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 t_{ij}^{*} denotes the j th age for animal i standardized to the range of -1 to 1, and let $\phi_m(t_{ij}^{*})$ be the m th Legendre

* Corresponding Author: Wenzhong Liu. Tel: +86-354-6289036, Fax: +86-354-6289295, E-mail: tglwzyc@public.yz.sx.cn
Received August 17, 2001; Accepted January 11, 2002

Table 1. Summary of the data set used

Terms	Body weight at				
	Birth	35 d	70 d	120 d	180 d
Number of animals	686	686	686	686	686
Number of base animals	108	108	108	108	108
Number of records per animal	5	5	5	5	5
Mean (kg)	1.14	6.32	17.72	39.59	70.57
Standard deviation (kg)	0.22	1.42	3.62	8.63	10.54
Coefficient of variation (%)	19.54	22.41	20.46	21.79	14.89

polynomial evaluated for t_{ij}^*. The RRM can then be written as (Meyer, 1998a)

$$y_{ij} = F + \sum_{m=0}^{k_A-1} \alpha_{im}\phi_m(t_{ij}^*) + \sum_{m=0}^{k_R-1} \gamma_{im}\phi_m(t_{ij}^*) + \varepsilon_{ij} \quad (1)$$

with y_{ij} the observation for animal i at time j, F some fixed effects, α_{im} and γ_{im} 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 ε_{ij} the measurement error (or temporary environmental effect) pertaining to y_{ij}.

Covariance structure

The covariance between two records for the same animal is then

$$Cov(y_{ij}, y_{ij'}) = \sum_{m=0}^{k_A-1}\sum_{l=0}^{k_A-1} \phi_m(t_{ij}^*)\phi_l(t_{ij'}^*)\, Cov(\alpha_{im}, \alpha_{il}) + \sum_{m=0}^{k_R-1}\sum_{l=0}^{k_R-1} \phi_m(t_{ij}^*)\phi_l(t_{ij'}^*) Cov(\gamma_{im}, \gamma_{il}) + Cov(\varepsilon_{ij}, \varepsilon_{ij'}) \quad (2)$$

Generally, measurement errors are assumed to be identically independent distributed with variance σ_ε^2, so that $Cov(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma_\varepsilon^2$ for $j = j'$ and 0 otherwise.

REML estimation

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

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}^*\alpha + \mathbf{Z}_D^*\gamma + \varepsilon \quad (3)$$

with **y** the vector of N observations measured on N_D animals, **b** the vector of fixed effects, α 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), γ the vector of $k_R \times N_D$ permanent environmental random regression coefficients, ε the vector of N measurement errors, and **X**, $\mathbf{Z}^*$ and $\mathbf{Z}_D^*$ denoting the corresponding 'design' matrices. The superscript '*' marks matrices incorporating orthogonal polynomial coefficients.

Analyses were carried out using program DXMRR (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(logL_1-logL_2) has a x^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) together with estimates of measurement errors ($\sigma_{\varepsilon_i}^2$, i=1, 5), are summarized in table 2. It can be

Table 2. Log restricted likelihoods (log$L$), Akaike's information criterion (AIC) and estimates of measurement errors ($\sigma_{\varepsilon_i}^2$) for different orders of polynomial fit (k)

Order of fit	log$L$	AIC	$\sigma_{\varepsilon 1}^2$	$\sigma_{\varepsilon 2}^2$	$\sigma_{\varepsilon 3}^2$	$\sigma_{\varepsilon 4}^2$	$\sigma_{\varepsilon 5}^2$
1	-12210.3	24434.6	1419.2	1063.0	446.7	20.1	1124.0
2	-6996.8	14015.6	72.2	0.7	10.2	11.4	116.7
3	-4196.6	8427.2	0.0	2.4	5.1	31.6	9.2
4	-3655.2	7360.4	0.0	0.7	4.3	8.6	8.8
5	-3542.6	7155.2	0.0	0.5	1.7	4.5	8.8

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

	Coefficients of A					Coefficients of R				
	k=1	k=2	k=3	k=4	k=5	k=1	k=2	k=3	k=4	k=5
0.0	0.00	8.41	6.35	7.90	10.37	8.91	12.25	4.43	15.35	13.21
0.1		10.43	9.83	12.50	13.68		15.45	3.99	25.34	19.70
0.2			3.65	2.60	-2.89			-0.32	-8.31	-9.30
0.3				-2.19	-3.09				-18.17	-13.13
0.4					3.30					2.73
1.1		13.36	16.05	22.81	32.62		20.42	5.46	48.11	63.42
1.2			6.40	6.00	14.74			1.59	-12.83	39.47
1.3				-4.48	-14.47				-35.55	-51.99
1.4					-10.10					-47.77
2.2			2.77	2.23	29.28			1.93	7.33	109.96
2.3				-1.14	-11.13				11.97	-39.46
2.4					-22.79					-100.63
3.3				1.16	11.56				29.22	46.68
3.4					11.30					47.40
4.4					20.68					97.55

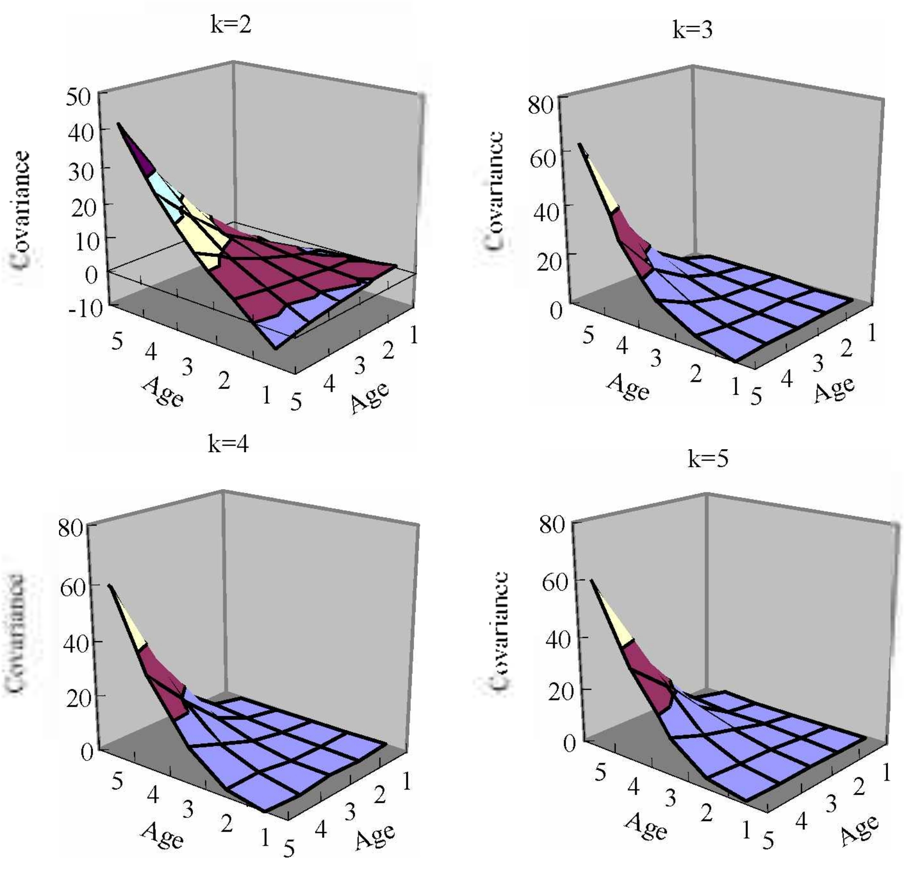

Figure 1. Additive genetic (co)variances (in kg^2) for orders of polynomial fit (k) of 2 to 5, in which 1, 2, 3, 4, 5 denoting 1, 35, 70, 120 and 180 days of age respectively.

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.

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 t_i denoting the i th standardized age, can be

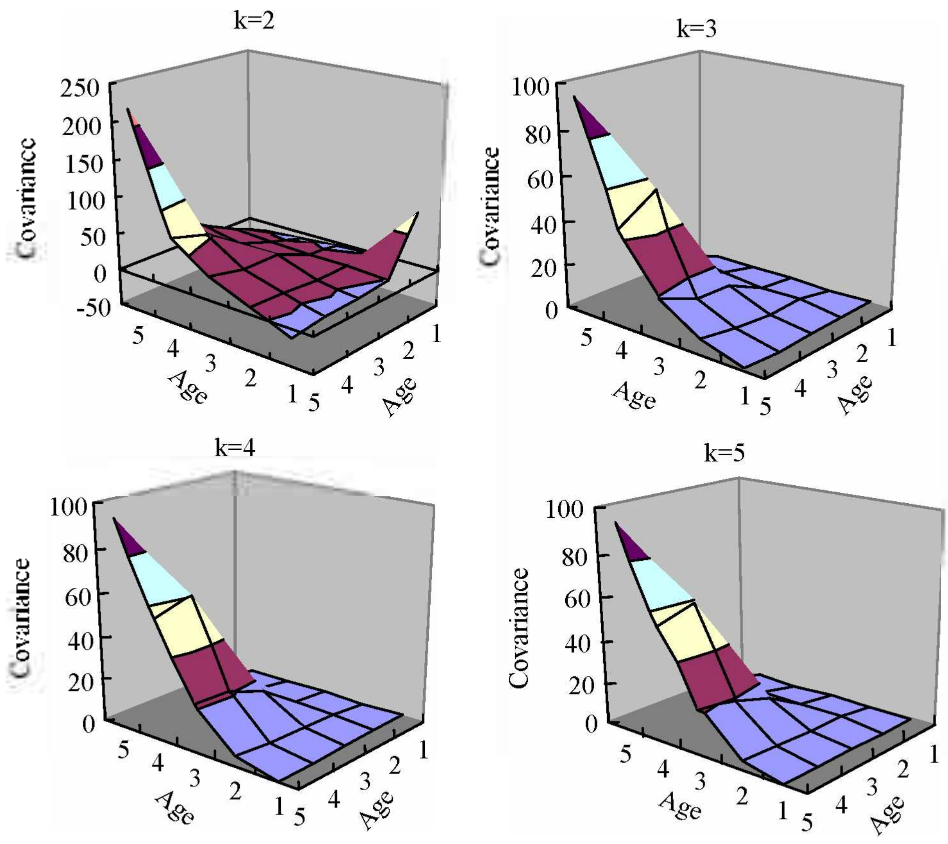

Figure 2. Phenotypic (co)variances (in kg^2) for orders of polynomial fit (k) of 2 to 5, in which 1, 2, 3, 4, 5 denoting 1, 35, 70, 120 and 180 days of age respectively.

written as

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

$$R(t_i,t_j)=\begin{bmatrix}1 & t_i & t_i^2 & t_i^3 & t_i^4\end{bmatrix}\begin{bmatrix}13.21 & 19.70 & -9.30 & -13.13 & 2.73\\ 19.70 & 63.42 & 39.47 & -51.99 & -47.77\\ -9.30 & 39.47 & 109.96 & -39.46 & -100.63\\ -13.13 & -51.99 & -39.46 & -46.68 & 47.40\\ 2.73 & -47.77 & -100.63 & 47.40 & 97.55\end{bmatrix}\begin{bmatrix}1\\ t_j\\ t_j^2\\ t_j^3\\ t_j^4\end{bmatrix}$$

Genetic and phenotypic (co)variances

The additive genetic and phenotypic (co)variances from the estimated CFs for different orders of polynomial fit are shown in figures 1 and 2. With the increase of age, both genetic and phenotypic covariance increases. For k=1 (not shown), all (co)variances are equal. Graphically, that is a plane parallel to the base. For k=2, (co)variances are linear functions of the ages, resulting in a tilted plane. Including quadratic ages for k=3 then gives a semi-parabolic surface. Considering cubic (k=4) or quartic (k=5) terms does not change its shape dramatically. For all orders of fit, the change trend of genetic and phenotypic covariance was very similar from k=3 onwards.

DISCUSSION

Often, a reduced order of fit involves fewer parameters and smoothes out differences in estimates of covariance (Kirkpatrick et al., 1990; Meyer and Hill, 1997), but from this study, a reduced order of fit did not describe the data adequately. This may be due to the selection of additional random effect. Using the same data set, Liu Wenzhong et al. (2001) studied influences of different additional random effects on estimating CFs. The results showed that a reduced order of fit (k=3) was feasible using animal rather than litter effect as additional random effect. All those results indicate that although litter effect is an important permanent environmental effect at early stage of growth, it cannot reflect all the permanent environmental effects in the entire growth period.

Because growth, especially early growth traits, were much influenced by maternal effects, it may be necessary to include the maternal effects in the model. Meyer (1999, 2000), Albuquerque and Meyer (2001) and Liu Wenzhong et al. (2001) have extended the model to include maternal 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

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: 2nd Int. Symp. Inf. Theory (Ed. B. Petrov and F. Csaki). Akad. Kiado, Budapest. pp. 267-281.

Albuquerque, L. G. and K. Meyer. 2001. Estimates of direct and maternal genetic effects for weights from birth to 600 days of age in Nelore cattle. J. Anim. Breed. Genet. 118:83-92.

Johnson, D. L. and R. Thompson. 1995. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. J. Dairy Sci. 78:449-456.

Kirkpatrick, M. and N. Heckman. 1989. A quantitative genetic model for growth, shape, and other infinite dimensional characters. J. Math. Biol. 27:429-450.

Kirkpatrick, M., D. Lofsvold and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979-993.

Kirkpatrick, M., W. G. Hill and R. Thompson. 1994. Estimating the covariance structure of traits during growth and aging, illustrated with lactation in dairy cattle. Genet. Res. 64:57-69.

Liu Wenzhong. 2001. ph D. thesis. China Agricultural University, Beijing, China.

Liu Wenzhong, Zhang Yuan and Zhou Zhongxiao. 2001. Influences of different additional random effects on estimating covariance functions. Hereditas (Beijing) 23:317-320.

Meyer, K. and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. Livst. Prod. Sci. 47:185-200.

Meyer, K. 1998a. Estimating covariance functions for longitudinal data using a random regression model. Genet. Sel. Evol. 30:221-240.

Meyer, K. 1998b. Modeling 'repeated' records: covariance functions and random regression models to analyze animal breeding data. Proc. 6th World Congr. Genet. Appl. Livest. Prod. 25:517-520.

Meyer, K. 1998c. "DXMRR"-A program to estimate covariance functions for longitudinal data by restricted maximum likelihood. Proc. 6th World Congr. Genet. Appl. Livest. Prod. 27:465-466.

Meyer, K. 1999. Estimates of direct and maternal genetic covariance functions for early growth of Australian beef cattle. 50th Ann. Meet. Eur. Assoc. Anim. Prod. Zurich. p. 296.

Meyer, K. 2001. Estimates of direct and maternal covariance functions for growth of Australian beef calves from birth to weaning. Genet. Sel. Evol. 33:487-514.

Van der Werf, J., M. Goddard and K. Meyer. 1998. The use of covariance functions and random regression for genetic evaluation of milk production based on test day records. J. Dairy Sci. 81:3300-3308.

Zhou Zhongxiao, Liu Wenzhong, Zhang Shaozeng, Zhang Yiming and Jia Yuzhi. 1999. Research on Shanxi lean-meat pig (Line SD-II) breeding. Proc. 10th Nat. Symp. Anim. Breed. Genet. pp. 401-404.