ORIGINAL ARTICLE

On identifiability of (co)variance components in animal models with competition effects

R.J.C. Cantet[1,2] & E.P. Cappa[1]

1 Departamento de Producción Animal Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina
2 Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

Summary

There is an increased interest in estimating the (co)variance components of additive animal models with direct and competition effects (AMC). However, some attempts to estimate the dispersion parameters in different animal species faced problems of convergence or inaccurate estimates when pen effects entered the model. We argue that the problem relates to lack of identifiability of the (co)variance components in some AMC. The check for identifiability of the dispersion parameters in mixed models with linear (co)variance structure requires that all the eigenvalues of the restricted maximum likelyhood information matrix $(I(\theta))$ be positive. We show, by way of simple numerical examples, that the singularity of $I(\theta)$ is due to confounding between fixed pen effects and the additive competition effects (SBVs). It is also observed that setting pen effects as random does not always remedy the collinearity with SBVs. An alternative AMC is presented in which the incidence matrix of the SBVs can be written as a function of the 'intensity of competition' (IC) among animals in the same pen. Examples are presented in which the ICs are related to time. The distribution of families of full and half sibs across pens also plays a role in the identifiability and asymptotic variances of the (co)variance components.

Introduction

Recently, there has been a growing interest in genetic evaluation models including additive genetic effects of competition (AMC). Bijma *et al.* (2007) and Allen *et al.* (2008) used the more general expression 'social effects' to refer to these associative or indirect effects. Although quantitative genetics theory dealing with interacting effects in plants and animals dates back to Griffing (1967), it is only in the recent past that social or competition effects have been introduced in genetic evaluation using a mixed model framework (Arango *et al.* 2005; Muir

2005; Van Vleck & Cassady 2005; Bijma *et al.* 2007; Allen *et al.* 2008; Cappa & Cantet 2008). The appeal of AMC relates to the increase in either total performance, or animal wellbeing, if genetic competition effects are selected against (Muir 2005; Van Vleck & Cassady 2005). However, estimation of the dispersion parameters was problematic when the AMC was fitted to animal data. In this respect, Arango *et al.* (2005) observed that convergence to a false local maximum occurred when the analysis was initiated with different starting values, and there was no clear indication of convergence for the estimate of the competition additive variance. Also, error

variance seemed to be overestimated when pen effects were left out from the model. In similar developments, the estimates obtained by Van Van Vleck & Cassady (2005) and Van Vleck et al. (2007) suggest that there was confounding between pen (i.e. management effect) and additive competition effects. Moreover, Bijma et al. (2007) found that the data structure used by Wolf (2003) provided 'either very little or no information at all to separate the direct and the associative heritable variance' in an experiment involving flies. On the other hand, Bijma et al. (2007) and Allen et al. (2008) are examples of successful estimation of the (co)variance components using AMC in layer chickens and hens. Also, no problems were observed when estimating the (co)variance components in an individual tree model (Cappa & Cantet 2008). A major difference between the later report and the previous ones using animal data is that elements of the incidence matrix for the random additive competition effects (Z_c) set by Cappa & Cantet (2008) differs from the one used with animal data. As this incidence matrix is related to the covariance structure of the AMC, a possible explanation for the problems observed when estimating the (co)variance components in the AMC using animal data is that the dispersion parameters may not be identifiable. By lack of identifiability of a set of parameters we mean the situation where multiple values of the parameters 'correspond to the same distribution of observable data' (Gustafson 2005). As a result, there are no unique estimates of the (co)variance components in the AMC when fixed pen effects enter into the model and the Z_c matrix reflects an equal weight for all competitors of a given individual. A similar conclusion on lack of identifiability in the AMC with fixed pen effects has been independently arrived at by Misztal & Rekaya (2007).

Usually animal breeders fit animal models to field data based on quantitative genetic theory, assuming that the model parameters are identifiable. Although breeders are usually acquainted with estimability of linear functions for parameters in the mean vector (Searle 1971; section 5.4), they have little or no indication of a method to check for identifiability of the dispersion parameters. Identifiability of the (co)variance components in mixed models with linear covariance structure, such as in the AMC, has been discussed by Brown (1984) and formalized by Jiang (1996). Our objective is to show how to check for identifiability of the dispersion parameters in animal models with competition effects using the developments set by Jiang (1996). Furthermore, by way of

simple numerical examples, we show how the incidence matrices X (including pen effects) and Z_c play a role on the identifiability of the (co)variance components in the AMC.

Methods

The animal model with competition effects

A model equation for the AMC is as follows:

$$\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} \;+\; \mathbf{Z}_d\,\mathbf{a}_d + \mathbf{Z}_c\,\mathbf{a}_c + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is the $n \times 1$ vector of records; $\mathbf{X}$ is a full rank $n \times p$ incidence matrix that relates the observations to the vector of fixed effects β, and e ($n \times 1$) is the random vector of errors distributed as $N_n(\mathbf{0}, \mathbf{I}_n\, \sigma^2_e)$, 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 (SBVs, as used by Bijma et al. 2007) is $\boldsymbol{a}_c$. We assume that the same q individuals in $\boldsymbol{a}_d$ have SBVs 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 $\mathbf{Z}_d$ and $\mathbf{Z}_c$, respectively. Each row of $\mathbf{Z}_d$ has all elements equal to 0 except for a 1 in the column belonging to $\boldsymbol{a}_{di}$. Matrix $\mathbf{Z}_c$ is described in detail below. The variances for direct and competition breeding values are σ^2_{Ad} and σ^2_{Ac}, respectively, whereas the covariance between both types of additive effects is σ_{AdAc}. Also, let 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 A_{ij}. We can write the total additive covariance matrix in a more compact manner as follows

$$\mathbf{Var}\begin{bmatrix}\mathbf{a}_d \\ \mathbf{a}_c\end{bmatrix} = \begin{bmatrix}\sigma^2_{Ad} & \sigma_{AdAc} \\ \sigma_{AdAc} & \sigma^2_{Ac}\end{bmatrix} \otimes \mathbf{A} = \mathbf{G}_0 \otimes \mathbf{A}$$

A point of note is that we have chosen not to include in (1) environmental competition effects (see for example expression (1) in Bijma et al. 2007). Had we done so, it would compound the problems of identifiability of the (co)variance components. Taking into account the random effects in model (1), the (co)variance matrix (V) of y is given by:

$$\begin{aligned}V = {}& \mathbf{Z}_d\mathbf{A}\mathbf{Z}'_d\,\sigma^2_{Ad} \;+\; \left(\mathbf{Z}_d\mathbf{A}\,\mathbf{Z}'_c \;+\; \mathbf{Z}_c\,\mathbf{A}\,\mathbf{Z}'_d\right)\sigma_{AdAc} \\ & +\; \mathbf{Z}_c\mathbf{A}\mathbf{Z}'_c\,\sigma^2_{Ac} \;+\; \mathbf{I}_n\,\sigma^2_e\end{aligned} \tag{2}$$

Although looking similar to other animal models, notably those including maternal effects, the (co)variance components in model (1) within the covariance matrix (2) may not be identifiable as we observe below.

Identifiability of the (co)variance components

To formalize identifiability of the dispersion parameters, let $\theta = [\theta_i]$ be the parametric vector of the (co)variance components ($i = 1, ..., k$) of a mixed model. Also let $V_1, ..., V_i, ..., V_k$ be the associated (co)variance structures, so that

$$Var(y) = V_1\theta_1 + V_2\theta_2 + ... + V_k\theta_k.$$

Jiang (1996) stated that a mixed model is identifiable of its (co)variance components under translation invariance (IDI), if: 1) matrices $V_1, V_2, ..., V_k$ are linearly independent; 2) there exists a matrix P such that $PX = 0$ and rank$[P] = n − \text{rank}[X]$. Then, $P = V^{-1} − V^{-1} X (X' V^{-1} X)^{-1} XV^{-1}$ (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 restricted maximum likelyhood (REML) information matrix ($I(\theta)$) is positive. As the inverse of the information matrix is the asymptotic covariance matrix of REML estimates (Harville 1977; Searle *et al.* 1992), the previous statement is tantamount to $I(\theta)$ 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 $I(\theta)$ (Harville 1977) is

$$I_{ij}(\theta) = tr\left(P\frac{\partial V}{\partial \theta_i} P\frac{\partial V}{\partial \theta_j}\right) = tr\left(PV_iPV_j\right) \quad (3)$$

For model (1) $I(\theta)$ is 4×4, and the covariance structures are $V_1 = Z_d A Z_d'$, $V_2 = Z_d A Z_c' + Z_c A Z_d'$, $V_3 = Z_c A Z'$, and $V_4 = I_n$. Diagonal elements of $I(\theta)$ are

$$I_{11}(\theta) = tr\left(PZ_dAZ_d'PZ_dAZ_d'\right)$$
$$I_{22}(\theta) = tr\left[P\left(Z_dAZ_c' + Z_cAZ_d'\right)P\left(Z_dAZ_c' + Z_cAZ_d'\right)\right]$$
$$I_{33}(\theta) = tr\left(PZ_cAZ_c'PZ_cAZ_c'\right)$$
$$I_{44}(\theta) = tr(PP).$$
(4)

Whereas off-diagonal elements of $I(\theta)$ are equal to

$$I_{12}(\theta) = tr\left[PZ_dAZ_d'P\left(Z_dAZ_c' + Z_cAZ_d'\right)\right],$$
$$I_{13}(\theta) = tr\left[PZ_dAZ_d'PZ_cAZ_c'\right] \quad I_{14}(\theta) = tr\left(PZ_dAZ_d'P\right)$$
$$I_{23}(\theta) = tr\left[P\left(Z_dAZ_c' + Z_cAZ_d'\right)PZ_cAZ_c'\right]$$
$$I_{24}(\theta) = tr\left[P\left(Z_dAZ_c' + Z_cAZ_d'\right)P\right]$$
$$I_{34}(\theta) = tr\left[PZ_cAZ_c'P\right]$$
(5)

Harville (1977) showed how to write down the elements of $I(\theta)$ in terms of matrices related to the mixed model equations of Henderson (1984). Boca & 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 numerical examples, we will Calculate $I(\theta)$ using expressions (4) and (5) to determine whether a given experimental design allows estimating separately the four (co)variance components.

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

$$I\begin{bmatrix}\sigma_A^2 \\ \sigma_e^2\end{bmatrix} = \begin{bmatrix} tr[PZAZ'PZAZ'] & tr[PZAZ'P] \\ tr[PZAZ'P] & tr[PP] \end{bmatrix} \quad (6)$$

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

$$V = ZAZ'\sigma_A^2 + I\sigma_e^2 = I\left(\sigma_A^2 + \sigma_e^2\right)$$

and the information matrix in (6) is

$$I\begin{bmatrix}\sigma_A^2 \\ \sigma_e^2\end{bmatrix} = \begin{bmatrix} tr[PP] & tr[PP] \\ tr[PP] & tr[PP] \end{bmatrix}$$

As all the elements of $I(\theta)$ are equal, the matrix is singular. This, in turn, indicates that there is not enough information in the data structure to estimate σ_A^2 and σ_e^2 separately. Jiang's (1996) condition can be easily check as the covariance structures of σ_A^2 and σ_e^2 are both equal to I and complete linear dependence occurs.

A numerical example of lack of identifiability in AMC

In this section we describe some of the difficulties involved when estimating all four (co)variance components from model (1). We will take the data and family structure of an example by Muir (2005) but modified for the purpose of making our point clear. Muir (2005) allocated animals randomly to pens. However, consider what would have happened had they use the following scenario: assume there are eight individuals with data divided equally into two 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 $\sigma_{Ad}^2 = \sigma_{Ac}^2 = \sigma_e^2 = 1$ and $\sigma_{AdAc} = -0.25$. Using the specification for (1) as in

Muir (2005), matrices X, Z_d and Z_c are respectively equal to

$$
X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad Z_d = I_8 \qquad Z_c = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{7}
$$

whereas the additive relationship matrix is

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

so that $I(\theta)$ is equal to

$$
\begin{bmatrix}
0.3092515 & -0.514861 & 0.4128936 & 0.5962941 \\
-0.514861 & 1.6515742 & 0.106991 & -1.162976 \\
0.4128936 & 0.106991 & 1.3456717 & 0.6259061 \\
0.5962941 & -1.162976 & 0.6259061 & 1.1862428
\end{bmatrix}
$$

The eigenvalues of $I(\theta)$ are 2.97, 1.52, 0, and 0; therefore, the matrix is singular and not all dispersion parameters can be estimated separately. A closer inspection of the example shows that matrices X and Z_c in (7) are collinear: each column of X is a linear combination of the columns of Z_c. The first column of X can be obtained by taking a third of the sum of the first four columns of Z_c, whereas the second column of X results of one third of the sum of the last four columns of Z_c. We attempted several small designs trying to avoid confounding families with pens, but found none in which the associated $I(\theta)$ was non-singular. The fact that contemporary formed by full-sibs produces a lack of identifiability of the (co)variance components was first shown by Bijma *et al.* (2007) when discussing the design used by Wolf (2003).

The animal model with competition effects and random contemporary groups

To avoid the trouble of collinearity between X and Z_c one may consider pen to be a random effect. However, this may be deceptive as the resulting model may not cope with the problem. Going back to the example, we took the pen effects to be random with a variance component of two, and set a common mean as the only fixed effect into the model. The eigenvalues of the ensuing $I(\theta)$ were equal to 2.97, 0.59, 0, 0 and 0. A further attempt to obtain identifiability consisted of arbitrarily setting the 2×2 variance-covariance matrix of pen effects to be equal to $\begin{bmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{bmatrix}$. However, this was incon sequential to identifiability as the eigenvalues of $I(\theta)$ were now 2.97, 0.79, 0, 0 and 0. Hence, setting pen effects as random does not always remedy the collinearity between X and Z_c. Therefore, for any particular data structure one will have to check whether the eigenvalues of $I(\theta)$ are positive to be sure the (co)variance components are identifiable.

The genetic data structure in the experiment by Wolf (2003)

Wolf (2003) performed an experiment to estimate the (co)variance components for the trait fly length in *Drosophila melanogaster*. Bijma *et al.* (2007) analysed the design of the genetic data structure in detail and concluded that it 'provides either very little or no information at all to separate' the direct and competition variances. They also produced simulated data of the genetic structure used by Wolf (2003) and found that the matrix of second derivatives was singular. We further exemplify the arguments of Bijma *et al.* (2007) with a small example. In the experiment

of Wolf (2003) pairs of flies were kept in tubes, which act like pens. All flies from the parental generation, either male or female, were unrelated. Any single male was mated to two females only, and the trait was measured in the progenies of these matings. Thus the family structure is composed of unrelated families, with the progeny of any family being either full-sibs or half-sibs. As a result, the additive covariances among animals in the progeny generation involved those of full-sibs, half-sibs or unrelated individuals, either in the same or in different tubes. The pairs inside any tube were full-sibs, half-sibs, or unrelated flies. In the latter case two half-sib families were combined in a pair of tubes. Bijma *et al.* (2007)

covariance between flies in different tubes, then Bijma *et al.* (2007) obtained $\frac{1}{2} \left(\sigma^2_{Ad} + \sigma^2_{Ac} + \sigma_{AdAc} \right)$. For the purpose of mimicking the experimental setup of Wolf (2003) the following example including 14 individuals in four unrelated families was prepared. Flies 1 to 4 are full-sibs from the mating of Sire 1 with Dam 1. The same sire mated with Dam 2 produces flies 5 and 6. Therefore, individuals 1 to 4 are half-sibs of 5 and 6. In the second family, Sire 2 mates with Dam 3 to produce full-sibs 7 and 8, and with Dam 4 to sire individuals 9 and 10. Thus, 7 and 8 are half-sibs of 9 and 10. Finally, fly 11 is a half-sib of 12 and 13 of 14. All other additive relationships were 0. The resulting A matrix is then equal to

$$
A = \begin{bmatrix}
1 & 0.5 & 0.5 & 0.5 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 1 & 0.5 & 0.5 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 0.5 & 1 & 0.5 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 0.5 & 0.5 & 1 & 0.25 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.25 & 0.25 & 0.25 & 0.25 & 1 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.25 & 0.25 & 0.25 & 0.25 & 0.5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.5 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 1 & 0.25 & 0.25 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.25 & 1 & 0.5 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0.25 & 0.5 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.25 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.25 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 1
\end{bmatrix}
$$

obtained the expectations of these covariances in terms of the (co)variance components for direct and competition effects (see Table A1 therein, p. 298). Some of these covariances are impossible to occur,

The pairs included in the tubes were: 1 with 2, 3 with 4, 5 with 6, 7 with 9, 8 with 10, 11 with 13, and 12 with 14. The resulting Z_c matrix is equal to

$$
Z_c = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}
$$

for example the covariance of full-sibs in the same tube, when the tube composition consists of a pair of half-sibs. If, instead, one is looking at the full-sib

As a consequence of this distribution, seven of the nine non-zero covariances described by Bijma *et al.* (2007) in Table A1 are present in the data

structure: (1) covariance between full-sibs in the same tube when tube composition consist of a pair of full-sibs (e.g. 1 and 2); (2) covariance of full-sib in different tubes when tubes are composed of pairs of full-sibs (e.g. 1 and 3); (3) covariance of half-sibs in different tubes when tubes are composed of full-sib pairs (e.g. between 1 and 5); (4) covariance between full-sibs in different tubes when tube composition consist of a pair of half-sibs (e.g. 7 and 8); (5) covariance of half-sibs in the same tube when tubes are composed of half-sib pairs (e.g. between 7 and 9); (6) covariance of half-sibs in different tubes when tubes are composed of pairs of half-sib (e.g. between 7 and 10); (7) covariance between unrelated individuals in the same tube when two half-sibs are located in different tubes and their respective competitors are half-sibs (e.g. between 11 and 13). A slight discrepancy with the covariances in Table A1 of Bijma *et al.* (2007) was found for the case of unrelated individuals located in different tubes. Although for most cases this covariance is equal to zero (e.g. between 1 and 14), if the unrelated individual compete with a half-sib of the other fly in the tube, then it is equal to $\frac{1}{2}\sigma_{AdAc}$ instead of 0 (e.g. between 11 and 14). In the resulting analysis, the eigenvalues of $I(\theta)$ were equal to 6.67, 3.66, 0.22, and 0, and the information matrix was singular.

An alternative animal model with competition effects

A closer look at the problem suggests that the key to avoid the collinearity between X and Z_c in model (1) are the non-zero elements of matrix Z_c. We now present a model with individual breeding values that has been used in tree breeding to fit SBVs. In doing so we write a_{cj} for the SBV of animal j that is expressed in the record of animal i. Let $j = 1, 2, \ldots,$ m_i be the competitors of i and $i = 1, 2, \ldots, m_j$ be the individuals with records that animal j competes with. Cappa & Cantet (2008) observed that the total additive genetic competition that is exerted over the phenotype of individual i from its competitors can be written as

$$f_{i1}\,\boldsymbol{a}_{c_1} + f_{i2}\,\boldsymbol{a}_{c_2} + \cdots + f_{im}\,\boldsymbol{a}_{c_m} = \sum_{j=1}^{m_i} f_{ij}\,\boldsymbol{a}_{c_j} \qquad (8)$$

The f_{ij}s in (8) are interpreted as the *intensities of competition* (IC) that the SBVs of 1, 2,..., m_i exert over the phenotype of $i(y_{ic_1 \ldots c_m})$, and are the

non-zero elements in row i of $\boldsymbol{Z}_c$. Thus, $y_{ic_1 \cdots c_m}$ is written as

$$\begin{aligned} y_{ic_1 \cdots c_{m_i}} = {} & x_i{}'\,\beta + \boldsymbol{a}_{d_i} + f_{i1}\,\boldsymbol{a}_{c_1} \\ & + f_{i2}\,\boldsymbol{a}_{c_2} + \cdots + f_{i\,m_i}\,\boldsymbol{a}_{c_{m_i}} + e_{ic_1 \cdots c_{m_i}} \end{aligned} \qquad (9)$$

Muir (2005) produced an expression akin to (9) (his equation (6) on p. 1250), although the values of his coefficients are unconstrained. Whereas Cappa & Cantet (2008) imposed two restriction, which we will denote as (I), in order to achieve identifiability of the (co)variance components. To introduce this restriction consider the following expression that is obtained in the Appendix [(A3)] for the additive genetic variance of the SBVs in model (9)

$$Var\left(\sum_{j=1}^{m_i} f_{ij}\,\boldsymbol{a}_{c_j}\right) = \left[\sum_{j=1}^{m_i} f_{ij}^2\left(1 + F_j\right) + 2\sum_{j\neq j'}^{m_i} f_{ij}f_{ij'}\,\boldsymbol{A}_{jj'}\right]\sigma_{Ac}^2 \qquad (10)$$

Observe that in the absence of inbreeding ($F_j = 0$, for all j) and genetic relationships among competitors ($A_{jj'} = 0$ for all $j \neq j'$), and regardless of the number of competing individuals, expression (10) is equal to

$$Var\left(\sum_{j=1}^{m_i} f_{ij}\,\boldsymbol{a}_{c_j}\right) = \left(\sum_{j=1}^{m_i} f_{ij}^2\right)\sigma_{Ac}^2 \qquad (11)$$

To standardize the variance of the SBVs within the phenotypic variance of animal i, we impose the restriction

$$\sum_{j=1}^{m_i} f_{ij}^2 = 1 \qquad (I)$$

This is to say that the sum of squares of the ICs of the competitors adds up to 1 within the phenotype of i. Thus, under no inbreeding of competitors, no relationship among competitors, and after restriction (I), expression (11) results to be equal to

$$Var\left(\sum_{j=1}^{m_i} f_{ij}\,\boldsymbol{a}_{c_j}\right) = \sigma_{Ac}^2 \qquad (12)$$

Expression (12) is not a definition of the additive variance for the SBVs but a consequence of imposing restriction (I) to the ICs, in the absence of

inbreeding and relationships among competitors. If $F_j > 0$ for any j and/or $A_{jj'} > 0$ for any $j \neq j'$, then expression (10) should be employed instead of (12), however the ICs still have to comply with the restriction (I).

We now show how the f_{ij}s can be calculated when the breeder has the opportunity to supervise the collection of data for estimating the (co)variance components. In doing so we follow the approach of Cappa & Cantet (2008) who showed how to obtain the IC for competing trees planted in a regular grid. The ICs are functions of some *external* variable. By external we mean independent of any intrinsic characteristic of the animal, especially of its breeding value. Just for the purpose of exemplifying 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, 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 at the start of the third period. Two

By equating (12) to (13), the ICs are such that

$$1 = \sum_{j=1}^{m_i} f_{ijk}^2 = n_1 f_{i1}^2 + \cdots + n_k f_{ik}^2 + \cdots + n_{m_i} f_{im_i}^2 \tag{14}$$

To solve (14) we have to use the relationships among ICs based on the fraction of time animals were competing. In 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 two animals in sequence 1-1-1 and four in sequence 1-2-1, then expression (14) for an animal in sequence 1-1-1 is equal to $f_{11}^2 + 4 f_{12}^2 = 1$. On replacing in the last expression with $f_{12} = (2/3) f_{11}$, we have $f_{11}^2 + 4 [(2/3) f_{11}]^2 = 1$, with solutions equal to $f_{11} = 3/5$ and $f_{12} = 2/5$. For an individual in sequence 1-2-1, (14) is now equal to $2 f_{21}^2 + 3 f_{22}^2 = 1$ 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 Z_c is then equal to

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

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 one-third less intense than 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

$$Var\left(\sum_{j=1}^{m_i} f_{ij} \, \boldsymbol{a}_{c_j} \right) = \left(n_1 f_{i1}^2 + \cdots + n_k f_{ik}^2 + \cdots + n_{m_i} f_{im_i}^2 \right) \sigma_{Ac}^2 \tag{13}$$

This type of structure for Z_c allows estimating the variance components, as we discuss below. The ICs can be interpreted as weighting factors expressing how intense pairs of animals compete in relation to all other animals, and can be chosen to represent extreme patterns in which only particular individuals display competition behaviour whereas the rest do not. Non-linear relationship between time and ICs may also be considered.

As a follow-up of our little example, 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 IC for animals in the same half-pen is greater than the one among individuals from different halves. For any animal in a half-pen expression (14) is equal to $f_{11}^2 + 2 f_{12}^2 = 1$. Assuming than competition between the two animals in the same half-pen is twice as intense as with the

two individuals in the other half-pen, we can impose the restriction $f_{11} = 2\,f_{12}$. 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 $f_{56} = f_{57} = 2\,f_{58}$, as 5 (sequence 1-1) compete with 8 (in sequence 2-2) less than with 6 or 7 that are in sequences 1-2 or 2-1. Solutions are $f_{56} = 2/3$ and $f_{58} = 1/3$, so that matrix Z_c is then equal to

$$Z_c = \begin{bmatrix} 0 & \sqrt{2/3} & \sqrt{1/6} & \sqrt{1/6} & 0 & 0 & 0 & 0 \\ \sqrt{2/3} & 0 & \sqrt{1/6} & \sqrt{2/3} & 0 & 0 & 0 & 0 \\ \sqrt{1/6} & \sqrt{1/6} & 0 & \sqrt{2/3} & 0 & 0 & 0 & 0 \\ \sqrt{1/6} & \sqrt{1/6} & \sqrt{2/3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/3 & 2/3 & 1/3 \\ 0 & 0 & 0 & 0 & 2/3 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 0 & 2/3 & 1/3 & 0 & 2/3 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 & 2/3 & 0 \end{bmatrix} \tag{15}$$

Now $I(\theta)$ is non-singular with eigenvalues 2.70, 1.65, 0.25, and 5×10^{-5}. Notice that the last eigenvalue is very small, which in turn indicates the presence of large asymptotic variances of the REML estimates when analyzing data with this design.

How can the asymptotic variances be improved upon? One possibility is to use another 'genetic design', thus modifying the A matrix. Suppose that four families with two full-sibs each are available so that and matrices X, Z_d and Z_c are the same as before.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Then, the eigenvalues of the resulting $I(\theta)$ are 2.64, 1.72, 0.28, and 0.12. A possible criterion to compare the efficiency of different experimental designs is D-optimality (Wald 1943), and amounts to maximizing the determinant of $I(\theta)$ ($|I(\theta)|$). If Muir (2005) had used a design where animals were not randomly allocated to pens, then the value of $|I(\theta)|$ would be equal to 3.86×10^{-5}, whereas for the design with four families of two full-sibs each and animals distributed across sub-pens $|I(\theta)| = 0.159$. The difference in efficiency is important: 4111 times ($0.1591/3.86 \times 10^{-5}$). 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 Z_c as in (15), keeping X, Z_d and A as before. Notice that 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 Z_c is block-diagonal with the same structure as in the upper block of (15). 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 Z_c is block-diagonal with the same structure as in the lower block of (15). Then, $|I(\theta)| = 0.178$ for the 1-1/2-2 design, and $|I(\theta)| = 0.119$ for the 1-2/2-1 design. Therefore, the 1-1/2-2 design is 12% more efficient than 1-1/1-2 and almost 50% more efficient than 1-2/2-1.

Discussion

The two major issues that were brought up in the current research are the check of identifiability of the dispersion parameters in animal models with linear covariance structures using $I(\theta)$, and the problems of identifiability in the AMC due to potential collinearity between X and Z_c. With regard to identifiability, Rothenberg (1971) showed that identifiability of an unknown parameter vector θ is equivalent to non-singularity of the information matrix $I(\theta)$. Brown (1984) observed that the variance components are identifiable in a mixed model if the covariance structures are linearly independent. However, Jiang (1996) put both ideas together by formalizing identifiability of the dispersion parameters in a mixed model with linearly independent covariance structures and by checking that the eigenvalues of $I(\theta)$ are positive. We have used this result in the current research to check for identifiability of the (co)variance components in the AMC. The expressions employed to calculate $I(\theta)$ were (4) and (5). For most

data sets used by animal breeders to estimate dispersion parameters with the AMC, the calculus of $I(\theta)$ by (4) and (5) is a sizeable computational task, if not an impossible one. We have obtained expressions for the elements of $I(\theta)$ from the AMC in terms of matrices that arise from solving the mixed model equations. These formulae still look formidable and difficult to compute, and they are available on request. Further research is needed to find computing formulae for the eigenvalues of $I(\theta)$ in models such as the AMC for large data sets. A point of note is that the covariance structures involved in the lemma 4.1 of Jiang (1996) are linear in the dispersion parameters. Therefore, the check of the eigenvalues of $I(\theta)$ can not be used with an animal model if the error structure follows, for example, an AR(1) process or the structured antedependence model. An alternative to an AR(1) model is the covariance structure (P) presented by Cantet *et al.* (2005) which can be seen as a linear approximation to an AR(1) process, with the added advantages that the variable may be measured at unequal time lags and that the inverse matrix is readily available without ever computing P.

The other issue the paper focused on is the identifiability of (co)variance components in the AMC for animal data, and the confounding between pen effects with SBVs. 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 fixed pen effects are in the model. Notice that treating pen as random might result in identifiable dispersion parameters but large asymptotic variance of REML estimates, and this will deliver inaccurate estimates of the parameters. For example, Van Vleck *et al.* (2007) fitted different AMC to data on average daily gains from Hereford bulls and found an estimate of the pen variance of 34.74. When additive competition effects were left out from the model estimated pen variance was equal to 108.08. Inclusion of environmental competition effects reduced estimated pen variance to 0.01. The confounding is more evident as environmental competition effects have an incident matrix with the same non-zero columns as Z_c, and there is an extra parameter to estimate. Also, Allen *et al.* (2008) obtained some approximate asymptotic standard errors of the REML estimates of σ^2_{Ac} and σ_{AdAc} that were larger than the estimates themselves (see Table 5, therein). Finally, treating CGs as random

may result in: (1) biased prediction of breeding values in the presence of selection, (2) erratic changes in accuracy if there is a non-random association of sires to CGs (Visscher & Goddard 1993).

Further improvement in identifiability or in the size of the asymptotic variances of the dispersion parameters can be attained by distributing full or half sib to different pens. However, one has to calculate the eigenvalues of $I(\theta)$ to be sure all (co)variance components are identifiable. Interestingly enough Bijma *et al.* (2007) produced a successful estimate of the (co)variance components for survival days in a population of laying hens using an AMC. In this case, cage (i.e. the pen effect) was not directly included in the model but the row position of the cage (a classification variable with 8 levels) and the level of the cage (top, middle and bottom). Cleverly enough, this parameterization of the fixed effects avoids confounding cage effects with matrix Z_c. However, a correlation was fitted for the residuals of animals housed in the same cage. When this correlation is positive the model is tantamount to fitting random cage effects, as discussed by Bijma *et al.* (2007) and Allen *et al.* (2008). Both papers show that there are situations in which random cage (or pen) effects allow identifying the (co)variance components. This parameterization permits estimating the environmental direct variance, and a linear combination of the competition environmental variance and the environmental covariance between both set of effects. In addition, Bijma *et al.* (2007) performed a Monte-Carlo computing experiment and obtained seemingly unbiased estimates of the (co)variance components without simulating any fixed effect. The use of different values for the ICs, as presented by Cappa & Cantet (2008), avoids the confounding between the pen effects and Z_c. However, if ICs are defined as a function of time, an experimental design that allow disentangling pen and competition effects should be defined before data collection. This may be difficult to achieve with field data. Any variable used to define the ICs should be independent of the SBVs, as the inference is conditional on the values of the ICs. Hadfield & Wilson (2007) have proposed modelling the additive competition effect as a regression on the inverse of the number of animals in the pen minus one. While this measure resembles the ICs, it falls short of a solution for identifiability as all individuals in the same pen will have the same IC. Then, any column associated with a pen effect in X is equal to the sum of all columns of Z_c related to the animals from that particular pen.

It is worth mentioning that identifiability may be an issue for the estimation of (co)variance components but not for the prediction of breeding values through BLUP via the mixed model equations (Henderson 1984), which is conditional on the 'true' dispersion parameters.

Acknowledgements

References

Allen E.D., Visscher J., Van Arendonk J.A.M., Bijma P. (2008) Survival of laying hens: genetic parameters for direct and associative effects in three purebred layer lines. *Poultry Sci.*, **87**, 233–239.

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

Bijma P., Muir W.M., Allen E.D., Wolf J.B., Van Arendonk J.A.M. (2007) Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics*, **175**, 289–299.

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

Brown K.G. (1984) On analysis of variance in the mixed model. *Ann. Stats.*, **12**, 1488–1499.

Cantet R.J.C., Birchmeier A.N., Canaza Cayo A.W., Fioretti C. (2005) Semiparametric animal models via penalized splines as alternatives to contemporary groups. *J. Anim. Sci.*, **83**, 2482–2494.

Cappa E.P., Cantet R.J.C. (2008) Direct and competition additive effects in tree breeding: Bayesian estimation from an individual tree mixed model. *Silvae Genetica*, **57**, 45–57.

Griffing B. (1967) Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Aust. J. Biol. Sci.*, **20**, 127–139.

Gustafson P. (2005) On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables. *Stat. Sci.*, **20**, 111–140.

Hadfield J.D., Wilson A.J. (2007) Multilevel selection 3: Modeling the effects of interacting individuals as a function of group size. *Genetics*, **175**, 667–668.

Harville D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.*, **72**, 320–340.

Henderson C.R. (1984) Applications of Linear Models to Animal Breeding. University of Guelph, Guelph, ON.

Jiang J. (1996) REML estimation: asymptotic behavior and related topics. *Ann. Stats.*, **24**, 255–286.

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

Misztal I., Rekaya R. (2007) Categorical expression of social competition. In: van der Honing Y. (Ed.), Book of Abstracts of the 58th Annual Meeting of the European Association of Animal Production, 58th Annual Meeting of the EAAP, 26–29 August, 2007, Dublin, 172.

Muir W.M. (2005) Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics*, **170**, 1247–1259.

Rothenberg T.J. (1971) Identification in parametric models. *Econometrica*, **39**, 577–591.

Searle S. (1971) Linear Models. Wiley, New York, USA.

Searle S, Casella G., McCullough C. (1992) Variance Components. Wiley, New York, USA.

Van Vleck L.D., Cassady J.P. (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., Koch R.M. (2007) Effect of competition on gain in feedlot bulls from Hereford selection lines. *J. Anim. Sci.*, **85**, 1625–1633.

Visscher P., Goddard M.E. (1993) Fixed and random contemporary groups. *J. Dairy Sci.*, **76**, 1444–1454.

Wald A. (1943) On the efficient design of statistical investigations. *Ann. Math. Stat.*, **14**, 134–140.

Wolf J.B. (2003) Genetic architecture and evolutionary constraint when the environment contains genes. *Proc. Natl Acad. Sci. USA*, **100**, 4655–4660.

Appendix A

Derivation of the total additive genetic variance with additive competition effects in model (9)
The additive genetic variance for direct and competition breeding values in model (9) is:

$$
\text{Var}\left(a_{d_i} + \sum_{j=1}^{m_i} f_{ij} a_{c_j}\right) = \text{Var}(a_{d_i}) + \text{Var}\left(\sum_{j=1}^{m_i} f_{ij} a_{c_j}\right)
$$
$$
+ 2 \, \text{Cov}\left(a_{d_i}, \sum_{j=1}^{m_i} f_{ij} a_{c_j}\right)
$$

$$(A1)$$

Following Kempthorne (1969, p. 349), the variance in the first term in (A1) is

$$
\text{Var}(a_{d_i}) = (1 + F_i)\sigma^2_{Ad} \tag{A2}
$$

For the second term in (A1) we have

$$
\text{Var}\left(\sum_{j=1}^{m_i} f_{ij} a_{c_j}\right) = \text{Var}(f_{i1} a_{c_1} + \cdots + f_{im} a_{c_m})
$$
$$
= \text{Var}(f_{i1} a_{c_1}) + \cdots + \text{Var}(f_{im} a_{cm}) + 2[\text{Cov}(f_{i1} a_{c_1}, f_{i2} a_{c_2}) + \cdots + \text{Cov}(f_{im-1} a_{cm-1}, f_{im} a_{cm})]
$$
$$
= f_{i1}^2 \, \text{Var}(a_{c_1}) + \cdots + f_{im}^2 \, \text{Var}(a_{cm}) + 2\left[f_{i1} f_{i2} \, \text{Cov}(a_{c_1}, a_{c_2}) + \cdots + f_{im-1} f_{im} \text{Cov}\left(a_{c_{m-1}}, a_{c_m}\right)\right]
$$
$$
= \sum_{j=1}^{m_i} f_{ij}^2 \, \text{Var}\left(a_{c_j}\right) + 2 \sum_{j \neq j'}^{m_i} f_{ij} f_{ij'} \, \text{Cov}\left(a_{c_j}, a_{c_{j'}}\right)
$$
$$
= \sum_{j=1}^{m_i} f_{ij}^2 (1 + F_j)\sigma^2_{Ac} + 2 \sum_{j \neq j}^{m_i} f_{ij} f_{ij'} A_{jj'} \sigma^2_{Ac}
$$

where $A_{jj'}$ indicates the relationship between competitors j and j'. Therefore

$$
\text{Var}\left(\sum_{j=1}^{m_i} f_{ij} a_{c_j}\right) = \left[\sum_{j=1}^{m_i} f_{ij}^2 (1 + F_j) + 2 \sum_{j \neq j'}^{m_i} f_{ij} f_{ij'} A_{jj'}\right]\sigma^2_{Ac} \tag{A3}
$$

For the third term in (A1) we use the covariance operator so that

$$
\text{Cov}\left(a_{d_i}, \sum_{j=1}^{m_i} f_{ij} a_{c_j}\right) = \text{Cov}(a_{d_i}, f_{i1} a_{c_1} + \cdots + f_{im} a_{c_m})
$$
$$
= \text{Cov}(a_{d_i}, f_{i1} a_{c_1}) + \cdots + \text{Cov}(a_{d_i}, f_{im} a_{c_m})
$$
$$
= f_{i1} \text{Cov}(a_{d_i}, a_{c_1}) + \cdots + f_{im} \text{Cov}(a_{d_i}, a_{c_m})
$$
$$
= \sum_{j=1}^{m_i} f_{ij} A_{ij} \, \sigma_{AdAc}
$$

$$(A4)$$

Replacing in (A1) with (A2), (A3) and (A4) gives the expression for the total additive variance within the variance of $y_{ij1..jm.}$ as follows

$$
\text{Var}\left(a_{d_i} + \sum_{j=1}^{m_i} f_{ij} a_{c_j}\right) = (1 + F_i)\sigma^2_{Ad}
$$
$$
+ \left[\sum_{j=1}^{m_i} (1 + F_j) + 2 \sum_{j \neq j}^{m_i} f_{ij} f_{ij'} A_{jj'}\right]\sigma^2_{Ac}
$$
$$
+ 2 \sum_{j=1}^{m_i} f_{ij} A_{ij} \sigma_{AdAc}
$$

$$(A5)$$