam of the ith animal. This equation can be derived by assuming that each progeny has only one parent known and that the parent has no other information; then the MME are set up for the animal and the parent and the equations for the animal are absorbed in those of the parent. The above equation will give an overestimate of the individual's contribution to its parents if it is in a contemporary group with many of its half-sibs. This can be discounted by weighting contributions with a factor dependent on the proportion of sibs in a subclass. Let $\mathbf{H}_i$ be a diagonal matrix of weights $w_m < 1$, with $w_m = \sqrt{(n_m - s_m)/n_m}$, where n_m is the total number of records in the subclass for trait m and s_m the total number of sibs of animal i in the subclass. Calculate $\mathbf{D}_i^{**} = \mathbf{H}_i\mathbf{D}_i^*\mathbf{H}_i$ and then replace $\mathbf{D}_i^*$ with $\mathbf{D}_i^{**}$.

3. Value of records on ancestors

In the third step, contributions from parents, ancestors and collateral relatives are accumulated for each animal, processing the pedigree from oldest to youngest. However, in step two, contributions from descendants were accumulated for all animals; hence $\mathbf{E}_j$ for parent j of animal i includes the contribution from animal i. The contributions of animal i have to be removed from $\mathbf{E}_j$ to avoid double counting. The corrected block is:

$$\mathbf{E}_j^* = \tfrac{1}{3}\mathbf{G}^{-1} - \tfrac{4}{9}\mathbf{G}^{-1}(-\mathbf{E}_i + \mathbf{F}_j + \tfrac{4}{3}\mathbf{G}^{-1})^{-1}\mathbf{G}^{-1}$$

where $\mathbf{F}_j$ is the sum of contributions from all sources of information for parent j. As parents are processed in the pedigree before progeny, $\mathbf{F}_j$ is always computed before the contribution of parent j to animal i is required. For animal i, $\mathbf{F}_i$ is:

$$\mathbf{F}_i = \sum_{j=1}^{t_i}\mathbf{E}_j^* + \mathbf{D}_i^* + \sum_{k=1}^{n_i}\mathbf{E}_k$$

with $t_i = 0$, 1 or 2 denoting the number of parents of animal i that are known and n_i, the number of progeny for animal i.

The matrix $\mathbf{T}_i$ of the approximate prediction error variance (PEV) and prediction error covariance (PEC) for the genetic effects for animal i is:

$$\mathbf{T}_i = (\mathbf{F}_i + \mathbf{G})^{-1}$$

The approximate reliability for a linear function of estimated breeding values for animal i then is:

$$r_i^2 = 1 - \mathbf{k}'\mathbf{T}_i\mathbf{k}/\mathbf{k}'\mathbf{G}\mathbf{k}$$

with $\mathbf{k}$ calculated as described in Chapter 7, Section 7.2.4.

Appendix E

E.1 Canonical Transformation: Procedure to Calculate the Transformation Matrix and its Inverse

The simplification of a multivariate analysis into n single trait analyses using canonical transformation involves transforming the observations of several correlated traits into new uncorrected traits (Section 5.2). The transformation matrix **Q** can be calculated by the following procedure, which has been illustrated by the **G** and **R** matrices for Example 5.1 in Section 5.1.2.

The **G** and **R** matrices are, respectively:

$$\begin{matrix} \\ \text{WWG} \\ \text{PWG} \end{matrix}\begin{bmatrix} 20 & 18 \\ 18 & 40 \end{bmatrix}$$

$$\begin{matrix} \\ \text{WWG} \\ \text{PWG} \end{matrix}\begin{bmatrix} 40 & 11 \\ 11 & 30 \end{bmatrix}$$

where WWG is the pre-weaning gain and PWG is the post-weaning gain.

1. Calculate the eigenvalues (**B**) and eigenvectors (**U**) of **R**:

$$\mathbf{R} = \mathbf{UBU}'$$

For the above **R**:

$$\mathbf{B} = \text{diag}(47.083, 22.917)$$

and:

$$\mathbf{U} = \begin{bmatrix} 0.841 & -0.541 \\ 0.541 & 0.841 \end{bmatrix}$$

2. Calculate **P** and **PGP**′:

$$\mathbf{P} = \mathbf{U}\sqrt{\mathbf{B}^{-1}}\,\mathbf{U}'$$

$$= \begin{bmatrix} 0.1642 & -0.0288 \\ -0.0288 & 0.1904 \end{bmatrix}$$

and:

$$\mathbf{PGP}' = \begin{bmatrix} 0.403 & 0.264 \\ 0.264 & 1.269 \end{bmatrix}$$

3. Calculate the eignenvalues (**W**) and eigenvectors (**L**) of **PGP**′:

$$\mathbf{PGP}' = \mathbf{LWL}'$$

$$\mathbf{W} = \text{diag}(0.3283,\ 1.3436)$$

$$\mathbf{L} = \begin{bmatrix} 0.963 & 0.271 \\ -0.271 & 0.963 \end{bmatrix}$$

4. The transformation matrix **Q** can be obtained as:

$$\mathbf{Q} = \mathbf{L}'\mathbf{P}$$

$$\mathbf{Q} = \begin{bmatrix} 0.1659 & -0.0792 \\ 0.0168 & 0.1755 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}^{-1} = \begin{bmatrix} 5.7651 & 2.6006 \\ -0.5503 & 5.4495 \end{bmatrix}$$

E.2 Canonical Transformation with Missing Records and Same Incidence Matrices

Ducrocq and Besbes (1993) presented a methodology for applying canonical transformation when all effects in the model affect all traits but there are missing traits for some animals. The principles of the methodology are briefly discussed and illustrated by an example.

Let **y**, the vector of observations, be partitioned as $\mathbf{y}' = [\mathbf{y}_v, \mathbf{y}_m]$ and $\mathbf{u} = [\mathbf{b}', \mathbf{a}']$, where $\mathbf{y}_v$ and $\mathbf{y}_m$ are vectors of observed and missing records, respectively, **b** is the vector of fixed effects and **a** is the vector of random effects. Assuming that the distribution of **y** given **u** is multivariate normal, Ducrocq and Besbes (1993) showed that the following expectation maximization (EM) algorithm gives the same solutions for **a** and **b** as when the usual multivariate MME are solved:

E step: at iteration k, calculate $\hat{\mathbf{y}}^{[k]} = \text{E}[\mathbf{y} \mid \mathbf{y}_v, \hat{\mathbf{u}}^{[k]}]$

M step: calculate $\hat{\mathbf{u}}^{[k+1]}$ = BLUE and BLUP solutions of **b** and **a**, respectively, given $\hat{\mathbf{y}}^{[k]}$

The E step implies doing nothing to observed records but replacing the missing observations by their expectation given the current solutions for **b** and **a** and the observed records. The equation for the missing records for animal i is:

$$\hat{\mathbf{y}}_{im}^{[k]} = \mathbf{x}'_{im}\mathbf{b}^{[k]} + \hat{\mathbf{a}}_{im}^{[k]} + \hat{\mathbf{e}}_{im}^{[k]} \qquad \text{[e.1]}$$

If $\mathbf{X}$ is the matrix that relates fixed effects to animals, $\mathbf{x}'_{im}$ denotes the row of $\mathbf{X}$ corresponding to missing records for animal i and $\hat{\mathbf{e}}_{im}^{[k]}$ is the regression of the residuals of missing records on the current estimates of the residuals for observed traits. Thus:

$$\hat{\mathbf{e}}_{im}^{[k]} = \mathrm{E}[\mathbf{e}_{im} \mid \mathbf{y}_{iv}, \mathbf{u} = \hat{\mathbf{u}}^{[k]}] = \mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}[\mathbf{y}_{iv} - \mathbf{x}'_{iv}\mathbf{b}^{[k]} - \hat{\mathbf{a}}_{iv}^{[k]}]$$

where $\mathbf{R}_{mv}$ and $\mathbf{R}_{vv}$ are submatrices obtained through partitioning of $\mathbf{R}$, the residual covariance matrix. $\mathbf{R}_{vv}$ represents the residual variance of observed traits and $\mathbf{R}_{mv}$ is the covariance between missing traits and observed traits. If three traits are considered, for example, and trait 2 is missing for animal i, then $\mathbf{R}_{vv}$ is the submatrix obtained by selecting in $\mathbf{R}$ the elements at intersection of rows 1 and 3 and columns 1 and 3. The submatrix $\mathbf{R}_{mv}$ is the element at the intersection of row 2 and columns 1 and 3. Once the missing observations have been estimated, records are now available on all animals and the analysis can be carried out as usual, applying canonical transformation as when all records are observed.

The application of the method in genetic evaluation involves the following steps at each iteration k, assuming $\mathbf{Q}$ is the transformation matrix to canonical scale and $\mathbf{Q}^{-1}$ the back-transforming matrix:

1. For each animal i with missing observations:
(1a) calculate $\hat{\mathbf{y}}_{im}^{[k]}$, given $\mathbf{b}^{[k]}$ and $\hat{\mathbf{a}}^{[k]}$ using [e.1];
(1b) transform $\hat{\mathbf{y}}_i$ to the canonical scale: $\hat{\mathbf{y}}_i^* = \mathbf{Q}\hat{\mathbf{y}}_i$.
2. Solve the mixed model equations to obtain solutions in the canonical scale: $\mathbf{b}^{*[k+1]}$ and $\hat{\mathbf{a}}^{*[k+1]}$.
3. Back-transform using $\mathbf{Q}^{-1}$ to obtain $\hat{\mathbf{b}}^{[k+1]}$ and $\hat{\mathbf{a}}^{[k+1]}$.
4. If convergence is not achieved, go to 1.

Ducrocq and Besbes (1993) showed that it is possible to update $\mathbf{y}$ (step 1) without back-transforming to the original scale (step 3) in each round of iteration. Suppose that the vector of observations for animal i with missing records, $\mathbf{y}_i$, is ordered such that observed records precede missing values: $\mathbf{y}'_i = [\mathbf{y}'_{iv}, \mathbf{y}'_{im}]$, and rows and columns of $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ are ordered accordingly. Partition $\mathbf{Q}$ as $(\mathbf{Q}_v \mid \mathbf{Q}_m)$ and $\mathbf{Q}^{-1}$ as:

$$\mathbf{Q}^{-1} = \begin{bmatrix} \mathbf{Q}^v \\ \mathbf{Q}^m \end{bmatrix}$$

then from [e.1], the equation for $\mathbf{Q}\hat{\mathbf{y}}_i$ or $\hat{\mathbf{y}}_i^*$ (see 1b) is:

$$\hat{\mathbf{y}}_i^* = \mathbf{Q}_v\mathbf{y}_{iv} + \mathbf{Q}_m[\mathbf{x}'_{im}\hat{\mathbf{b}}^{[k]} + \hat{\mathbf{a}}_{im}^{[k]} + \mathbf{R}_{mv}\mathbf{R}_{vv}^{-1}(\mathbf{y}_{iv} - \mathbf{x}'_{iv}\hat{\mathbf{b}}^{[k]} - \hat{\mathbf{a}}_{iv}^{[k]})] \quad [e.2]$$

However:

$$\begin{bmatrix} \hat{\mathbf{b}}_{iv} \\ \hat{\mathbf{b}}_{im} \end{bmatrix} = \mathbf{Q}^{-1}\hat{\mathbf{b}}^* = \begin{bmatrix} \mathbf{Q}^v\hat{\mathbf{b}}^* \\ \mathbf{Q}^m\hat{\mathbf{b}}^* \end{bmatrix}$$

and a similar expression exists for $\hat{\mathbf{a}}$. Substituting these values for $\mathbf{b}$ and $\hat{\mathbf{a}}$ in equation [e.2]:

$$\hat{\mathbf{y}}_i^* = (\mathbf{Q}_v + \mathbf{Q}_m \mathbf{R}_{mv} \mathbf{R}_{vv}^{-1}) \mathbf{y}_{iv} + (\mathbf{Q}_m \mathbf{Q}^m - \mathbf{Q}_m \mathbf{R}_{mv} \mathbf{R}_{vv}^{-1} \mathbf{Q}^v)(\mathbf{x}_i' \hat{\mathbf{b}}^{*[k]} + \hat{\mathbf{a}}^{*[k]})$$
$$= \mathbf{Q}_1 \mathbf{y}_{iv} + \mathbf{Q}_2 (\mathbf{x}_i' \hat{\mathbf{b}}^{*[k]} + \hat{\mathbf{a}}^{*[k]}) \qquad \text{[e.3]}$$

with $\mathbf{Q}_1 = \mathbf{Q}_v + \mathbf{Q}_m \mathbf{R}_{mv} \mathbf{R}_{vv}^{-1}$ and $\mathbf{Q}_2 = (\mathbf{Q}_m \mathbf{Q}^m - \mathbf{Q}_m \mathbf{R}_{mv} \mathbf{R}_{vv}^{-1} \mathbf{Q}^v)$.

Thus, for an animal with missing records, $\hat{\mathbf{y}}_i^*$ in [e.3] is the updated vector of observation transformed to canonical scale (steps 1a and 1b above) and this is calculated directly without back-transformation to the original scale (step 3). The matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ in [e.3] depend on the missing pattern and, if there are n missing patterns, n such matrices of each type must be set up initially and stored for use at each iteration.

E.2.1 Illustration

Using the same genetic parameters and data as for Example 5.3, the above methodology is employed to estimate sex effects and predict breeding values for pre-weaning weight and post-weaning gain iterating on the data (see Section 8.3].

From Section E.1, $\mathbf{Q}$ is:

$$\mathbf{Q} = \begin{bmatrix} 0.1659 & -0.0792 \\ 0.0168 & 0.1755 \end{bmatrix} \quad \text{and} \quad \mathbf{Q}^{-1} = \begin{bmatrix} 5.7651 & 2.6006 \\ -0.5503 & 5.4495 \end{bmatrix}$$

Partitioning $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ as specified above gives the following matrices:

$$\mathbf{Q}_v = \begin{bmatrix} 0.1659 \\ 0.0168 \end{bmatrix}, \quad \mathbf{Q}_m = \begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix}$$

$$\mathbf{Q}^v = [5.7651 \quad 2.6006] \quad \text{and} \quad \mathbf{Q}^m = [-0.5503 \quad 5.4495]$$

From the residual covariance matrix in Section E.1:

$$\mathbf{R}_{mv} \mathbf{R}_{vv}^{-1} = \tfrac{11}{40} = 0.275$$

The matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$, respectively, are:

$$\mathbf{Q}_1 = \begin{bmatrix} 0.1659 \\ 0.0168 \end{bmatrix} + \begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix} 0.275 = \begin{bmatrix} 0.1441 \\ 0.0654 \end{bmatrix}$$

and:

$$\mathbf{Q}_2 = \left[\begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix} [-0.5503 \quad 5.4495] - \begin{bmatrix} -0.0792 \\ 0.1755 \end{bmatrix} 0.275 [5.7651 \quad 2.6006] \right]$$
$$= \begin{bmatrix} 0.1691 & -0.3750 \\ -0.3748 & -0.8309 \end{bmatrix}$$

Employing steps 1 to 4 given earlier to the data in Example 5.3, using the various transformation matrices given above and solving for sex and animal solutions by iterating on the data (see Chapter 13, Section 13.3),

gave the following solutions on the canonical scale at convergence. The solutions on the original scale are also presented.

	Canonical scale		Original scale	
Effects	VAR1	VAR2	WWG	PWG
*Sex**				
1	0.180	1.265	4.326	6.794
2	0.124	1.108	3.598	5.968
Animal				
1	0.003	0.053	0.154	0.288
2	–0.006	–0.010	–0.059	–0.054
3	0.003	–0.030	–0.062	–0.163
4	0.002	0.007	0.027	0.037
5	–0.010	–0.097	–0.307	–0.521
6	0.001	0.088	0.235	0.477
7	–0.011	–0.084	–0.280	–0.452
8	0.013	0.076	0.272	0.407
9	0.009	0.010	0.077	0.051

*1 = male, 2 = female.
VAR1 = $\mathbf{Q}y_1$, VAR2 = $\mathbf{Q}y_2$ with WWG = y_1 and PWG = y_2.

These are similar to the solutions obtained from the multivariate analysis in Section 5.3 or the application of the Cholesky transformation in Section 5.4. The advantage of this methodology is that the usual univariate programs can easily be modified to incorporate missing records.

The prediction of the missing record (PWG) for animal 4 using solutions on canonical and original scales at convergence is illustrated below.

Using [e.1]:

$$\begin{aligned}\hat{y}_{42} &= \hat{b}_{12} + \hat{a}_{42} - R_{mv}R_{vv}^{-1}(y_{41} - \hat{b}_{11} - \hat{a}_{41}) \\ &= 6.794 + 0.037 + 0.275(4.5 - 4.326 - 0.027) \\ &= 6.9\end{aligned}$$

where y_{ij} and $\hat{a}_{ij}$ are the record and estimated breeding value, respectively, for animal i and trait j and b_{kj} is the fixed effect solution for level k for trait j.

Using [e.2]:

$$\begin{bmatrix}\hat{y}_{41}^{*} \\ \hat{y}_{42}^{*}\end{bmatrix} = \mathbf{Q}_1 y_{41} + \mathbf{Q}_2(\mathbf{x}'\hat{\mathbf{b}}^{*} + \hat{\mathbf{a}}_4^{*})$$

$$\begin{bmatrix}\hat{y}_{41}^{*} \\ \hat{y}_{42}^{*}\end{bmatrix} = \begin{bmatrix}0.648 \\ 0.294\end{bmatrix} + \mathbf{Q}_2\begin{bmatrix}0.180 \\ 1.265\end{bmatrix} + \mathbf{Q}_2\begin{bmatrix}0.002 \\ 0.007\end{bmatrix}$$

$$= \begin{bmatrix}0.648 \\ 0.294\end{bmatrix} + \begin{bmatrix}-0.445 \\ 0.990\end{bmatrix} = \begin{bmatrix}0.203 \\ 1.284\end{bmatrix}$$

These predicted records for animal 4 are on the canonical scale and they are used for the next round of iteration if convergence has not been achieved. These predicted records can be transformed to the original scale as:

$$\begin{bmatrix} \hat{y}_{41} \\ \hat{y}_{42} \end{bmatrix} = \mathbf{Q}^{-1} \begin{bmatrix} 0.203 \\ 1.284 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 6.9 \end{bmatrix}$$

The record for WWG is as observed and the predicted missing record for PWG is the same as when using [e.1].

E.3 Cholesky Decomposition

Any positive semi-definite symmetric matrix **R** can be expressed in the form **TT**′, where **T** is a lower triangular matrix. The matrix **T** can be calculated using the following formulae.

The *i*th diagonal element of **T** is calculated as:

$$t_{ii} = r_{ii} - \sqrt{\sum_{j=1}^{i-1} t_{ij}^2}$$

and the lower off-diagonal element of the *i*th row and the *k*th column of **T** as:

$$t_{ik} = \frac{1}{t_{kk}} \left(r_{ik} - \sum_{j=1}^{k-1} t_{ij} t_{kj} \right)$$

Appendix F: Procedure for Computing Deregressed Breeding Values

The deregressed breeding values (DRP) of bulls used in multi-trait across-country evaluations (MACE) are obtained by solving [5.22] for **y** considering data from only one country at a time. Jairath *et al*. (1998) presented an algorithm for calculating deregressed proof (DRP). For instance, equation [5.22] for country *i* can be written as:

$$\begin{pmatrix} \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{1} & \mathbf{1}'\mathbf{R}_i^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{R}_i^{-1}\mathbf{1} & \mathbf{R}_i^{-1}+\mathbf{A}_{nn}^{-1}\alpha_i & \mathbf{A}_{np}^{-1}\alpha_i & \mathbf{A}_{ng}^{-1}\alpha_i \\ \mathbf{0} & \mathbf{A}_{pn}^{-1}\alpha_i & \mathbf{A}_{pp}^{-1}\alpha_i & \mathbf{A}_{pg}^{-1}\alpha_i \\ \mathbf{0} & \mathbf{A}_{gn}^{-1}\alpha_i & \mathbf{A}_{gp}^{-1}\alpha_i & \mathbf{A}_{gg}^{-1}\alpha_i \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\mu}}_i \\ \mathbf{Q}\mathbf{g}_i+\hat{\mathbf{s}}_i \\ \hat{\mathbf{p}}_i \\ \hat{\mathbf{g}}_i \end{pmatrix} = \begin{pmatrix} \mathbf{1}'\mathbf{R}_i^{-1}\mathbf{y}_i \\ \mathbf{R}_i^{-1}\mathbf{y}_i \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \quad [f.1]$$

where $\mathbf{p}_i$ is the vector of identified parents without estimated breeding value (BV) and $\mathbf{A}_{ij}^{-1}$ are blocks of the inverse of the relationship (see Chapter 3, Section 3.5) with $j = n$, p and g for animals with records, ancestors and genetic groups, respectively, and $\alpha_i = (4 - h_i^2)/h_i^2$, the ratio of residual variance to sire variance for the *i*th country. The deregression of estimated BV involves solving [f.1] for y_i. The constant $\boldsymbol{\mu}_i$ and vectors $\mathbf{s}_i$, $\mathbf{p}_i$, $\mathbf{g}_i$ and $\mathbf{y}_i$ are unknown but $\mathbf{a}_i$, the vector of genetic evaluations for sires, is known as well as matrices $\mathbf{Q}$, $\mathbf{R}_i^{-1}$ and $\mathbf{A}_{ij}^{-1}$. Let $\mathbf{a}_i = \mathbf{1}\boldsymbol{\mu}_i + \mathbf{Q}\mathbf{g}_i + \mathbf{s}_i$. The following iterative procedure can be used to compute the vector of DRP, $\mathbf{y}_i$:

1. Set $\mathbf{1}\boldsymbol{\mu}_i$, $\mathbf{p}_i$, $\mathbf{s}_i$ and $\mathbf{g}_i$ to 0.
2. Calculate $\mathbf{Q}\mathbf{g}_i + \mathbf{s}_i = \mathbf{a}_i - \mathbf{1}\boldsymbol{\mu}_i$
3. Compute:

$$\begin{pmatrix} \hat{\mathbf{p}}_i \\ \hat{\mathbf{g}}_i \end{pmatrix} = -\begin{pmatrix} \mathbf{A}_{pp}^{-1} & \mathbf{A}_{pg}^{-1} \\ \mathbf{A}_{gp}^{-1} & \mathbf{A}_{gg}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A}_{pn}^{-1} \\ \mathbf{A}_{gn}^{-1} \end{pmatrix} (\mathbf{Q}\hat{\mathbf{g}}_i + \hat{\mathbf{s}}_i)$$

4. Generate:

$$\mathbf{R}_i^{-1}\mathbf{y}_i = \mathbf{R}_i^{-1}\mathbf{1}\boldsymbol{\mu}_i + (\mathbf{R}_i^{-1} + \mathbf{A}_{nn}^{-1})(\mathbf{Q}\hat{\mathbf{g}}_i + \mathbf{s}_i) + \mathbf{A}_{pn}^{-1}\hat{\mathbf{p}}_i\alpha_i + \mathbf{A}_{gn}^{-1}\hat{\mathbf{g}}_i\alpha_i$$

and $\mathbf{1}'\mathbf{R}_i^{-1}\mathbf{y}_i$.

5. Now calculate:

$$\boldsymbol{\mu}_1 = (\mathbf{1}'\mathbf{R}_i^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{R}_i^{-1}\mathbf{y}_i$$

6. Continue at step 2 until convergence is achieved.
7. Then compute DRP as $\mathbf{y}_i = \mathbf{R}_i(\mathbf{R}_i^{-1}\mathbf{y}_i)$.

Using the data for country 1 in Example 5.7, the deregression steps above are illustrated in the first iteration. For country 1, $\alpha_1 = 206.50/20.5 = 10.0732$ and, considering only the bulls with evaluations, $\mathbf{R}_1 = \text{diag}(0.0172, 0.0067, 0.0500, 0.0400)$. The pedigree structure (see Example 5.7) used for the deregression of breeding values in country 1 is:

Bull	Sire	MGS	MGD
1	5	G2	G3
2	6	7	G4
3	5	2	G4
4	1	G2	G4
5	G1	G2	G3
6	G1	G2	G3
7	G1	G2	G3

MGS, maternal grandsire; MGD, maternal grandam.

The matrix $\mathbf{A}_1^{-1}$ was calculated according to the rules in Section 5.6.2.

In the first round of iteration, the transpose of the vector $\mathbf{Qg}_1 + \mathbf{s}_1$ in step 2 above is:

$$(\mathbf{Qg}_1 + \mathbf{s}_1)' = (9.0 \quad 10.1 \quad 15.8 \quad -4.7)$$

The vector of solutions for $\mathbf{p}_1$ and $\mathbf{g}_1$ in step 3 is computed as:

$$\begin{bmatrix}\hat{\mathbf{p}}_1\\ \hat{\mathbf{g}}_1\end{bmatrix} = \begin{bmatrix} 17.094 & 0.000 & 0.000 & -5.037 & -0.839 & -0.839 & 1.831\\ 0.000 & 13.736 & 1.831 & -5.037 & -2.518 & -2.518 & 1.831\\ 0.000 & 1.831 & 10.989 & -5.037 & -2.518 & -2.518 & 0.916\\ -5.037 & -5.037 & -5.037 & 8.555 & 3.777 & 3.777 & 0.000\\ -0.839 & -2.518 & -2.518 & 3.777 & 4.568 & 2.728 & 0.839\\ -0.839 & -2.518 & -2.518 & 3.777 & 2.728 & 3.728 & 0.000\\ 1.831 & 1.831 & 0.916 & 0.000 & 0.839 & 0.000 & 3.671 \end{bmatrix}^{-1}$$

$$\begin{bmatrix} 6.716 & -1.831 & 7.326 & 0.000\\ 0.000 & 7.326 & 0.000 & 0.000\\ 0.000 & 3.663 & 0.000 & 0.000\\ 0.000 & 0.000 & 0.000 & 0.000\\ 1.679 & 0.000 & 0.000 & 3.357\\ 3.357 & 0.000 & 0.000 & 0.000\\ -1.679 & 2.747 & 3.663 & 3.357 \end{bmatrix} \begin{bmatrix} 9.0\\ 10.1\\ 15.8\\ -4.7 \end{bmatrix} = \begin{bmatrix} 16.330\\ 12.861\\ 12.622\\ 23.481\\ -9.801\\ 12.375\\ -0.564 \end{bmatrix}$$

The transpose of the vector $(\mathbf{R}_1^{-1}\mathbf{y}_1)$ in step 4 is (30.2 9.0 10.1 15.8) and

$$1'\mathbf{R}_1^{-1}\mathbf{y}_1 = 2235.50$$

Therefore in the first round of iteration (step 4):

$$\boldsymbol{\mu}_1 = 2235.50/253 = 8.835$$

Convergence was achieved after about six iterations. The transpose of the vector $(\mathbf{R}_1^{-1}\mathbf{y}_1)$ after convergence is:

$$(\mathbf{R}_1^{-1}\mathbf{y}_1)' = (563.928 \quad 1495.751 \quad 385.302 \quad -214.278)$$

with $\mathbf{R}_1^{-1}$ = diag(0.0172, 0.0067, 0.050, 0.04), the transpose of the vector of DRB calculated in step 7 is:

$$\mathbf{y}_1' = (9.7229 \quad 9.9717 \quad 19.2651 \quad -8.5711)$$

Appendix G: Calculating Φ, a Matrix of Legendre Polynomials Evaluated at Different Ages or Time Periods

The matrix **Φ** is of order t (the number days in milk or ages) by k (where k is the order of fit) with element $\phi_{ij} = \phi_j(a_t)$ equals the jth Legendre polynomial evaluated at the tth standardized age or days in milk (DIM). Thus a_t is the tth DIM or age standardized to the interval for which the polynomials are defined. Kirkpatrick *et al.* (1990, 1994) used Legendre polynomials which span the interval –1 to +1. Defining d_{min} and d_{max} as the first and latest DIM on the trajectory, DIM d_t can be standardized to a_t as:

$$a_t = -1 + 2(a_t - d_{\min})/(d_{\max} - d_{\min})$$

In matrix notation, $\mathbf{\Phi} = \mathbf{M\Lambda}$, where **M** is the matrix containing the polynomials of the standardized DIM values and **Λ** is a matrix of order k containing the coefficients of Legendre polynomials. The elements of **M** can be calculated as $m_{ij} = (a_i^{(j-1)},\ i = 1,\ldots,t;\ j = 1,\ldots,k)$. For instance, given that $k = 5$ and that $t = 3$ (three standardized DIM), **M** is:

$$\mathbf{M} = \begin{bmatrix} 1 & a_1 & a_1^2 & a_1^3 & a_1^4 \\ 1 & a_2 & a_2^2 & a_2^3 & a_2^4 \\ 1 & a_3 & a_3^2 & a_3^3 & a_3^4 \end{bmatrix}$$

Using the milk yield data in Table 7.1 as an illustration, with ten DIM, the vector of standardized DIM is:

$$\mathbf{a}' = [-1.0 \quad -0.7778 \quad -0.5556 \quad -0.3333 \quad -0.1111 \quad 0.1111 \quad 0.3333 \quad 0.5556 \quad 0.7778 \quad 1.0]$$

and **M** is:

$$\mathbf{M} = \begin{bmatrix} 1.0000 & -1.0000 & 1.0000 & -1.0000 & 1.0000 \\ 1.0000 & -0.7778 & 0.6049 & -0.4705 & 0.3660 \\ 1.0000 & -0.5556 & 0.3086 & -0.1715 & 0.0953 \\ 1.0000 & -0.3333 & 0.1111 & -0.0370 & 0.0123 \\ 1.0000 & -0.1111 & 0.0123 & -0.0014 & 0.0002 \\ 1.0000 & 0.1111 & 0.0123 & 0.0014 & 0.0002 \\ 1.0000 & 0.3333 & 0.1111 & 0.0370 & 0.0123 \\ 1.0000 & 0.5556 & 0.3086 & 0.1715 & 0.0953 \\ 1.0000 & 0.7778 & 0.6049 & 0.4705 & 0.3660 \\ 1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \end{bmatrix}$$

Next, the matrix $\mathbf{\Lambda}$ of Legendre polynomials needs to be computed. The *j*th Legendre polynomial evaluated at age t ($P_j(t)$), can in general be evaluated by the formula given by Abramowitz and Stegun (1965). In general, for the j integral, the polynomial:

$$P_j(t) = \frac{1}{2^j} \sum_{r=0}^{j/2} \frac{(-1)^r (2j-2r)!}{r!(j-r)!(j-2r)!} t^{j-2r}$$

where $j/2 = (j-1)/2$ if j is odd. The first five Legendre polynomials therefore are:

$$P_0(t) = 1; \quad P_1(t) = t; \quad P_2(t) = \tfrac{1}{2}(3t^2 - 1)$$

$$P_3(t) = \tfrac{1}{2}(5t^3 - 3t); \quad \text{and} \quad P_4(t) = \tfrac{1}{8}(35t^4 - 30t^2 + 3)$$

The normalized value (that is, the integration of $P_j^2(t)$ from –1 to 1 added to 1) of the *j*th Legendre polynomial evaluated at age t ($\phi_j(t)$), can be obtained as:

$$\phi_j(t) = \sqrt{\frac{2n+1}{2}} P_j(t)$$

Thus:

$$\phi_0(t) = \sqrt{\tfrac{1}{2}} P_0(t) = 0.7071; \quad \phi_1(t) = \sqrt{\tfrac{3}{2}} P_1(t) = 1.2247(t)$$

$$\phi_2(t) = \sqrt{\tfrac{5}{2}} P_2(t) = 2.3717(t^2) - 0.7906$$

$$\phi_3(t) = \sqrt{\tfrac{7}{2}} P_3(t) = 4.6771(t^3) - 2.8067(t)$$

$$\phi_4(t) = \sqrt{\tfrac{9}{2}} P_4(t) = 9.2808(t^4) - 7.9550(t^2) + 0.7955$$

Therefore, for $k = 5$ in Example 7.1, $\mathbf{\Lambda}$ is:

$$\mathbf{\Lambda} = \begin{bmatrix} 0.7071 & 0.0000 & -0.7906 & 0.0000 & 0.7955 \\ 0.0000 & 1.2247 & 0.0000 & -2.8067 & 0.0000 \\ 0.0000 & 0.0000 & 2.3717 & 0.0000 & -7.9550 \\ 0.0000 & 0.0000 & 0.0000 & 4.6771 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 9.2808 \end{bmatrix}$$

and $\mathbf{\Phi} = \mathbf{M\Lambda}$ is:

$$\mathbf{\Phi} = \begin{bmatrix} 0.7071 & -1.2247 & 1.5811 & -1.8704 & 2.1213 \\ 0.7071 & -0.9525 & 0.6441 & -0.0176 & -0.6205 \\ 0.7071 & -0.6804 & -0.0586 & 0.7573 & -0.7757 \\ 0.7071 & -0.4082 & -0.5271 & 0.7623 & 0.0262 \\ 0.7071 & -0.1361 & -0.7613 & 0.3054 & 0.6987 \\ 0.7071 & 0.1361 & -0.7613 & -0.3054 & 0.6987 \\ 0.7071 & 0.4082 & -0.5271 & -0.7623 & 0.0262 \\ 0.7071 & 0.6804 & -0.0586 & -0.7573 & -0.7757 \\ 0.7071 & 0.9525 & 0.6441 & 0.0176 & -0.6205 \\ 0.7071 & 1.2247 & 1.5811 & 1.8704 & 2.1213 \end{bmatrix} \quad [g.1]$$

Appendix H: Computing the Covariance Matrix of Additive Genetic Effect of Marked QTL (MQTL) when Paternal or Maternal Origin of Marker Alleles cannot be Determined and Marker Information is Incomplete

Wang *et al.* (1995) generalized [8.4] to accommodate situations where the paternal or maternal origin of marker alleles cannot be determined and where marker genotypes of some individuals are unknown. The covariance matrix, $\mathbf{G}_v$, between additive effects of MQTL alleles was derived by considering the covariance of alleles between individuals and within individuals separately.

H.1 Covariance Between Individuals

Assume that *s* and *d* are parents of *i*, and *j* is not a direct descendant of *i*; then [8.4] can be written as:

$$\text{cov}(v_i^{ki}, v_j^{kj} | \mathbf{M}_{obs}) = P(Q_i^{ki} \equiv Q_j^{kj} | \mathbf{M}_{obs})\sigma_v^2 \qquad [h.1]$$

where *ki* and *kj* can either be 1 (paternal MTQL allele) or 2 (maternal MQTL allele), $\mathbf{M}_{obs}$ is the marker genotype and$P(Q_i^{ki} \equiv Q_j^{kj} | \mathbf{M}_{obs})$ is the probability that Q_i^{ki} is identical by descent (IBD) to Q_j^{kj}, given $\mathbf{M}_{obs}$. This probability can be expressed as:

$$\begin{aligned} P(Q_i^{ki} \equiv Q_j^{kj} | \mathbf{M}_{obs}) = {} & P(Q_i^{ki} \leftarrow Q_s^1, Q_s^1 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \\ & + P(Q_i^{ki} \leftarrow Q_s^2, Q_s^2 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \\ & + P(Q_i^{ki} \leftarrow Q_d^1, Q_d^1 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \\ & + P(Q_i^{ki} \leftarrow Q_d^2, Q_d^2 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \end{aligned} \qquad [h.2]$$

where, for instance, $P(Q_i^{ki} \leftarrow Q_s^1,\ Q_s^1 \equiv Q_j^{kj})$ means that Q_i^{ki} descended from Q_s^1 and Q_s^1 was IBD to Q_j^{kj}. The conditional sampling of Q_i^{ki} from *s* or

d is independent of the alleles of j being IBD to alleles in s or d, given that j is not a direct descendant of i and marker genotypes of s and d are known.

Thus the probability in [h.2] can be computed recursively as:

$$\begin{aligned} P(Q_i^{ki} \equiv Q_j^{kj} | \mathbf{M}_{obs}) = {} & P(Q_i^{ki} \leftarrow Q_s^1 | \mathbf{M}_{obs}) P(Q_s^1 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \\ & + P(Q_i^{ki} \leftarrow Q_s^2 | \mathbf{M}_{obs}) P(Q_s^2 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \\ & + P(Q_i^{ki} \leftarrow Q_d^1 | \mathbf{M}_{obs}) P(Q_d^1 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \\ & + P(Q_i^{ki} \leftarrow Q_d^2 | \mathbf{M}_{obs}) P(Q_d^2 \equiv Q_j^{kj} | \mathbf{M}_{obs}) \end{aligned} \quad [h.3]$$

where, in general, $P(Q_i^{ki} \leftarrow Q_p^{kp})$ is the conditional probability that allele Q_i^{ki} in progeny i descended from allele Q_p^{kp} in parent $p = s$ or d for ki, $kp = 1$ or 2. This conditional probability is referred to as the probability of descent for a QTL allele (PDQ). There are eight PDQs for each individual, as Wang *et al.* (1995) showed, and these are given later as elements of the matrix **B** in equation [h.6]. Each PDQ can be computed as:

$$\begin{aligned} P(Q_i^{ki} \leftarrow Q_p^{kp} | \mathbf{M}_{obs}) = {} & (1-\rho) P(M_i^{ki} \leftarrow M_p^1 | \mathbf{M}_{obs}) \\ & + \rho P(M_i^{ki} \leftarrow M_p^2 | \mathbf{M}_{obs}) \end{aligned} \quad [h.4]$$

for $ki = 1$ or 2 and $p = s$ or d, where $\rho = r$ when $kp = 1$ and $\rho = 1 - r$ when $kp = 2$ and r is the recombination rate. In general, the probability $P(M_i^{ki} \leftarrow M_p^{kp} | \mathbf{M}_{obs})$ in the above equation is referred to as the probability of descent for a marker allele (PDM) and it defines the conditional probability that marker allele M_i^{ki} in progeny i descended from marker allele M_p^{kp}, given the marker genotype and pedigree. There are eight PDMs for an individual and these are given later as elements of the matrix **S** in equation [h.7]. The computation of the PDMs is illustrated in the next section. The PDQs and PDMs associated with unknown parents are undefined. For an individual i, equation [h.4] can be written in matrix notation as:

$$\mathbf{B}_i = \mathbf{S}_i \mathbf{R} \quad [h.5]$$

where:

$$\mathbf{B}_i = \begin{pmatrix} P(Q_i^1 \leftarrow Q_s^1 | \mathbf{M}_{obs}) & P(Q_i^1 \leftarrow Q_s^2 | \mathbf{M}_{obs}) & P(Q_i^1 \leftarrow Q_d^1 | \mathbf{M}_{obs}) & P(Q_i^1 \leftarrow Q_d^2 | \mathbf{M}_{obs}) \\ P(Q_i^2 \leftarrow Q_s^1 | \mathbf{M}_{obs}) & P(Q_i^2 \leftarrow Q_s^2 | \mathbf{M}_{obs}) & P(Q_i^2 \leftarrow Q_d^2 | \mathbf{M}_{obs}) & P(Q_i^1 \leftarrow Q_d^2 | \mathbf{M}_{obs}) \end{pmatrix}$$

[h.6]

$$\mathbf{S}_i = \begin{pmatrix} P(M_i^1 \leftarrow M_s^1 | \mathbf{M}_{obs}) & P(M_i^1 \leftarrow M_s^2 | \mathbf{M}_{obs}) & P(M_i^1 \leftarrow M_d^1 | \mathbf{M}_{obs}) & P(M_i^1 \leftarrow M_d^2 | \mathbf{M}_{obs}) \\ P(M_i^2 \leftarrow M_s^1 | \mathbf{M}_{obs}) & P(M_i^2 \leftarrow M_s^2 | \mathbf{M}_{obs}) & P(M_i^2 \leftarrow M_d^2 | \mathbf{M}_{obs}) & P(M_i^1 \leftarrow M_d^2 | \mathbf{M}_{obs}) \end{pmatrix}$$

[h.7]

and:

$$\mathbf{R} = \begin{pmatrix} 1-r & r & 0 & 0 \\ r & 1-r & 0 & 0 \\ 0 & 0 & 1-r & r \\ 0 & 0 & r & 1-r \end{pmatrix} \quad [h.8]$$

Thus, given the PDMs and **R**, the PDQs can be calculated.

H.1.1 Computing PDMs

Given that G_i, G_s and G_d are the marker genotypes of an individual i and its parents s and d, then $P(M_i^{ki} \leftarrow M_p^{kp} | \mathbf{M}_{obs}) = P(M_i^{ki} \leftarrow M_p^{kp} | G_s, G_d, G_i)$ can be computed as:

$$P(M_i^{ki} \leftarrow M_p^{kp} | G_s, G_d, G_i) = P(M_i^{ki} \leftarrow M_p^{kp}, G_i | G_s, G_d) / P(G_i | G_s, G_d) \quad [h.9]$$

The numerator and denominator of [h.9] can be calculated using Mendelian principles. For instance, if $G_s = M_s^1 M_s^2 = A_1A_1, G_d = M_d^1 M_d^2 = A_1A_2$ and $G_i = M_i^1 M_i^2 = A_1A_2$, then $P(G_i | G_s, G_d) = \frac{1}{2}$, as two of the four possible progeny genotypes are of type A_1A_2 and:

$$P(M_i^1 \leftarrow M_s^1, G_i | G_s, G_d) = \tfrac{1}{4}$$

This is so because the probability that M_i^1 descended from the sire marker genotype is 0.5 or from the dam's marker genotype is 0.5. Since these two events are independent:

$$P(M_i^1 \leftarrow M_s^1, G_i | G_s, G_d) = 0.5(0.5) = 0.25$$

Thus, using [h.9]:

$$P(M_i^1 \leftarrow M_s^1 | G_s, G_d, G_i) = 0.25 / 0.5 = 0.5$$

Following similar arguments:

$$P(M_i^1 \leftarrow M_d^1 | G_s, G_d, G_i) = 0.5/0.5 = 1 \text{ as } P(M_i^2 \leftarrow M_d^2, G_i | G_s, G_d) = 0.5$$

The eight PDMs for each individual i are stored in $\mathbf{S}_i$. The elements of the matrix $\mathbf{S}_i$ in equation [h.7] for the above example and other examples given by Wang *et al.* (1995) are:

s	d	i	$\mathbf{S}_i(1,1)$	$\mathbf{S}_i(1,2)$	$\mathbf{S}_i(1,3)$	$\mathbf{S}_i(1,4)$	$\mathbf{S}_i(2,1)$	$\mathbf{S}_i(2,2)$	$\mathbf{S}_i(2,3)$	$\mathbf{S}_i(2,4)$
A_1A_1	A_1A_2	A_1A_2	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0	0	0
A_1A_2	A_1A_2	A_1A_2	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
A_1A_1	A_1A_1	A_1A_1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

H.2 Covariance Within Individuals

For an individual i with parents s and d, the conditional covariance between the additive effects v_i^1 and v_i^2 of MQTL alleles Q_i^1 and Q_i^2, given $\mathbf{M}_{obs}$, can be written as:

$$\text{cov}(v_i^1, v_i^2 | \mathbf{M}_{obs}) = P(Q_i^1 \equiv Q_i^2 | \mathbf{M}_{obs}) \sigma_v^2 = f_i \sigma_v^2 \quad [h.10]$$

where $f_i = P(Q_i^1 \equiv Q_i^2 | \mathbf{M}_{obs})$ is the conditional inbreeding coefficient of individual i for the MQTL. It is the probability that, given $\mathbf{M}_{obs}$, the two homologous alleles at the MQTL in individual i are identical by descent.

The two homologous alleles at the MQTL, Q_i^1 and Q_i^2, in individual i, descended from one of the following parental pairs: (Q_s^1, Q_d^1), (Q_s^1, Q_d^2), (Q_s^2, Q_d^1) or (Q_s^2, Q_d^2). Assume that $T_{ks,kd}$ denotes the event that the pair of alleles in i descended from the parental pair (Q_s^{ks}, Q_d^{kd}) for k_s, $k_d = 1$ or 2, then f_i can be written (Wang *et al.*, 1995) as:

$$f_i = \sum_{ks=1}^{2} \sum_{kd=1}^{2} P(Q_s^{ks} \equiv Q_d^{kd} | \mathbf{M}_{obs}) P(T_{kskd} | \mathbf{M}_{obs}) \qquad [h.11]$$

As Wang *et al.* (1995) showed, the $P(T_{ks,kd} | \mathbf{M}_{obs})$ in the above equation can be expressed in terms of PDQs as:

$$\begin{aligned} P(T_{ks,kd} | \mathbf{M}_{obs}) = {} & P(Q_i^1 \leftarrow Q_s^{ks} | \mathbf{M}_{obs}) P(Q_i^2 \leftarrow Q_d^{kd} | \mathbf{M}_{obs}) / [P(Q_i^1 \leftarrow Q_s^1 | \mathbf{M}_{obs}) \\ & + P(Q_i^1 \leftarrow Q_s^2 | \mathbf{M}_{obs})] \\ & + P(Q_i^1 \leftarrow Q_d^{kd} | \mathbf{M}_{obs}) P(Q_i^2 \leftarrow Q_d^{ks} | \mathbf{M}_{obs}) \\ & / [P(Q_i^1 \leftarrow Q_d^1 | \mathbf{M}_{obs}) + P(Q_i^1 \leftarrow Q_d^2 | \mathbf{M}_{obs})] \end{aligned} \qquad [h.12]$$

For example, with $ks = 1$ and $kd = 2$

$$P(T_{12} | \mathbf{M}_{obs}) = B_i(1,1) B_i(2,4) / B_i(1,1) B_i(1,2) + B_i(1,4) B_i(2,1) / B_i(1,3) B_i(1,4)$$

where $B_i(l,k)$ are elements of B_i in [h.5]. If one of the denominators in [h.12] is zero, then the entire corresponding term is set to zero.

H.3 Constructing the Matrix G_v

Using equations [h.5] and [h.11], the matrix $\mathbf{G}_v$ with elements $g(i,j)$ can be constructed recursively as:

$$\begin{aligned} g\left(\tfrac{1}{i}, j\right) = {} & B_i(1,1) g\left(\tfrac{1}{s}, j\right) + B_i(1,2) g\left(\tfrac{2}{s}, j\right) \\ & + B_i(1,3) g\left(\tfrac{1}{d}, j\right) + B_i(1,4) g\left(\tfrac{2}{d}, j\right) \end{aligned} \qquad [h.13]$$

$$\begin{aligned} g\left(\tfrac{2}{i}, j\right) = {} & B_i(2,1) g\left(\tfrac{1}{s}, j\right) + B_i(2,2) g\left(\tfrac{2}{s}, j\right) \\ & + B_i(2,3) g\left(\tfrac{1}{d}, j\right) + B_i(2,4) g\left(\tfrac{2}{d}, j\right) \end{aligned} \qquad [h.14]$$

for $j = 1, \ldots, \frac{1}{i} - 1$, where $B_i(l,k)$ are defined as in [h.6], s and d are the sire and dam of animal i and $g(\frac{1}{i}, \frac{2}{i}) = f_i$ is defined in equation [h.11].

H.3.1 An illustration

Using the pedigree in Example 8.1, with the assumed genotypes in the table below, the calculation of $\mathbf{G}_v$, given that the paternal or maternal origin of marker alleles is undetermined, is illustrated. It is also assumed that marker $r = 0.1$ and $\sigma_v^2 = 1$.

Calf	Sire	Dam	Marker genotype
1	–	–	A_1A_1
2	–	–	A_2A_2
3	1	2	A_1A_2
4	1	3	A_1A_2

5	4	3	A_1A_2

First, calculate the $\mathbf{S}_i$ in [h.7] for animals with parents identified. The matrices $\mathbf{S}_i$, for $i = 3, 5$, are shown below:

$$\mathbf{S}_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \mathbf{S}_4 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{S}_5 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

Using equation [h.5], the corresponding $\mathbf{B}_i$, for $i = 3, 5$, are:

$$\mathbf{B}_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \mathbf{B}_4 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix} \quad \text{and}$$

$$\mathbf{B}_5 = \begin{pmatrix} 0.45 & 0.05 & 0.45 & 0.05 \\ 0.05 & 0.45 & 0.05 & 0.45 \end{pmatrix}$$

Using equations [h.13] and [h.14], the matrix $\mathbf{G}_V$ for the example pedigree is:

$$\mathbf{G}_V = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.500 & 0.050 & 0.453 & 0.073 \\ 0.000 & 1.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.500 & 0.050 & 0.453 & 0.073 \\ 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.450 & 0.048 & 0.428 \\ 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.500 & 0.000 & 0.450 & 0.048 & 0.428 \\ 0.500 & 0.500 & 0.000 & 0.000 & 1.000 & 0.000 & 0.500 & 0.100 & 0.680 & 0.120 \\ 0.000 & 0.000 & 0.500 & 0.500 & 0.000 & 1.000 & 0.000 & 0.900 & 0.095 & 0.855 \\ 0.500 & 0.500 & 0.000 & 0.000 & 0.500 & 0.000 & 1.000 & 0.050 & 0.678 & 0.098 \\ 0.050 & 0.050 & 0.450 & 0.450 & 0.100 & 0.900 & 0.050 & 1.000 & 0.163 & 0.863 \\ 0.453 & 0.453 & 0.048 & 0.048 & 0.680 & 0.095 & 0.678 & 0.163 & 1.000 & 0.167 \\ 0.073 & 0.073 & 0.428 & 0.428 & 0.120 & 0.855 & 0.098 & 0.863 & 0.167 & 1.000 \end{bmatrix}$$

The calculation for the first three animals can be illustrated as follows. For the first two animals, parents are unknown; therefore the first four rows of $\mathbf{G}_V$ have ones in their diagonal elements. For animal 3:

$$g\left(\begin{smallmatrix} 1 & 1 \\ 3 & 1 \end{smallmatrix}\right) = (1/2)g\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 1 & 1 \\ 2 & 1 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 2 & 1 \\ 2 & 1 \end{smallmatrix}\right) = 0.5$$

$$g\left(\begin{smallmatrix} 1 & 2 \\ 3 & 1 \end{smallmatrix}\right) = (1/2)g\left(\begin{smallmatrix} 1 & 2 \\ 1 & 1 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 2 & 2 \\ 1 & 1 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 2 & 2 \\ 2 & 1 \end{smallmatrix}\right) = 0.5$$

$$g\left(\begin{smallmatrix} 1 & 1 \\ 3 & 2 \end{smallmatrix}\right) = (1/2)g\left(\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 1 & 1 \\ 2 & 2 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 2 & 1 \\ 2 & 2 \end{smallmatrix}\right) = 0$$

$$g\left(\begin{smallmatrix} 1 & 2 \\ 3 & 2 \end{smallmatrix}\right) = (1/2)g\left(\begin{smallmatrix} 1 & 2 \\ 1 & 2 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 2 & 2 \\ 1 & 2 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 1 & 2 \\ 2 & 2 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 2 & 2 \\ 2 & 2 \end{smallmatrix}\right) = 0$$

$$g\left(\begin{smallmatrix} 2 & 1 \\ 3 & 1 \end{smallmatrix}\right) = (0)g\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 1 & 1 \\ 2 & 1 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 2 & 1 \\ 2 & 1 \end{smallmatrix}\right) = 0$$

$$g\left(\begin{smallmatrix} 2 & 2 \\ 3 & 1 \end{smallmatrix}\right) = (0)g\left(\begin{smallmatrix} 1 & 2 \\ 1 & 1 \end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix} 2 & 2 \\ 1 & 1 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 1, & 2 \\ 2 & 1 \end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix} 2 & 2 \\ 2 & 1 \end{smallmatrix}\right) = 0$$

$$g\left(\begin{smallmatrix}2 & 1\\ 3 & 2\end{smallmatrix}\right) = (0)g\left(\begin{smallmatrix}1 & 1\\ 1 & 2\end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix}2 & 1\\ 1 & 2\end{smallmatrix}\right) + (1/2)\,g\left(\begin{smallmatrix}1 & 1\\ 2 & 2\end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix}2 & 1\\ 2 & 2\end{smallmatrix}\right) = 0.5$$

$$g\left(\begin{smallmatrix}2 & 2\\ 3 & 2\end{smallmatrix}\right) = (0)g\left(\begin{smallmatrix}1 & 2\\ 1 & 2\end{smallmatrix}\right) + (0)g\left(\begin{smallmatrix}2 & 2\\ 1 & 2\end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix}1 & 2\\ 2 & 2\end{smallmatrix}\right) + (1/2)g\left(\begin{smallmatrix}2 & 2\\ 2 & 2\end{smallmatrix}\right) = 0.5$$

Since the parents of animal 3 are unrelated, $g\left(\begin{smallmatrix}1 & 2\\ 3 & 3\end{smallmatrix}\right) = 1$.

The calculation of f_i for animal 4 is illustrated as the parents are related. From [h.11], to compute f_4, $P(Q_1^{k1} \equiv Q_3^{k3} | \mathbf{M}_{obs})$ and $P(T_{k1,k3} | \mathbf{M}_{obs})$ for k_1, $k_3 = 1$ or 2 are needed. The probabilities $P(Q_1^{k1} \equiv Q_3^{k3} | \mathbf{M}_{obs})$ have already been calculated earlier in computing $\mathbf{G}_v$ as:

$$P(Q_1^1 \equiv Q_3^1 | \mathbf{M}_{obs}) = g\left(\begin{smallmatrix}1 & 1\\ 1 & 3\end{smallmatrix}\right) = 0.5$$

$$P(Q_1^1 \equiv Q_3^2 | \mathbf{M}_{obs}) = g\left(\begin{smallmatrix}1 & 2\\ 1 & 3\end{smallmatrix}\right) = 0$$

$$P(Q_1^2 \equiv Q_3^1 | \mathbf{M}_{obs}) = g\left(\begin{smallmatrix}2 & 1\\ 1 & 3\end{smallmatrix}\right) = 0.5$$

$$P(Q_1^2 \equiv Q_3^2 | \mathbf{M}_{obs}) = g\left(\begin{smallmatrix}2 & 2\\ 1 & 3\end{smallmatrix}\right) = 0$$

Probabilities $P(T_{k1,k3} | \mathbf{M}_{obs})$ can be computed, using [h.11], as:

$$\begin{aligned} P(T_{11} | \mathbf{M}_{obs}) &= B_4(1,1)B_4(2,3)/B_4(1,1) + B_4(1,2) \\ &\quad + B_4(1,3)B_4(2,1)/B_4(1,3) + B_4(1,4) \\ &= 0.5\ (0.1)/(0.5 + 0.5) + 0 = 0.05 \end{aligned}$$

Similarly:

$$P(T_{12} | \mathbf{M}_{obs}) = 0.45, P(T_{21} | \mathbf{M}_{obs}) = 0.05 \quad \text{and} \quad P(T_{22} | \mathbf{M}_{obs}) = 0.45$$

Therefore:

$$f_4 = \sum_{k1=1}^{2} \sum_{k3=1}^{2} P(Q_1^{k1} \equiv Q_3^{k3} | \mathbf{M}_{obs}) P(T_{k1k3} | \mathbf{M}_{obs}) = 0.05$$

The extension of the above algorithm for computing $\mathbf{G}_v$ when QTL is bracketed by two markers is given by Pong-Wong *et al.* (2001).

H.4 Computing Inverse of $\mathbf{G}_v$

Using the same principles discussed in Chapter 8, Section 8.4, the inverse of the $\mathbf{G}_v$ matrix can be calculated as:

$$\mathbf{G}_{v,i}^{-1} = \begin{pmatrix} \mathbf{G}_{v,i-1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \mathbf{W}_i \qquad [h.15]$$

where $\mathbf{W}_i$ is a matrix containing the contributions of individual i to $\mathbf{G}_{v,i}^{-1}$ and is:

$$\mathbf{W}_i = \begin{pmatrix} \mathbf{B}'\mathbf{D}_i^{-1}\mathbf{B}_i & -\mathbf{B}_i'\mathbf{D}_i^{-1} \\ -\mathbf{D}_i^{-1}\mathbf{B}_i' & \mathbf{D}_i^{-1} \end{pmatrix} \qquad [h.16]$$

where $\mathbf{D}_i = \mathbf{C}_i - \mathbf{B}_i\mathbf{C}_{sd}\mathbf{B}_i'$, with $\mathbf{C}_{sd}$ being the 4 by 4 conditional gametic relationship matrix for parents of i, s and d, the elements of which are in $\mathbf{G}_{v,i-1}$, $\mathbf{B}_i$ is as defined in [h.6] and $\mathbf{C}_i$ is:

$$\mathbf{C}_i = \begin{pmatrix} 1 & f_i \\ f_i & 1 \end{pmatrix}$$

Thus, when $f_i = f_s = f_d = 0$, then:

$$\mathbf{D}_i = \mathbf{I}_2 - \mathbf{B}_i\mathbf{B}'_i \qquad \text{[h.17]}$$

The algorithm for computing $\mathbf{G}_V^{-1}$ therefore is as follows:

Set $\mathbf{G}_V^{-1}$ to a null matrix. For individual i, for $i = 1,\ldots,n$:

1. If both parents are unknown, add 1 to positions $\begin{pmatrix} 1 & 1 \\ i & i \end{pmatrix}$ and $\begin{pmatrix} 2 & 2 \\ i & i \end{pmatrix}$ of $\mathbf{G}_V^{-1}$.
2. If at least one parent is known, then:
   (a) compute $\mathbf{B}_i$ using [h.5];
   (b) compute $\mathbf{D}_i$ as defined in [h.16] if there is inbreeding, or [h.17] if there is no inbreeding;
   (c) compute $\mathbf{W}_i$ using [h.16] and add the elements to the relevant positions in $\mathbf{G}_V^{-1}$.

For the example pedigree, $\mathbf{G}_V^{-1}$ is:

$$\mathbf{G}_V^{-1} = \begin{bmatrix}
2.000 & 1.000 & 0.000 & 0.000 & -1.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 \\
1.000 & 2.000 & 0.000 & 0.000 & -1.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.500 & 0.500 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.500 & 1.500 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
-1.000 & -1.000 & 0.000 & 0.000 & 2.617 & 0.707 & 0.561 & -0.349 & -1.210 & -0.325 \\
0.000 & 0.000 & -1.000 & -1.000 & 0.707 & 7.456 & 0.207 & -4.044 & -0.226 & -2.098 \\
-1.000 & -1.000 & 0.000 & 0.000 & 0.561 & 0.207 & 2.561 & 0.207 & -1.210 & -0.325 \\
0.000 & 0.000 & 0.000 & 0.000 & -0.349 & -4.044 & 0.207 & 6.511 & -0.226 & -2.098 \\
0.000 & 0.000 & 0.000 & 0.000 & -1.210 & -0.226 & -1.210 & -0.226 & 2.667 & 0.207 \\
0.000 & 0.000 & 0.000 & 0.000 & -0.325 & -2.098 & -0.325 & -2.098 & 0.207 & 4.640
\end{bmatrix}$$

H.5 Incomplete Marker Data

When marker genotypes of parents are unknown, Wang *et al.* (1995) showed that equation [h.3] does not hold; therefore, the algorithm in Section H.3 for calculating $\mathbf{G}_V$ cannot be applied. Given incomplete marker information, the conditional gametic relationship matrix, $\mathbf{G}_{V|\mathbf{M}_{obs}}$ can be computed as:

$$\mathbf{G}_{V|\mathbf{M}_{obs}} = \sum_{\omega\in\Omega} \mathbf{G}_{V|\omega,\mathbf{M}_{obs}} P(\omega|\mathbf{M}_{obs}) \qquad \text{[h.18]}$$

where Ω is the set of all possible marker genotype configurations for individuals with unknown genotypes, $\mathbf{G}_{V|\omega,\mathbf{M}_{obs}}$ is the conditional gametic relationship matrix, given marker genotypes ω for individuals with unknown genotypes and $\mathbf{M}_{obs}$ for individuals with known genotypes, $P(\omega|\mathbf{M}_{obs})$ being the conditional probability of individuals with unknown genotypes having marker genotypes ω, given $\mathbf{M}_{obs}$. Although $P(\omega|\mathbf{M}_{obs})$ in [h.18] can be computed efficiently (Elston and Stewart, 1971), the use of [h.18] is not efficient when a large number of individuals have unknown genotypes. Wang *et al.* (1995) presented an approximation in which the PDMs (in

equation [h.6]) for individual i having parents s and d with unknown marker genotypes can be computed as:

$$P(M_i^1 \leftarrow M_s^1 | \mathbf{M}_{obs}) = \sum_{G_s} \sum_{G_d} \sum_{G_i} P(M_i^1 \leftarrow M_s^1 | G_s, G_d, G_i) P(G_s, G_d, G_i | \mathbf{M}_{obs})$$

where each summation is over all possible genotypes at the marker locus. If G_s, G_d or G_i is not missing, then the corresponding summation is dropped. However, with a large number of individuals with a missing marker genotype, computing $P(G_s, G_d, G_i | \mathbf{M}_{obs})$ can take a long time, but it can be approximated by conditioning only on marker information of 'close' relatives of i, s and d, such as the parents, sibs or offspring.

In practice, simulation techniques, usually Markov chain Monte Carlo (MCMC) methods (Chapter 12), are employed to calculate $P(\omega | \mathbf{M}_{obs})$. One such method is the 'single site' approach (Sheehan, 1990), which involves updating the individual's genotype, conditioned upon the individual's phenotype and current genotypes of the parents, mates and progeny. However, this suffers from poor 'mixing' qualities for a complex pedigree and irreducibility of the chains can only be ensured for biallelic loci. More complex sampling techniques have been presented that address the problem of the irreducibility of the chain (Lin *et al.*, 1994; Lund and Jensen, 1998), but this cannot always be guaranteed.

Thompson (1994) and Thompson and Heath (1999) presented an alternative sampling strategy based on utilizing segregation indicators. It is based on the principle that segregation events (separation of alleles at a locus during meiosis) determine the inheritance of genetic materials from parent to progeny. The segregation indicators give possible allelic pathways through the pedigree. Using their notation, the segregation indicator (S_{ij}) equals 0 if inherited at the ith segregation and the jth locus is the parent's maternal allele. Otherwise, $S_{ij} = 1$ if the inherited allele at segregation i at locus j is the parent's paternal allele. The set of segregation indicators for the n segregations in the pedigree and the m loci, where these loci may be marker loci and/or QTL, is represented by $\mathbf{s} = \{S_{ij}, i = 1, \ldots, n, j = 1, \ldots, m\}$. The segregation indicators can be used to estimate IBD probabilities between any pair of individuals in the pedigree from a large number of s with probability $P(\mathbf{s} | \mathbf{M}_{obs})$. Using the above principles, a multiple-site segregation sampler was developed by Thompson and Heath (1999). This is implemented in the QTL mapping software, LOKI. The user supplies LOKI with the pedigree structure, marker genotypes, marker positions and MQTL positions for which the IBD matrices are to be calculated.

References

Abramowitz, M. and Stegun, I.A. (1965) *Handbook of Mathematical Functions*. Dover, New York.

Albuquerque, L.G. and Meyer, K. (2001) Estimates of covariance functions for growth from birth to 630 days in Nelore cattle. *Journal of Animal Science* 79, 2776–2789.

Andersen, S. and Pedersen, B. (1996) Growth and food intake curves for group-housed gilts and castrated male pigs. *Animal Science* 63, 457–464.

Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

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

Boichard, D., Bonaiti, B., Barbat, A. and Mattalia, S. (1995) Three methods to validate the estimation of genetic trend for dairy cattle. *Journal of Dairy Science* 78, 431–437.

Boichard, D., Fritz, S., Rossignol, M.N., Boscher, M.Y., Malafosse, A. and Colleau, J.J. (2002) Implementation of marker-assisted selection in French dairy cattle. *7th World Congress on Genetics Applied to Livestock Production, August 2002. Montpellier*, France, Communication no. 22-03.

Brotherstone, S. and Hill, W.G. (1991) Dairy herd life in relation to linear type traits and production. 2 Genetic analyses for pedigree and non-pedigree cows. *Animal Production* 53, 289–297.

Brotherstone, S., White, I.M.S. and Meyer, K. (2000) Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. *Animal Science* 70, 407–415.

Bulmer, M.G. (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford, UK.

Cameron, N.D. (1997) *Selection Indices and Prediction of Genetic Merit in Animal Breeding*. CAB International, Wallingford, UK.

Cantet, R.J.C. and Smith, C. (1991) Reduced animal model for marker assisted selection using best linear unbiased prediction. *Genetics, Selection, Evolution* 23, 221–233.

Cockerham, C.C. (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39, 859–882.

Dempster, A.P., Laird, N.M. and Rubin, D.B (1977) Maximum likelihood from incomplete data via EM algorithm. *Journal of Royal Statistics Society, Series B* 39, 1–38.

Ducrocq, V. (1992) Solving animal model equations through an approximate incomplete Cholesky decomposition. *Genetics, Selection, Evolution* 24, 193–209.

Ducrocq, V. and Besbes, B. (1993) Solutions of multiple trait animal models with missing data on some traits. *Journal of Animal Breeding and Genetics* 110, 81–89.

Elston, R.C. and Stewart, J. (1971) A general model for the genetic analysis of pedigree data. *Human Heredity* 21, 523–542.

Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn. Longman, London.

Fernando, R.L. and Grossman, M. (1989) Marker-assisted selection using best linear unbiased prediction. *Genetics, Selection, Evolution* 21, 467–477.

Fernando, R.L., Billingsley, R.D. and Gianola, D. (1983) Effects of method of scaling on heritability estimates and sire evaluations for frame size at weaning in Angus cattle. *Journal of Animal Science* 56, 1047–1056.

Fisher, R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society*, *Edinburgh* 52, 399–433.

Fiske, W.F. and Banos, G. (2001) Weighting factors of sire daughter information in international genetic evaluations. *Journal of Dairy Science* 84, 1759–1767.

Foulley, J.L., Gianola, D. and Thompson, R. (1983) Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genetics, Selection, Evolution* 15, 401–424.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.

Gianola, D. (1982) Theory and analysis of threshold characters. *Journal of Animal Science* 54, 1079–1096.

Gianola, D. and Foulley, J.L. (1983) Sire evaluation for ordered categorical data with a threshold model. *Genetics, Selection, Evolution* 15, 201–224.

Gilmour, A.R., Thompson, R. and Cullis, B. (1995) Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440–1450.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. and Thompson, R. (2003) *ASREML UserGuide Release 1.0*. VSN International, Hemel Hempstead, UK, 310 pp.

Goddard, M.E. (1985) A method of comparing sires evaluated in different countries. *Livestock Production Science* 13, 321–331.

Goddard, M.E. (1992) A mixed model for the analyses of data on multiple genetic markers. *Theoretical and Applied Genetics* 83, 878–886.

Goddard, M.E. (1999) New technology to enhance genetic improvement of pigs. *Manipulating Pig Production* 7, 44–52.

Goddard, M.E. and Hayes, B.J. (2002) Optimisation of response using molecular data. *7th World Congress on Genetics Applied to Livestock Production, August 2002. Montpellier, France*, Communication no. 22-01.

Groeneveld, E. (1990) *PEST User's Manual*. Institute of Animal Husbandry and Animal Behaviour, Federal Agricultural Research Centre, Neustadt, Germany.

Groeneveld, E. and Kovac, M. (1990) A generalised computing procedure for setting up and solving mixed models. *Journal of Dairy Science* 73, 513–531.

Groeneveld, E., Kovac, M. and Wang, T. (1990) Pest, a general purpose BLUP package for multivariate prediction and estimation. In: *Proceedings of the 4th World Congress on Genetics Applied to Livestock, Edinburgh*, 13, pp. 488–491.

Grundy, B., Villanueva, B. and Woolliams, J.A. (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetical Research Cambridge* 72, 159–168.

Guo, Z. and Swalve, H.H. (1997) Comparison of different lactation curve sub-models in test day models. *Proceedings of the Annual Interbull Meeting, Vienna. Interbull Bulletin* 16, 75–79.

Haldane, J.B.S. (1919) The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* 8, 299.

Henderson, C.R. (1949) Estimation of changes in herd environment. *Journal of Dairy Science* 32, 709 (Abstract).

Henderson, C.R. (1950) Estimation of genetic parameters. *Annals of Mathematical Statistics* 21, 309.

Henderson, C.R. (1953) Estimation of variance and covariance components. *Biometrics* 9, 226–252.

Henderson, C.R. (1963) Selection index and expected geentic advance. In: Hanson, W.D. and Robinson, H.F. (eds) *Statistical Genetics and Plant Breeding*. Publication 982, National Academy of Sciences, National Research Council, Washington, DC, pp. 141–163.

Henderson, C.R. (1973) Sire evaluation and genetic trends. In: *Proceedings of the Animal Breeding and Genetics Symposium in Honour of J.L. Lush*. American Society for Animal Science, Blackburgh, Champaign, Illinois, pp. 10–41.

Henderson, C.R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447.

Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in predicting of breeding values. *Biometrics* 32, 69–83.

Henderson, C.R. (1984) *Applications of Linear Models in Animal Breeding.* University of Guelph Press, Guelph, Canada.

Henderson, C.R. (1985) Best linear unbiased prediction of non-additive genetic merits in non-inbred populations. *Journal of Animal Science* 60, 111–117.

Henderson, C.R. and Quaas, R.L. (1976) Multiple trait evaluation using relatives' records. *Journal of Animal Science* 43, 1188–1197.

Henderson, C.R., Kempthorne, O., Searle, S.R. and von Krosigk, C.M. (1959) The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192–218.

Heringstad, B., Chang, Y.M., Gianola, D. and Klemetsdal, G. (2002) Analysis of clinical mastitis in Norwegian cattle with a longitudinal threshold model. In: *7th World Congress Applied to Livestock Production, August 2002, Montpellier, France*, Communication no. 20–03.

Hoeschele, I. and VanRaden, P.M. (1991) Rapid inverse of dominance relationship matrices for noninbred populations by including sire by dam subclass effects. *Journal of Dairy Science* 74, 557–569.

IMSL (1980) *Library Reference Manual*. IMSL, Houston, Texas.

Interbull (2000) *National Genetic Evaluation Programmes for Dairy Production Traits Practised in Interbull Member Countries 1999–2000*. Department of Animal Breeding and Genetics, Uppsala, Sweden, Bulletin 24.

Jairath, L., Dekkers, J.C.M., Schaeffer, L.R., Liu, Z., Burnside, E.B. and Kolstad, B. (1998) Genetic evaluation for herd life in Canada. *Journal of Dairy Science* 81, 550–562.

Jamrozik, J. and Schaeffer, L.R. (1997) Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. *Journal of Dairy Science* 80, 762–770.

Jamrozik, J., Schaeffer, L.R. and Dekkers, J.C.M. (1997) Genetic evaluation of dairy cattle using test day yields and random regression model. *Journal of Dairy Science* 80, 1217–1226.

Jensen, J. and Madsen, P. (1997) *A User's Guide to DMU. A Package for Analysing Multivariate Mixed Models*. Research Centre, National Institute of Animal Science, Foulum, Denmark.

Jensen, J., Wang, C.S., Sorensen, D.A. and Gianola, D. (1994) Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agriculturæ Scandinavica* 44, 193–201.

Kemp, R.A. (1985) The effects of positive assortative mating and preferential treatment of progeny on the estimation of breeding values. Unpublished PhD Thesis, University of Guelph, Guelph, Canada.

Kempthorne, O. and Nordskog, A.W. (1959) Restricted selection indexes. *Biometrics* 15, 10–19.

Kennedy, B.W. (1989) *Animal Model BLUP*. Erasmus Intensive Graduate Course, Trinity College, Dublin.

Kirkpatrick, M., Lofsvold, D. and Bulmer, M. (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979–993.

Kirkpatrick, M., Hill, W.G. and Thompson, R. (1994) Estimating the covariance of traits during growth and ageing, illustrated with lactation in dairy cattle. *Genetics Research Cambridge* 64, 57–69.

Lidauer, M., Strandén, I., Mäntysaari, E.A., Pösö, J. and Kettunen, A. (1999) Solving large test-day models by iteration on the data and preconditioned conjugate gradient. *Journal of Dairy Science* 82, 2788–2796.

Lin, S., Thompson, E. and Wijsman, E. (1994) Finding noncommunicating sets for Markov chain Monte Carlo estimations on pedigrees. *American Journal of Human Genetics* 54, 695–704.

Liu, Z., Reinhardt, F. and Reents, R. (2002) The multiple trait effective daughter contribution method applied to approximate reliabilities of estimated breeding values from a random regression test day model for genetic evaluation in dairy cattle. In: *7th World Congress Applied to Livestock Production*, *August 2002, Montpellier, France*, Communication no. 20–15.

Lohuis, M.M. and Dekkers, J.C.M. (1998) Merits of borderless evaluations. In: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production Science*, *Armidale, Australia*, XXVI, pp. 169–172.

Lund, M. and Jensen, C.S. (1998) Multivariate updating of genotypes in a Gibbs sampling algorithm in the mixed inheritance model. In: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production Science*, *Armidale, Australia*, pp. 25, 521–524.

Mantysaari, E.A. (1999) Derivation of multiple trait reduced random regression model for the first lactation test day records of milk, protein and fat. In: Proceedings of the 50th Annual Meeting of EAAP, *August 1999, Zurich, Switzerland*, 26 pp.

Maron, M.J. (1987) *Numerical Analysis*. MacMillian, New York.

Meijering, A. and Gianola, D. (1985) Linear versus nonlinear methods of sire evalutions for categorical traits: a simulation study. *Génétique, Sélection, Évolution* 17, 115–132.

Meuwissen, T.H.E. (1997) Maximising the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* 75, 934–940.

Meuwissen, T.H.E. and Luo, Z. (1992) Computing inbreeding coefficients in large populations. *Génétique, Sélection, Évolution* 24, 305–313.

Meyer, K. (1989) Approximate accuracy of genetic evaluation under an animal model. *Livestock Production Science* 21, 87–100.

Meyer, K. (1999) Estimates of genetic and phenotypic covariance functions for post-weaning growth and mature weight of beef cows. *Journal of Animal Breeding and Genetics* 116, 181–205.

Meyer, K. and Hill, W.G. (1997) Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. *Livestock Production Science* 47, 185–200.

Meyer, K. and Tier, B. (2003) Approximate prediction covariances among multiple estimated breeding values for individuals. *Proceedings of the Annual Interbull Meeting, Rome. Interbull Bulletin* 31, 133–136.

Misztal, I. and Gianola, D. (1988) Indirect solution of mixed model equations. *Journal of Dairy Science*, 71 (Suppl. 2), 99–106.

Mrode, R.A. and Swanson, G.J.T. (1995) Comparison of the efficiency of the repeatability model on observed or transformed yields to a multivariate analysis. In: *Summaries of Papers, British Society of Animal Science Winter Meeting, March 1995*, paper no. 129.

Mrode, R.A. and Swanson, G.J.T. (1999) Simplified equations for evaluations of bulls in the Interbull international evaluation system. *Livestock Production Science* 62, 43–52.

Mrode, R.A. and Swanson, G.J.T. (2004) Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. *Livestock Production Science* 86, 253–260.

Pander, B.L., Hill, W.G. and Thompson, R. (1992) Genetic parameters of test day records of British Holstein heifers. *Animal Production* 55, 11–21.

Patterson, H.D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.

Pong-Wong, R., George, A.W., Woolliams, J.A. and Haley, C.S. (2001) A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genetics, Selection, Evolution* 33, 453–471.

Ptak, E. and Schaeffer, L.R. (1993) Use of test day yields for genetic evaluation of dairy sires and cows. *Livestock Production Science* 34, 23–34.

Quaas, R.L. (1976) Computing the diagonal elements of a large numerator relationship matrix. *Biometrics* 32, 949–953.

Quaas, R.L. (1984) *Linear Prediction in BLUP School Handbook: Use of Mixed Models for Prediction and Estimation of (Co)variance Components.* Animal Genetics and Breeding Unit, University of New England, New South Wales, Australia, 76 pp.

Quaas, R.L. (1995) Fx algorithms. An unpublished note.

Quaas, R.L. and Pollak, E.J. (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *Journal of Animal Science* 51, 1277–1287.

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

Robertson, A. and Rendel, J.M. (1954) The performance of heifers got by artificial insemination. *Journal of Agricultural Science, Cambridge* 44, 184–192.

Ronningen, K. and Van Vleck, L.D. (1985) Selection index theory with practical applications. In: Chapman, A.B. (ed.) *General and Quantitative Genetics*. World Animal Science, A4, Elsevier Science, Oxford, UK.

Schaeffer, L.R. (1984) Sire and cow evaluation under multiple trait models. *Journal of Dairy Science* 67, 1567–1580.

Schaeffer, L.R. (1994) Multiple-country comparison of dairy sires. *Journal of Dairy Science* 77, 2671–2678.

Schaeffer, L.R. (2004) Application of random regression models in animal breeding. *Livestock Production Science* 86, 35–45.

Schaeffer, L.R. and Dekkers, J.C.M. (1994) Random regression in animal models for test-day production in dairy cattle. In: *Proceedings of the 5th World Congress on Applied Livestock Production*, *Guelph, Canada, XVIII*, pp. 443–446.

Schaeffer, L.R. and Kennedy, B.W. (1986) Computing solutions to mixed model equations. In: *Third World Congress on Genetics Applied to Livestock Production, Lincoln, Nebraska*, p. 392.

Schaeffer, L.R. and Wilton, J.W. (1987) RAM computing strategies and multiple traits. Prediction of genetic value for beef cattle. In: *Proceedings Workshop II*. Winrock, International, Kansas City, Missouri.

Schaeffer, L.R., Wilton, J.W. and Thompson, R. (1978) Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* 34, 199–208.

Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J. and van Doormaal, B.J. (2000) Experience with a test day model. *Journal of Dairy Science* 83, 1135–1144.

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

Sheehan, N. (1990) Genetic restoration on complex pedigrees. PhD thesis, University of Washington, Seattle, Washington.

Sigurdsson, A. and Banos, G. (1995) Dependent variables in international sire evaluations. *Acta Agriculturae Scandinavica* 45, 209–217.

Sorensen, D. and Gianola, D. (2002) *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer-Verlag, New York.

Sorensen, D.A. and Kennedy, B.W. (1983) The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. *Theoretical and Applied Genetics* 66, 217–220.

Sorensen, D.A., Andersen, S. and Gianola, D. (1995) Bayesian inference in threshold models using Gibbs sampling. *Genetics, Selection, Evolution* 27, 229–249.

Strandén, I. and Lidauer, M. (1999) Solving large mixed linear models using preconditioned conjugate gradient iteration. *Journal of Dairy Science* 82, 2779–2787.

Szyda, J., Liu, Z., Reinhardt, F. and Reents, R. (2003) Incorporation of QTL information into routine estimation of breeding values for German Holstein dairy cattle. *Proceedings of the 54th Annual metting of the European Association for Animal Production*, Rome, 2003, p. 11.

Thompson, E.A. (1994) Monte Carlo estimation of multilocus autozygosity probabilities. In: *Proceedings of the 1994 Interface Conference, Fairfax Station, Virginia*, pp. 498–506.

Thompson, E.A. and Heath, S.C. (1999) Estimation of conditional multilocus gene identity among relatives. In: Sheillier-Moseiwitch, F., Donelly, P. and

Watterman, M. (eds) *Statistics in Molecular Biology*. IMS Lecture Note Series, Springer-Verlag, New York, pp. 95–113.

Thompson, R. (1977a) The estimation of heritability with unbalanced data. II. Data available on more than two generations. *Biometrics* 33, 497–504.

Thompson, R. (1977b) Estimation of quantitative genetic parameters. In: Pollak, E., Kempthorne, O. and Bailey, T.B. (eds) *Proceedings of the International Conference on Quantitative Genetics*. Iowa State University Press, Ames, Iowa, pp. 639–657.

Thompson, R. (1979) Sire evaluation. *Biometrics* 35, 339–353.

Thompson, R. and Meyer, K. (1986) A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Production Science* 15, 299–313.

Tier, B. and Solkner, J. (1993) Analysing gametic variation with an animal model. *Theoretical and Applied Genetic*s 85, 868–872.

Van Arendonk, J.A.M., Tier, B. and Kinghorn, B.P. (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics* 137, 319–329.

VanRaden, P.M. and Hoeschele, I. (1991) Rapid inverse of additive by additive relationship matrices by including sire–dam combination effects. *Journal of Dairy Science* 74, 570–579.

VanRaden, P.M. and Wiggans, G.R. (1991) Derivation, calculation and use of national animal model information. *Journal of Dairy Science* 74, 2737–2746.

Veerkamp, R.F., Hill, W.G., Stott, A.W., Brotherstone, S. and Simm, G. (1995) Selection for longevity and yield in dairy cows using transmitting abilities for type and yield. *Animal Science* 61, 189–197.

Visscher, P.M. (1991) Estimation of genetic parameters in dairy cattle using an animal model and implications for genetic evaluation. Unpublished PhD thesis, University of Edinburgh, UK.

Wang, C.S., Rutledge, J.J. and Gianola, D. (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genetics, Selection, Evolution* 25, 41–62.

Wang, C.S., Rutledge, J.J. and Gianola, D. (1994) Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genetics, Selection, Evolution* 26, 91–115.

Wang, T., Fernando, R.L., van der Beek, S., Grossman, M. and van Arendonk, J.A.M. (1995) Covariance between relatives for a marked quantitative trait locus. *Genetics, Selection, Evolution* 27, 251–274.

Wang, Y., Miller, S.P., Schenkel, F.S., Wilton, J.W. and Boettcher, P.J. (2002) Performance of a linear-threshold model to evaluate calving interval and birth weight in a multibreed beef population. In: *7th World Congress Applied to Livestock Production, August 2002, Montpellier, France*, Communication no. 17–10.

Weigel, K.A. and Rekaya, R. (2000) A multiple-trait model herd cluster model for international dairy sire evaluation. *Journal of Dairy Science* 83, 815–821.

Weigel, K.A., Rekaya, R., Zwald, N.R. and Fikse, W.F. (2001) International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records. *Journal of Dairy Science* 84, 2789–2795.

Weller, J.I. (2001) *Quantitative Trait Loci Analysis in Animals*. CAB International, Wallingford, UK.

Westell, R.A. and Van Vleck, L.D. (1987) Simultaneous genetic evaluation of sires and cows for a large population of dairy cattle. *Journal of Dairy Science* 70, 1006–1017.

Westell, R.A., Quaas, R.L. and Van Vleck, L.D. (1988) Genetic groups in an animal model. *Journal of Dairy Science* 71, 1310–1318.

White, I.M.S., Thompson, R. and Brotherstone, S. (1999) Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science* 82, 632–638.

Wiggans, G.R., VanRaden, P.M. and Powell, R.L. (1992) A method for combining United States and Canadian bull proofs. *Journal of Dairy Science* 75, 2834–2839.

Willham, R.L. (1963) The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* 19, 18–27.

Wright, S. (1922) Coefficients of inbreeding and relationship. *American Naturalist* 56, 330–338.

Index