Estimation of Genetic and Phenotypic Covariance Functions for Body Weight as Longitudinal Data of SD-II Swine Line

Wenzhong Liu*, Guoqing Cao, Zhongxiao Zhou and Guixian Zhang
College of Animal Science and Technology, Shanxi Agricultural University, Taigu 030801, P. R. China

ABSTRACT: Growth records over six generations of 686 pigs in SD-II Swine Line were used to estimate the genetic and phenotypic covariance functions for body weight as longitudinal data. A random regression model with Legendre polynomials of age as independent variables was used to estimate the (co)variances among the regression coefficients, thus the coefficients of genetic and permanent environmental covariance functions by restricted maximum likelihood employing the average information algorithm. The results showed that, using litter effect as additional random effect, a reduced order of fit did not describe the data adequately. For all five orders of fit, however, the change trends of genetic and phenotypic (co)variances were very similar from κ=3 onwards. (Asian-Aust. J. Anim. Sci. 2002. Vol 15, No. 5 : 622-626)

Key Words: Swine, Covariance Functions, Longitudinal Data, Restricted Maximum Likelihood, Random Regression Model

INTRODUCTION

An animal's phenotype changes with age. A trait that changes with age can be represented as a trajectory, that is, a function of time. Often, such traits are measured on the same individual(s) at various times or ages. Such records are commonly referred to as longitudinal data and such trajectories as infinite-dimensional characters. The infinite-dimensional model offers several advantages over earlier attempts to adapt quantitative genetics to growth trajectories (Kirkpatrick and Heckman, 1989; Kirkpatrick et al., 1990).

Covariance functions (CFs) have been recognized as a suitable alternative to the conventional multivariate mixed model to describe genetic and phenotypic variation for longitudinal data (Meyer, 1998a). In essence, CFs are the infinite-dimensional equivalents to covariance matrices in a traditional, finite-multivariate analysis (Kirkpatrick et al., 1990).

As the name indicates, a CF describes the covariance between records taken at certain ages as a function of these ages. A suitable class of functions is the family of orthogonal polynomials (Kirkpatrick et al., 1990).

Kirkpatrick et al. (1990, 1994) modeled CFs using orthogonal polynomials of age, choosing Legendre polynomials, and described a generalized least-squares procedure to determine the coefficients of a CF from an estimated covariance matrix. However, Meyer and Hill (1997) thought that this was often not available or computationally expensive to obtain. These authors showed that the coefficients of CFs could be estimated directly from the data by restricted maximum likelihood (REML) through a simple reparameterisation of existing, finite-dimensional multivariate REML algorithms. In the general case, however, this approach required a multivariate mixed model matrix proportional to the number of ages in the data to be set up and factored, even for a reduced order of fit. This severely limited practical applications, especially for data records at all ages. Meyer (1998a,b) described another alternative procedure for the estimation of CFs. It was shown that the CF model was equivalent to a random regression model (RRM) with polynomials of age as independent variables, and that REML estimates of the coefficients of the CF could be obtained as (co)variances among the regression coefficients.

SD-II Line is a synthesized specialized dam line of Duroc and Shanxi Black Swine using a closed nucleus breeding system. This paper estimates the genetic and phenotypic CFs for body weight of the SD-II Line using a RRM so as to reflect the changes of genetic and phenotypic (co)variances in continuous time scale. The resultant genetic parameters from estimated CFs can be further used in genetic evaluation.

MATERIALS AND METHODS

Data

Data on body weight collected from the SD-II Line for six generations were used for this study. The data set comprised 3430 records from 686 pigs, each weighed at birth, 35, 70, 120 and 180 days of age. (table 1). Detailed descriptions of the selection and management procedures followed in this line were given by Zhou et al. (1999).

Random regression model

Let $i^*_j$ denotes the $j$th age for animal $i$ standardized to the range of -1 to 1, and let $\phi_m(i^*_j)$ be the $m$th Legendre
ESTIMATES OF COVARIANCE FUNCTIONS

623

polynomial evaluated for \( t_i^* \). The RRM can then be written as (Meyer, 1998a)

\[
[y_i] = \mathbf{F} + \sum_{m=1}^{k_A} \alpha_m \phi_m(t_i) + \sum_{m=1}^{k_R} \gamma_m \phi_m(t_i) + e_i
\]

with \( y_i \) the observation for animal \( i \) at time \( j \), \( \mathbf{F} \) some fixed effects, \( \alpha_m \) and \( \gamma_m \) representing the \( m \)th additive genetic and permanent environmental random regression coefficients for animal \( i \), respectively, \( k_A \) and \( k_R \) denoting the respective orders of fit and \( e_i \) the measurement error (or temporary environmental effect) pertaining to \( y_i \).

Covariance structure

The covariance between two records for the same animal is then

\[
\text{Cov}(y_i, y_j) = \sum_{m=1}^{k_A} \sum_{l=1}^{k_R} \phi_m(t_i) \phi_l(t_j) \text{Cov}(\alpha_m, \alpha_l) + \sum_{m=1}^{k_A} \sum_{l=1}^{k_R} \phi_m(t_i) \phi_l(t_j) \text{Cov}(\gamma_m, \gamma_l) + \text{Cov}(e_i, e_j)
\]

(2)

Generally, measurement errors are assumed to be identically independent distributed with variance \( \sigma_i^2 \), so that

\[
\text{Cov}(e_i, e_j) = \sigma_i^2 \quad \text{for} \quad j = j \quad \text{and} \quad 0 \quad \text{otherwise}.
\]

REML estimation

Considering all animals, equation (1) can be written in matrix form as

\[
y = \mathbf{Xb} + \mathbf{Z}^T \mathbf{a} + \mathbf{Z}^T \mathbf{e}
\]

with \( y \) the vector of \( N \) observations measured on \( N_D \) animals, \( b \) the vector of fixed effects, \( a \) the vector of \( k_A \times N_A \) additive genetic random regression coefficients (\( N_A \geq N_D \) denoting the total number of animals in the analysis, including parents without records), \( \gamma \) the vector of \( k_R \times N_D \) permanent environmental random regression coefficients, \( e \) the vector of \( N \) measurement errors, and \( \mathbf{X}, \mathbf{Z} \) and \( \mathbf{Z}^T \) denoting the corresponding ‘design’ matrices. The superscript ‘\( * \)’ marks matrices incorporating orthogonal polynomial coefficients.

Analyses were carried out using program DEXPAMR (Meyer, 1998c), which employed the average information REML of Johnson and Thompson (1995) for parameter estimation. A separate measurement error variance component for each days of age (five variances) was fitted. Fixed effects fitted were generation (6 levels), sex (2 levels) and litter size (13 levels). Additional random effect fitted was litter with 220 levels. Additive genetic and permanent environmental CFs were fitted to the same order throughout (\( k_A = k_R = k \) ). Orders of fit considered ranged from 1 to 5. A likelihood ratio test such that \(-2(\log L_1 - \log L_2)\) has a \( \chi^2 \) distribution with \( k_2 - k_1 \) degrees of freedom, where \( k_i \) is the different order of fit and \( L_i \) is the maximum value of the restricted likelihood function, was carried out in order to select the appropriate order of fit. Alternatively, Akaike’s information criterion (Akaike, 1973) that is not dependent on the number of free parameters in the whole model was also used.

RESULTS

Likelihoods and information criterion

Log restricted likelihoods (log \( L \)), Akaike’s information criterion (AIC) and estimates of measurement errors (\( \sigma_i^2 \), \( i = 1, 5 \)), are summarized in table 2. It can be

<table>
<thead>
<tr>
<th>Order of fit</th>
<th>log ( L )</th>
<th>AIC</th>
<th>( \sigma_{e1}^2 )</th>
<th>( \sigma_{e2}^2 )</th>
<th>( \sigma_{e3}^2 )</th>
<th>( \sigma_{e4}^2 )</th>
<th>( \sigma_{e5}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-12210.3</td>
<td>24434.6</td>
<td>1419.2</td>
<td>1063.0</td>
<td>446.7</td>
<td>20.1</td>
<td>1124.0</td>
</tr>
<tr>
<td>2</td>
<td>-6996.8</td>
<td>14015.6</td>
<td>72.2</td>
<td>0.7</td>
<td>10.2</td>
<td>11.4</td>
<td>116.7</td>
</tr>
<tr>
<td>3</td>
<td>-4196.6</td>
<td>8427.2</td>
<td>0.0</td>
<td>2.4</td>
<td>5.1</td>
<td>31.6</td>
<td>9.2</td>
</tr>
<tr>
<td>4</td>
<td>-3655.2</td>
<td>7360.4</td>
<td>0.0</td>
<td>0.7</td>
<td>4.3</td>
<td>8.6</td>
<td>8.8</td>
</tr>
<tr>
<td>5</td>
<td>-3542.6</td>
<td>7155.2</td>
<td>0.0</td>
<td>0.5</td>
<td>1.7</td>
<td>4.5</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Table 1. Summary of the data set used

<table>
<thead>
<tr>
<th>Terms</th>
<th>Body weight at</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Birth</td>
</tr>
<tr>
<td>Number of animals</td>
<td>686</td>
</tr>
<tr>
<td>Number of base animals</td>
<td>108</td>
</tr>
<tr>
<td>Number of records per animal</td>
<td>5</td>
</tr>
<tr>
<td>Mean (kg)</td>
<td>1.14</td>
</tr>
<tr>
<td>Standard deviation (kg)</td>
<td>0.22</td>
</tr>
<tr>
<td>Coefficient of variation (%)</td>
<td>19.54</td>
</tr>
</tbody>
</table>
seen that increasing k from 1 to 5 yields significant increases in log L and decreases in measurement error. Both likelihood ratio tests and Akaike’s information criteria indicate that the appropriate order of fit is 5 and that a reduced order of fit cannot describe the data adequately.

Table 3. Estimates of the coefficients of the genetic CF, A, and the permanent environmental CF, R for different orders of fit

<table>
<thead>
<tr>
<th>k=1</th>
<th>k=2</th>
<th>k=3</th>
<th>k=4</th>
<th>k=5</th>
<th>Coefficients of R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>k=1</td>
</tr>
<tr>
<td>0.0</td>
<td>0.00</td>
<td>8.41</td>
<td>6.35</td>
<td>7.90</td>
<td>10.37</td>
</tr>
<tr>
<td>0.1</td>
<td>10.43</td>
<td>9.83</td>
<td>12.50</td>
<td>13.68</td>
<td>15.45</td>
</tr>
<tr>
<td>0.2</td>
<td>3.65</td>
<td>2.60</td>
<td>-2.89</td>
<td>-0.32</td>
<td>-3.09</td>
</tr>
<tr>
<td>0.3</td>
<td>-2.19</td>
<td>3.30</td>
<td>2.73</td>
<td>3.30</td>
<td>2.73</td>
</tr>
<tr>
<td>0.4</td>
<td>13.36</td>
<td>16.05</td>
<td>22.81</td>
<td>32.62</td>
<td>32.62</td>
</tr>
<tr>
<td>1,1</td>
<td>6.40</td>
<td>6.00</td>
<td>14.74</td>
<td>1.59</td>
<td>1.59</td>
</tr>
<tr>
<td>1,2</td>
<td>1.3</td>
<td>-4.48</td>
<td>-14.47</td>
<td>-35.55</td>
<td>-35.55</td>
</tr>
<tr>
<td>1,3</td>
<td>2.2</td>
<td>2.23</td>
<td>29.28</td>
<td>19.33</td>
<td>19.33</td>
</tr>
<tr>
<td>1,4</td>
<td>2.77</td>
<td>2.79</td>
<td>11.56</td>
<td>29.22</td>
<td>29.22</td>
</tr>
<tr>
<td>2,2</td>
<td>2.77</td>
<td>2.23</td>
<td>29.28</td>
<td>1.93</td>
<td>1.93</td>
</tr>
<tr>
<td>2,4</td>
<td>-2.27</td>
<td>11.56</td>
<td>29.22</td>
<td>46.68</td>
<td>46.68</td>
</tr>
<tr>
<td>3,3</td>
<td>1.16</td>
<td>11.30</td>
<td>20.68</td>
<td>47.40</td>
<td>47.40</td>
</tr>
<tr>
<td>3,4</td>
<td>1.16</td>
<td>11.30</td>
<td>20.68</td>
<td>47.40</td>
<td>47.40</td>
</tr>
<tr>
<td>4,4</td>
<td>1.16</td>
<td>11.30</td>
<td>20.68</td>
<td>47.40</td>
<td>47.40</td>
</tr>
</tbody>
</table>

Figure 1. Additive genetic (co)variances (in kg²) for orders of polynomial fit (k) of 2 to 5, in which 1, 2, 3, 4, 5 denoting 1, 35, 70, 120 and 180 days of age respectively.

**Coefficients of CFs**

Estimates of the coefficients of the genetic CF, A and the permanent environmental CF, R are listed in Table 3. Thus, under the suitable order of fit, estimated CFs for SD-II Line, with ti denoting the ith standardized age, can be
written as

\[
\begin{bmatrix}
10.37 & 13.68 & -2.89 & -3.09 & 3.30 \\
13.68 & 32.62 & 14.74 & -14.47 & -10.10 \\
-2.89 & 14.74 & 29.28 & -11.13 & -22.79 \\
-3.09 & -14.47 & -11.13 & 11.56 & 11.30 \\
3.30 & -10.10 & -22.79 & 11.30 & 20.68 \\
\end{bmatrix}
\begin{bmatrix}
k = 1 \\
k = 2 \\
k = 3 \\
k = 4 \\
k = 5 \\
\end{bmatrix}
\]

\[
A(t_i, t_j) = \begin{bmatrix}
10.37 & 13.68 & -2.89 & -3.09 & 3.30 \\
13.68 & 32.62 & 14.74 & -14.47 & -10.10 \\
-2.89 & 14.74 & 29.28 & -11.13 & -22.79 \\
-3.09 & -14.47 & -11.13 & 11.56 & 11.30 \\
3.30 & -10.10 & -22.79 & 11.30 & 20.68 \\
\end{bmatrix} 
\begin{bmatrix}
k = 1 \\
k = 2 \\
k = 3 \\
k = 4 \\
k = 5 \\
\end{bmatrix}
\]

\[
R(t_i, t_j) = \begin{bmatrix}
19.70 & 63.42 & 39.47 & -51.99 & -47.77 \\
2.73 & -47.77 & -100.63 & 47.40 & 97.55 \\
\end{bmatrix} 
\begin{bmatrix}
k = 1 \\
k = 2 \\
k = 3 \\
k = 4 \\
k = 5 \\
\end{bmatrix}
\]
genetic and permanent environmental effects. For data used in this paper, a further study is needed for estimation of CFs modeling maternal effects, besides direct genetic effect.

RRM and CFs model observations (namely means) and (co)variances over time respectively. Their combination can effectively explain the genetic and environmental variations in continuous scale. For this reason, CFs have been used in beef cattle (eg. Meyer, 1999; 2001), dairy cattle (eg. Van der Werf et al., 1998), Zebu cattle (eg. Albuquerque and Meyer, 2001), swine (Liu Wenzhong et al., 2001), and goats (Liu Wenzhong, 2001). Because of the availability of ready-made software, extensive application of CFs to longitudinal data in other species is expected.

REFERENCES


CATTLE REMEMBER LOCATIONS OF PREFERRED FOOD
CATTLE REMEMBER LOCATIONS OF PREFERRED FOOD