On estimating (co)variance components in animal models with competition effects[1]

R.J.C.Cantet [a, b]; E. P. Cappa[a]

[a.] Departamento de Producción Animal, Universidad de Buenos Aires,

Avenida San Martín 4453, C1417DSE Buenos Aires, Argentina

[b.] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

1. Corresponding author: Rodolfo J .C. Cantet, rcantet@agro.uba.ar

Fax: 54-011-4514-8000, extension 8785, Phone: 54-011-4514-8000, ext. 8784

ABSTRACT

There is an increased interest in estimating the (co)variance components of additive animal models with direct and competition effects (AMC). However, all attempts to estimate the dispersion parameters in different animal species faced problems of convergence and highly variable estimates. The trouble is related with the lack of asymptotic identifiability in certain AMC. This property is observed when calculating the 4 × 4 information matrix ($\boldsymbol{I}(\boldsymbol{\theta})$) for the AMC REML likelihood, and its smallest eigenvalue goes to zero. The singularity of $\boldsymbol{I}(\boldsymbol{\theta})$ is due to confounding between the fixed pen (contemporary group) effect and the additive competition effects. The incidence matrix of additive competition effects ($\boldsymbol{Z}_c$) can be written as a function of the “intensity of competition” (IC) among animals in a contemporary group. The IC values can be interpreted as weighting factors expressing how intense pairs of animals compete in relation to all other animals. The sum of squares of the IC values in any row of $\boldsymbol{Z}_c$ should add up to 1 (standardization) in order for the phenotypic variance of any given observation not to be affected by the number of competitors. Moreover, data sets to estimate the (co)variance components in the AMC should be obtained with some sort of design in order for the (co)variance components to be asymptotically identifiable. Examples are presented in which the IC’s are related to either time or number of competitors in the pen.

Key Words: competition effects, identifiability, covariance components.

Introduction

Recently, there has been a growing interest in genetic evaluation models including additive genetic effects of competition (AMC). Although quantitative genetics theory dealing with interacting effects in plants and animals dates back to Griffing (1967), it is only recently (Muir, 2005; Van Vleck and Cassady, 2005; Arango et al, 2005; Cappa and Cantet, 2007) that competition effects have been introduced in genetic evaluation using a mixed model framework. The appeal of AMC models relates to the increase in either total performance, or animal wellbeing, if genetic competition effects are selected against (Muir, 2005; Van Vleck and Cassady, 2005).

A major drawback with the AMC in animals is that all reported estimates of the dispersion parameters seem to suffer from either inaccuracy or problems of convergence of the (co)variance components (Van Vleck and Cassady, 2005; Arango et al, 2005; Van Vleck et al, 2007). These problems have not been observed when estimating the (co)variance components in an individual tree model (Cappa and Cantet, 2007). There is a major difference between the later estimates to those parameter estimates with animals in the elements of the incidence matrix for the random additive competition effects ($\boldsymbol{Z}_c$). The objective of this presentation is to show, by way of simple examples, the role of matrix $\boldsymbol{Z}_c$ in the identifiability of the (co)variance components in AMC. A second goal is to display the effects of the distribution of the additive relationships of animals across pens on the asymptotic variance of the estimated dispersion parameters.

Theory

The equation for the animal model with competition effects is as follows:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_d\boldsymbol{a}_d + \boldsymbol{Z}_c\boldsymbol{a}_c + \boldsymbol{e} \qquad [1]$$

where $\boldsymbol{y}$ is the $n \times 1$ vector of records; $\boldsymbol{X}$ is a full rank $n \times p$ incidence matrix that relates the observations to the vector of fixed effects $\boldsymbol{\beta}$, and $\boldsymbol{e}$ ($n \times 1$) is the random vector of i.i.d. errors distributed as $N_n(\boldsymbol{0}, I_n\, \sigma^2_e)$, being σ^2_e the error variance. The vector of direct breeding values is $\boldsymbol{a}_d$ and the vector of breeding values for competition effects is $\boldsymbol{a}_c$. We assume that the same q individuals in $\boldsymbol{a}_d$ have competition breeding values in $\boldsymbol{a}_c$, and that individual breeding values are ordered by date of birth the same way in both vectors. Direct and competition breeding values are related to $\boldsymbol{y}$ by the $n \times q$ incidence matrices $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_c$, respectively. Each row of $\boldsymbol{Z}_d$ has all elements equal to 0 except for a 1 in the column belonging to $\boldsymbol{a}_{di}$. Matrix $\boldsymbol{Z}_c$ is described in detail below. The variances for direct and competition breeding values are $\sigma^2_{\mathrm{A}d}$ and $\sigma^2_{\mathrm{A}c}$, respectively, whereas the covariance between both types of additive effects is $\sigma_{\mathrm{A}d\mathrm{A}c}$. Also, let $\boldsymbol{A}$ be the $q \times q$ relation matrix with diagonal elements equal to $1 + F_i$, with F_i being the inbreeding coefficient of i, and off-diagonals elements equal to the additive relationships $\boldsymbol{A}_{ij}$. Now, we are able to write the total additive covariance matrix in a more compact manner as follows

$$\mathrm{Var}\begin{bmatrix}\boldsymbol{a}_d \\ \boldsymbol{a}_c\end{bmatrix} = \begin{bmatrix}\sigma^2_{\mathrm{A}d} & \sigma_{\mathrm{A}d\mathrm{A}c} \\ \sigma_{\mathrm{A}d\mathrm{A}c} & \sigma^2_{\mathrm{A}c}\end{bmatrix} \otimes \boldsymbol{A} = \boldsymbol{G}_0 \otimes \boldsymbol{A}$$

Taking into account the random effects in model [1], the (co)variance matrix ($\boldsymbol{V}$) of $\boldsymbol{y}$ is given by:

$$\boldsymbol{V} = \boldsymbol{Z}_d \boldsymbol{A} \boldsymbol{Z}_d^{\prime} \sigma^2_{\mathrm{Ad}} + \left(\boldsymbol{Z}_d \boldsymbol{A} \boldsymbol{Z}_c^{\prime} + \boldsymbol{Z}_c \boldsymbol{A} \boldsymbol{Z}_d^{\prime} \right) \sigma_{\mathrm{AdAc}} + \boldsymbol{Z}_c \boldsymbol{A} \boldsymbol{Z}_c^{\prime} \sigma^2_{\mathrm{Ac}} + \boldsymbol{I}_n \sigma^2_e \qquad [1B]$$

Key elements in model [1] are the non-zero elements of matrix $\boldsymbol{Z}_c$. We write $\boldsymbol{a}_{cj}$ for the breeding value of competition effects of animal *j*. Cappa and Cantet (2007) observed that the total additive genetic competition that is exerted over individual *i* from competitors $j_1, j_2, \ldots, j_{m_i}$ can be written as

$$f_{i1} \boldsymbol{a}_{c_1} + f_{i2} \boldsymbol{a}_{c_2} + \ldots + f_{im} \boldsymbol{a}_{c_m} = \sum_{j=1}^{m_i} f_{ij} \boldsymbol{a}_{c_j} \qquad [2]$$

The f_{ij} elements in [2] are interpreted as the *intensities of competition* (IC) that the breeding values of $j_1, j_2, \ldots, j_m$ exert over the phenotype of *i* ($y_{ij_1..j_m}$). To obtain the IC values, Cappa and Cantet (2007) considered that in the absence of inbreeding and of genetic relationships among competitors and regardless of the number of competing animals, the variance of the breeding values of the competitors in the observation of *i* is equal to

$$\mathrm{Var}\left(\sum_{j=1}^{m_i} f_{ij} \boldsymbol{a}_{c_j} \right) = \sigma^2_{\mathrm{Ac}} \qquad [3]$$

This implies that the potential genetic strength ($\boldsymbol{a}_c$) for competing with m_i animals is split m_i times. Cappa and Cantet (2007) observed further how to obtain the IC for competing trees planted in a regular grid. Whichever the source of competition is, note that

$$\mathrm{Var}\left(\sum_{j=1}^{m_i} f_{ij} \boldsymbol{a}_{c_j} \right) = \left(\sum_{j=1}^{m_i} f^2_{ijk} \right) \sigma^2_{\mathrm{A}c} \qquad [4]$$

As the variance of any breeding value is (1 + *F*) σ^2_{A} (Kempthorne, 1969, page 349), when inbreeding is null (*F* = 0) expression [4] implies that

$$\sum_{j=1}^{m_i} f^2_{ijk} = 1 \qquad [5]$$

Thus, to standardize the variance of the breeding values for competition effects in the phenotypic variance of any animal *i*, the f_{ij}'s of its competitors should be chosen so that the sums of squares of the IC values add up to 1.

We now show how the f_{ij}'s can be calculated. Suppose the intensity of competition is related to the *time* the animals spend together in the pen. Thus, some individuals may spend the entire post weaning period in the same pen, while others may be rotated among pens during *m* periods. As in crossover designs (Bate and Jones, 2006), animals are assigned to *sequences* of occupancy times of the pens. For example, for *m* = 3, the sequence 1-2-1 means that the animals spent the first period in pen 1, the second in pen 2, and finally the individual goes back to pen 1 during the third period. Two individuals in the sequence 1-1-1 compete all the time. However, competition of an animal in 1-1-1 with another individual assigned to sequence 1-2-1 is two third as intense as the competition with an animal in the same sequence. The amount of time spent in any period does not have to be equal to the previous periods, but each change of pen should be counted as a new period. The idea is to calculate the IC as a function of the total proportion of shared time. By letting n_k to be the number of competitors of animal *i* during sequence *k*, we have that

$$\mathrm{Var}\left(\sum_{j=1}^{m_i} f_{ij} \boldsymbol{a}_{c_j} \right) = \left(n_1 \; f^2_{i1} + \ldots + n_k \; f^2_{ik} + \ldots + n_{m_i} \; f^2_{i m_i} \right) \sigma^2_{\mathrm{Ac}} \qquad [6]$$

By equating [4] to [6], the ICs are such that

$$1=\sum_{j=1}^{m_i} f_{ijk}^2 = n_1\ f_{i1}^2 + \ldots + n_k\ f_{ik}^2 + \ldots + n_{m_i}\ f_{i\,m_i}^2 \qquad [7]$$

To solve [7] we have to use the relationships among ICs based on the fraction of time animals were competing. For the example above with two sequences, f_{12} = (2/3) f_{11}, as an animal in the 1-1-1 pattern compete one third of the time more with another individual in the same sequence, as compared with an animal in the 1-2-1 pattern. Suppose there are 2 animals in sequence 1-1-1 and 4 in sequence 1-2-1, then expression [7] for an animal in sequence 1-1-1 is equal to $f^2_{11} + 4 f^2_{12} = 1$. On replacing in the last expression with f_{12} = (2/3) f_{11}, we have $f^2_{11} + 4\,[(2/3)\, f_{11}]^2 = 1$, with solutions equal to $f_{11} = 3/5$ and $f_{12} = 2/5$. For an individual in sequence 1-2-1, [7] is now equal to $2 f^2_{21} + 3 f^2_{22} = 1$ with the restriction that f_{21} = (2/3) f_{22}, and the solution is $f_{22} = {}^3\!/\!\sqrt{35}$ and $f_{21} = {}^2\!/\!\sqrt{35}$. By letting the records of animals in the 1-1-1 sequence to come first than those from animals of the sequence 1-2-1, matrix $\boldsymbol{Z}_c$ is then equal to

$$\boldsymbol{Z}_c = \begin{bmatrix} 0 & 3/5 & 2/5 & 2/5 & 2/5 & 2/5 \\ 3/5 & 0 & 2/5 & 2/5 & 2/5 & 2/5 \\ 2/\sqrt{35} & 2/\sqrt{35} & 0 & 3/\sqrt{35} & 3/\sqrt{35} & 3/\sqrt{35} \\ 2/\sqrt{35} & 2/\sqrt{35} & 3/\sqrt{35} & 0 & 3/\sqrt{35} & 3/\sqrt{35} \\ 2/\sqrt{35} & 2/\sqrt{35} & 3/\sqrt{35} & 3/\sqrt{35} & 0 & 3/\sqrt{35} \\ 2/\sqrt{35} & 2/\sqrt{35} & 3/\sqrt{35} & 3/\sqrt{35} & 3/\sqrt{35} & 0 \end{bmatrix}$$

This type of structure in $\boldsymbol{Z}_c$ allows a feasible estimation of the variance components, as we discuss below.

Identifiability of the variance components

Let $\boldsymbol{\theta} = [\theta_i]$ be the parametric vector of the (co)variance components (i = 1, .., k) of a mixed model, and $\boldsymbol{V}_1,\ldots,\boldsymbol{V}_i,\ldots,\boldsymbol{V}_k$ the associated (co)variance structures, so that Var($\boldsymbol{y}$) = $\boldsymbol{V}_1\,\theta_1 + \boldsymbol{V}_2\,\theta_2 + \ldots + \boldsymbol{V}_k\,\theta_k$. Jiang (1996) defined that a mixed model is identifiable of its (co)variance components under the translation invariance class (IDI), if: 1) matrices $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots \boldsymbol{V}_k$ are linearly independent; 2) there exists a matrix $\boldsymbol{P}$ such that $\boldsymbol{PX} = \boldsymbol{0}$ and rank[$\boldsymbol{P}$] = n – rank[$\boldsymbol{X}$]. The matrix with such properties is $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\,\boldsymbol{V}^{-1}$ (Harville, 1977). Jiang (1996, lemma 4.1) further observed that a mixed model belong to the IDI class, if and only if, the smallest eigenvalue of the REML information matrix ($\boldsymbol{I}(\boldsymbol{\theta})$) is positive. As the inverse of the information matrix is the asymptotic covariance matrix of REML estimates (Searle et al, 1992), the previous statement is tantamount to $\boldsymbol{I}(\boldsymbol{\theta})$ being positive definite, or equivalently, admitting an inverse. This property becomes useful to check whether the (co)variance components in model [1] are identifiable. The expression for the i,j element of $\boldsymbol{I}(\boldsymbol{\theta})$ (Harville, 1977, page 326) is

$$\boldsymbol{I}_{ij}(\boldsymbol{\theta}) = \operatorname{tr}\left(\boldsymbol{P}\ \frac{\partial \boldsymbol{V}}{\partial\,\theta_i}\ \boldsymbol{P}\ \frac{\partial\ \boldsymbol{V}}{\partial\,\theta_j}\right) = \operatorname{tr}\left(\boldsymbol{P}\ \boldsymbol{V}_i\ \boldsymbol{P}\ \boldsymbol{V}_j\right) \qquad [8]$$

For model [1] $\boldsymbol{I}(\boldsymbol{\theta})$ is 4 × 4, $\boldsymbol{V}_1 = \boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_d'$, $\boldsymbol{V}_2 = \boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_c' + \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_d'$, $\boldsymbol{V}_3 = \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_c'$, and $\boldsymbol{V}_4 = \boldsymbol{I}_n$. Diagonal elements are

$$\boldsymbol{I}_{11}(\boldsymbol{\theta}) = \operatorname{tr}\left(\boldsymbol{P}\ \boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_d'\boldsymbol{P}\ \boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_d'\right),$$

$$\boldsymbol{I}_{22}(\theta) = \operatorname{tr}\left[\boldsymbol{P}\left(\boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_c' + \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_d'\right)\boldsymbol{P}\ \left(\boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_c' + \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_d'\right)\right] \qquad [9]$$

$$\boldsymbol{I}_{33}(\theta) = \operatorname{tr}\left(\boldsymbol{P}\ \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_c'\boldsymbol{P}\ \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_c'\right) \qquad \boldsymbol{I}_{44}(\theta) = \operatorname{tr}\left(\boldsymbol{P}\,\boldsymbol{P}\right).$$

Whereas off-diagonal elements are

$$\boldsymbol{I}_{12}(\theta) = \operatorname{tr}\left[\boldsymbol{P}\boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_d'\boldsymbol{P}\ \left(\boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_c' + \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_d'\right)\right],$$

$$\boldsymbol{I}_{13}(\theta) = \operatorname{tr}\left[\boldsymbol{P}\boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_d'\boldsymbol{P}\ \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_c'\right]\ \boldsymbol{I}_{14}(\theta) = \operatorname{tr}\left(\boldsymbol{P}\ \boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_d'\boldsymbol{P}\right) \qquad [10]$$

$$\boldsymbol{I}_{23}(\theta) = \operatorname{tr}\left[\ \boldsymbol{P}\ \left(\boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_c' + \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_d'\right)\boldsymbol{P}\boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_c'\right]$$

$$\boldsymbol{I}_{24}(\theta) = \operatorname{tr}\left[\ \boldsymbol{P}\ \left(\boldsymbol{Z}_d\boldsymbol{A}\boldsymbol{Z}_c' + \boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_d'\right)\boldsymbol{P}\right]\ \ \boldsymbol{I}_{34}(\theta) = \operatorname{tr}\left[\boldsymbol{P}\boldsymbol{Z}_c\boldsymbol{A}\boldsymbol{Z}_c'\boldsymbol{P}\right]$$

Harville (1977) showed how to write down the elements of $\boldsymbol{I}(\boldsymbol{\theta})$ in terms of matrices related to the mixed model equations of Henderson (1984). Boca and Cantet (2004) used this approach to calculate asymptotic variances of REML estimates from an animal model with additive and dominance effects. To deal with small examples, we will calculate $\boldsymbol{I}(\boldsymbol{\theta})$ using expressions [9] and [10] to, in order to find out whether a given experimental design allows estimating separately the four (co)variance components.

To get insight into how $\boldsymbol{I}(\boldsymbol{\theta})$ helps to identify the (co)variance components, consider first the simple additive animal model with direct effects only and homogeneous error variance. The $\boldsymbol{I}(\boldsymbol{\theta})$ is now of order 2×2 and equal to

$$\boldsymbol{I}\begin{bmatrix}\sigma^2_{\mathrm{A}}\\ \sigma^2_e\end{bmatrix} = \begin{bmatrix}\operatorname{tr}\left[\boldsymbol{P}\boldsymbol{Z}\ \boldsymbol{A}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z}\ \boldsymbol{A}\boldsymbol{Z}'\right] & \operatorname{tr}\left[\boldsymbol{P}\boldsymbol{Z}\ \boldsymbol{A}\boldsymbol{Z}'\boldsymbol{P}\right]\\ \operatorname{tr}\left[\boldsymbol{P}\boldsymbol{Z}\ \boldsymbol{A}\boldsymbol{Z}'\boldsymbol{P}\right] & \operatorname{tr}\left[\boldsymbol{P}\,\boldsymbol{P}\right]\end{bmatrix} \qquad [11]$$

When there are q unrelated individuals each one having one record, $\boldsymbol{Z} = \boldsymbol{I}_q$ and $\boldsymbol{A} = \boldsymbol{I}_q$. All assumptions considered, the covariance matrix of the data is equal to:

$$\boldsymbol{V} = \boldsymbol{Z}\ \boldsymbol{A}\boldsymbol{Z}'\,\sigma^2_{\mathrm{A}} + \boldsymbol{I}\ \sigma^2_e = \boldsymbol{I}\ \left(\sigma^2_{\mathrm{A}} + \sigma^2_e\right)$$

and the information matrix in [11] is

$$\boldsymbol{I}\begin{bmatrix}\sigma^2_{\mathrm{A}}\\ \sigma^2_e\end{bmatrix} = \begin{bmatrix}\operatorname{tr}\left[\boldsymbol{P}\,\boldsymbol{P}\right] & \operatorname{tr}\left[\boldsymbol{P}\,\boldsymbol{P}\right]\\ \operatorname{tr}\left[\boldsymbol{P}\,\boldsymbol{P}\right] & \operatorname{tr}\left[\boldsymbol{P}\,\boldsymbol{P}\right]\end{bmatrix}$$

As all the elements of $\boldsymbol{I}(\boldsymbol{\theta})$ are equal, the matrix is singular. This, in turn, indicates that there is not enough information in the data structure to estimate σ^2_{A} and σ^2_e separately.

Designs for identifying the (co)variance components in models with competition

In this section we will give simple examples showing the difficulties involved in estimating all four (co)variance components from model [1]. Consider first the example given by Muir (2005). There are 8 individuals with data divided equally into 2 pens. Moreover, all animal in each pen are full-sibs and each animal in any pen is a half-sib of any animal in the other pen. For simplicity, let $\sigma^2_{\mathrm{A}d} = \sigma^2_{\mathrm{A}c} = \sigma^2_e = 1$ and $\sigma_{\mathrm{A}d\mathrm{A}c} = -0.25$. Using the specification for [1] as in Muir (2005), matrices $\boldsymbol{X}$, $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_c$ are respectively equal to

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{Z}_d = \boldsymbol{I}_8 \qquad \boldsymbol{Z}_c = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \qquad [12]$$

whereas the additive relationship matrix is

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 1 & 0.5 & 0.5 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.5 & 1 & 0.5 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.5 & 0.5 & 1 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 1 & 0.5 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.5 & 1 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.5 & 0.5 & 1 & 0.5 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

The eigenvalues of $\boldsymbol{I}(\boldsymbol{\theta})$ are 2.97, 1.52, 0, and 0; therefore, the matrix is singular and not all dispersion parameters can be separately estimated. Alternatively, suppose that pens are divided in halves with a central feeder shared by all animals in the pen. Now, it is reasonable to assume that the intensity of competition for animals in the same half-pen is greater than the one among individuals in different halves. For any animal in a half-pen expression [7] is equal to $f^2_{12} + 2\,f^2_{12} = 1$. Assuming than competition between the two pigs in the same half-pen is twice as intense as with the two pigs in the other half-pen, we can pose the restriction that $f_{11} = 2\,f_{12}$. After solving we obtain $f_{11} = \sqrt{2/3}$ and $f_{12} = \sqrt{1/6}$. For the other pen with animals 5 to 8, individuals 5 and 8 remains in the same half-pen during the entire feeding period (sequence 1-1), whereas 6 and 7 switch half-pens at the middle of the trial (sequence 1-2). The ICs for animal 5 are such that $f^2_{56} + f^2_{57} + f^2_{58} = 1$ with the restriction that $f_{56} = f_{57} = 2f_{58}$, as 5 (1-1) compete with 8 (2-2) less than with 6 or 7 that are in the sequences 1-2 or 2-1. Solutions are $f_{56} = 2/3$ and $f_{58} = 1/3$, so that matrix $\boldsymbol{Z}_c$ is then equal to

$$\boldsymbol{Z}_c = \begin{bmatrix} 0 & \sqrt{2/3} & \sqrt{1/6} & \sqrt{1/6} & 0 & 0 & 0 & 0 \\ \sqrt{2/3} & 0 & \sqrt{1/6} & \sqrt{2/3} & 0 & 0 & 0 & 0 \\ \sqrt{1/6} & \sqrt{1/6} & 0 & \sqrt{2/3} & 0 & 0 & 0 & 0 \\ \sqrt{1/6} & \sqrt{1/6} & \sqrt{2/3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 2/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 2/3 \\ 0 & 0 & 0 & 0 & 1/3 & 1/3 & 2/3 & 0 \end{bmatrix} \qquad [13]$$

Now $\boldsymbol{I}(\boldsymbol{\theta})$ is non-singular with eigenvalues 2.81, 2.02, 0.17, and 3.86×10^{-5}. Notice that the last eigenvalue is very small, which in turn indicates large asymptotic variances of the REML estimates when analyzing data with this design.

Can the asymptotic variances be improved upon? One possibility is to use another ‘genetic design’, thus modifying the $\boldsymbol{A}$ matrix. Suppose that 4 families with 2 full-sibs each are available so that $\boldsymbol{A} = \boldsymbol{I}_2 \otimes \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$, and matrices $\boldsymbol{X}$, $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_c$ are the same as before. Then, the eigenvalues of the resulting $\boldsymbol{I}(\boldsymbol{\theta})$ are 2.64, 1.72, 0.28, and 0.12. A useful criterion to compare the efficiency of different experimental designs is *D*-optimality (Wald, 1943), and amounts to maximizing the determinant of $\boldsymbol{I}(\boldsymbol{\theta})$ ($|\boldsymbol{I}(\boldsymbol{\theta})|$). The value of $|\boldsymbol{I}(\boldsymbol{\theta})|$ for the design with the genetic structure used by Muir (2005) is equal to 3.86×10^{-5}, whereas for the design with 4 families of 2 full-sibs each and animals distributed across sub-pens $|\boldsymbol{I}(\boldsymbol{\theta})| = 0.159$. This is a difference in efficiency of more than 4111 times ($0.1591/3.86\times10^{-5}$). Though the example may seem artificial, it serves to illustrate two facts: 1) if pen is a fixed effect in the model, animals in the same pen should have different ICs in order for the covariance components to be identifiable; 2) the distribution of families of full and half sibs across pens plays a role in the efficiency of the design. Further examples are two alternatives to the design with $\boldsymbol{Z}_c$ as in [13], keeping $\boldsymbol{X}$, $\boldsymbol{Z}_d$ and $\boldsymbol{A}$ as before. Notice that $\boldsymbol{Z}_c$ has diagonal blocks with different structure. In the first design, animals are not rotated across half-pens but stay in the same half-pen during the entire feeding period (1-1/2-2). Then, matrix $\boldsymbol{Z}_c$ is block-diagonal with the same structure as in the upper block of [13]. In the other design animals in both pens are rotated across half-pens at the middle of the period (1-2/2-1), so that $\boldsymbol{Z}_c$ is block-diagonal with the same structure as in the lower block of [13]. Then, $|\boldsymbol{I}(\boldsymbol{\theta})| = 0.178$ for 1-1/2-2 design, and $|\boldsymbol{I}(\boldsymbol{\theta})| = 0.119$ for the 1-2/2-1 design. Thus, the 1-1/2-2 design was 12% more efficient than 1-1/1-2, and almost 50% more efficient than the 1-2/2-1 design.

Discussion

While data structures used here are small and somewhat complex, they serve to illustrate the problem of estimating (co)variance components in models with competition effects. Any data set used for estimating the dispersion parameters will suffer from lack of identifiability when all animals in the same pen share the same IC factor and pen is treated as a fixed effect. Additive relationships may ameliorate the problem but one has to go to the trouble of calculating $\boldsymbol{I}(\boldsymbol{\theta})$ to be sure all (co)variance components are identifiable. In turn, distribution of families across pens or half-pens will affect the asymptotic variances of the dispersion parameters.

References

Arango, J., I. Misztal, S. Tsuruta, M. Culbertson, and W. Herring 2005. Estimation of variance components including competitive effects of Large White growing gilts. J. Anim. Sci. 83: 1241-1246.

Boca, R. T., and R. J. C. Cantet. 2004. On the precision of REML estimates of additive and dominance variance components in maize using relationship matrices. Maydica. 49: 295-302

Bate, S. T., and B. Jones. 2006. The construction of nearly balanced and nearly strongly balanced uniform cross-over designs. J. Statistical Planning and Inference. 136: 3248-3267.

Harville, D. 1997. Matrix algebra from a statistician’s perspective. Springer-Verlag, New York, NY.

Henderson, C. R. 1984. Applications of linear models to animal breeding. University of Guelph, Guelph, ON.

Kempthorne, O. 1969. An Introduction to Genetic Statistics. The Iowa state University Press, Ames, IA.

Muir, W. M. 2005. Incorporation of competitive effects in forest tree or animal breeding programs. Genetics 170: 1247-1259.

Searle, S, Casella, G., and C. McCullough. 1992. Variance Components. Wiley, New York, NY.

Van Vleck, L. D., and J. P. Cassady. 2005. Unexpected estimates of variance components with a true model containing genetic competition effects. J. Anim. Sci. 83: 68–74.

Van Vleck, L. D., Cundif, L. V. and R. M. Koch. 2007. Effect of competition on gain in feedlot bulls from Hereford selection lines. J. Anim. Sci. 85:1625-1633.

Wald, A. 1943. On the efficient design of statistical investigations. The Annals of Math. Stats. 14:134-140.

Willham, R. L. 1963. The covariance between relatives for characters composed of components contributed by related individuals. Biometrics 19: 18-27.

Wright, A. J. 1986. Individual and group selection with competition. Theoretical and Applied Genetics 72: 256-263.