Note

Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML

Troy M. Fischer[a,c][*], Arthur R. Gilmour[b,c],
Julius H.J. van der Werf[a,c]

[a] School of Rural Science and Agriculture, University of New England, Armidale,
NSW, 2351, Australia
[b] NSW Agriculture, Orange Agricultural Institute, Orange, NSW, 2800, Australia
[c] Australian Sheep Industry CRC

Abstract – Approximate standard errors (ASE) of variance components for random regression coefficients are calculated from the average information matrix obtained in a residual maximum likelihood procedure. Linear combinations of those coefficients define variance components for the additive genetic variance at given points of the trajectory. Therefore, ASE of these components and heritabilities derived from them can be calculated. In our example, the ASE were larger near the ends of the trajectory.

random regression / heritability / approximate standard error / genetic parameter / residual maximum likelihood

1. INTRODUCTION

Random regression (RR) has been widely used for genetic analysis of longitudinal data from many of the major animal breeding industries world wide and has also been implemented into routine large scale animal breeding applications [4]. Estimates of derived genetic parameters such as heritability at given points along the trajectory are commonly published from such studies and comment is often made about the accuracy and robustness of RR models. However, there have been no attempts to quantify the accuracy of such estimates for different parts of the trajectory from RR analyses using residual

[*] Corresponding author: tfischer@une.edu.au

maximum likelihood (REML) methods. In contrast, Meyer [9] published confidence intervals of genetic parameter estimates derived from Bayesian analyses using Gibbs sampling. With REML estimation by the average information algorithm, approximate variances of variance components are obtained from the inverse of the information matrix. Variance components as well as heritabilities at given trajectory points can be calculated from variances of random regression coefficients and therefore approximate standard errors (ASE) of these derived parameters can also be obtained. The aims of this note are to describe how to calculate ASE for genetic parameter estimates derived from RR models and to apply the method to a field data set.

2. MATERIALS AND METHODS

2.1. Random regression model

Consider a variance-covariance (VCV) matrix $\mathbf{G}_0$ of rank t for repeated measurements of weight at t given trajectory points (*e.g.* ages). Under the covariance function (CF) approach defined by Kirkpatrick *et al.* [5], $\mathbf{G}_0$ is modelled with a reduced number of parameters. The genetic CF of order k, where $k \leq t$, can be estimated from $\mathbf{G}_0$ such that:

$$\hat{\mathbf{G}} = \mathbf{\Phi} \, \mathbf{K} \, \mathbf{\Phi}' \tag{1}$$

where $\hat{\mathbf{G}}$ is an approximation of $\mathbf{G}_0$. Meyer [8] showed that $\mathbf{K}$ can be estimated directly from data using RR. The matrix $\mathbf{K}$ of order k contains the variance components for the RR coefficients in the model. The matrix $\mathbf{\Phi}$ of order $t \times k$ contains orthogonal polynomial coefficients evaluated at t standardised trajectory points (ages) with elements $\phi_{ij} = \phi_j(x_i)$, being the jth polynomial coefficient for the ith point x_i [6]. The covariance structure for the environmental effects is fitted as an unstructured $t \times t$ covariance matrix. This yields the model:

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{b} + \mathbf{Z}_i \alpha_i + \mathbf{e}_i \tag{2}$$

where $\mathbf{y}_i$ is the vector of t_i observations measured on animal i, $\mathbf{b}$ is a vector of fixed effects, α_i a vector of additive genetic RR coefficients and $\mathbf{e}_i$ a vector of residual errors pertaining to $\mathbf{y}_i$. $\mathbf{X}_i$ and $\mathbf{Z}_i$ are design matrices relating $\mathbf{b}$ and α_i to $\mathbf{y}_i$, where $\mathbf{Z}_i$ contains the elements $\mathbf{\Phi}$ pertaining to ages in the data. Extending the model to n individuals, the corresponding variances are defined as $\text{var}(\alpha) = \mathbf{K} \otimes \mathbf{A}$, where $\mathbf{K}$ contains the additive genetic variances and covariances for the RR coefficients, $\mathbf{A}$ is the numerator relationship matrix among

individuals and the symbol $\otimes$ denotes direct product. The solution for $\mathbf{K}$ can be used as in equation (1) to calculate the variances and covariances among defined trajectory points.

2.2. Calculation of standard error of parameters derived from RR coefficients

Consider a genetic variance covariance matrix, $\hat{\mathbf{G}}$, derived from equation (1), $\hat{\mathbf{G}} = \mathbf{\Phi}\,\hat{\mathbf{K}}\,\mathbf{\Phi}'$, where $\mathbf{\Phi}$ has dimension $t \times k$, $\hat{\mathbf{K}}$ has dimension $k \times k$ and $\hat{\mathbf{G}}$ is $t \times t$. We can write the elements of $\hat{\mathbf{G}}$ in vector form, such that the variances and covariances of these parameters can be summarized in a matrix. Hence, equation (1) can also be written as

$$\text{vec}\left(\hat{\mathbf{G}}\right) = \mathbf{\Phi} \otimes \mathbf{\Phi} \quad \text{vec}\left(\hat{\mathbf{K}}\right) \tag{3}$$

where $\mathbf{\Phi} \otimes \mathbf{\Phi}$ has dimension $(t \times t) \times (k \times k)$, vec($\hat{\mathbf{K}}$) is the vector form of $\hat{\mathbf{K}}$ of dimensions $(k \times k) \times 1$ achieved by stacking the columns of $\hat{\mathbf{K}}$ below one another, and similarly vec($\hat{\mathbf{G}}$) is the vector form of $\hat{\mathbf{G}}$ of dimensions $(t \times t) \times 1$. It can be checked for a small example that equations (1) and (3) are equivalent, written in matrix and vector form respectively. The variance of estimates in $\hat{\mathbf{G}}$ can be calculated in a similar manner whereby

$$\text{var}\left(\text{vec}\left(\hat{\mathbf{G}}\right)\right) = (\mathbf{\Phi} \otimes \mathbf{\Phi})\ \text{var}\left(\text{vec}\left(\hat{\mathbf{K}}\right)\right)\ (\mathbf{\Phi} \otimes \mathbf{\Phi})' \tag{4}$$

where var(vec($\hat{\mathbf{K}}$)) has dimensions $(k{\times}k){\times}(k{\times}k)$ and var(vec($\hat{\mathbf{G}}$)) is a $(t{\times}t)$ by $(t{\times}t)$ matrix. Var(vec($\hat{\mathbf{K}}$)) can be approximated from the appropriate elements of the inverse of the average information matrix in a REML procedure (*e.g.* as given in the *.vvp file in ASReml) [3]. The same principles apply to other random effects in the RR model, and the covariance between variances of different random effects. Hence, this methodology can be extended to the matrices estimated for other random regression effects and the covariance between random effects. Subsequently these matrices are summed as in equation (5) to build a matrix containing estimates of variance of phenotypic (co)variance components, var(vec($\hat{\mathbf{P}}$)), which also has dimension $(t \times t) \times (t \times t)$.

$$\text{var}\left(\text{vec}\left(\hat{\mathbf{P}}\right)\right) = \text{var}\left(\text{vec}\left(\hat{\mathbf{G}}\right)\right) + \text{var}\left(\text{vec}\left(\hat{\mathbf{E}}\right)\right) + 2\left(\text{cov}\left(\text{vec}\left(\hat{\mathbf{G}}\right), \text{vec}\left(\hat{\mathbf{E}}\right)\right)\right). \tag{5}$$

For functions of variance components (such as heritabilities) a Taylor series expansion can be used to approximate the variance of a variance ratio as detailed by Lynch and Walsh [7]. For the ratio of genetic to phenotypic variance,

we get:

$$\mathrm{var}\left(\hat{g}_{i,i}/\hat{p}_{i,i}\right) = \mathrm{var}\left(\hat{h}_i^2\right)$$

$$\approx \left(\hat{p}_{i,i}^2 \mathrm{var}\left(\hat{g}_{i,i}\right) + \hat{g}_{i,i}^2 \mathrm{var}\left(\hat{p}_{i,i}\right) - 2\hat{g}_{i,i}\hat{p}_{i,i}\mathrm{cov}\left(\hat{g}_{i,i}, \hat{p}_{i,i}\right)\right)/\hat{p}_{i,i}^4 \quad (6)$$

where $\hat{g}_{i,i}$ and $\hat{p}_{i,i}$ are elements of $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}$, $\mathrm{var}(\hat{g}_{i,i})$, $\mathrm{var}(\hat{p}_{i,i})$ and $\mathrm{cov}(\hat{g}_{i,i}, \hat{p}_{i,i})$ represent variance and covariance of genetic and phenotypic variance at time i. The ASE for the heritability estimate at time i (for univariate and RR estimates) is obtained by taking the square root of equation (6).

2.3. Example of application of method to RR coefficients estimated from field data
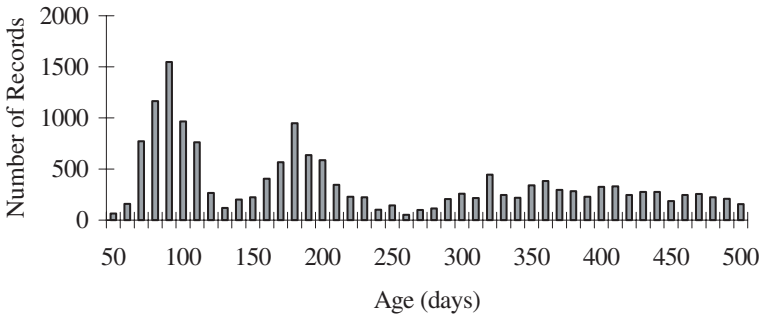
**Figure 1.** Number of records at different ages.

A VCV matrix for additive genetic and phenotypic effects for weight over a 450-day trajectory was derived based on the analysis performed by Fischer *et al.* [2]. Data for this analysis originated from the LAMBPLAN database and consisted of 16 826 weight records on 5 420 Poll Dorset sheep. The number of records at different ages is represented in Figure 1.

Fischer *et al.* [2] used RR to estimate CF coefficients for direct and maternal genetic and environmental effects. The model also included heterogeneous residual variance across ages of measurement. ASReml [3] was used for this analysis. Based on a third order CF for additive genetic effects, a VCV matrix ($\hat{\mathbf{G}}$) was constructed for weights at 10 equidistant ages (*i.e.* defining $\mathbf{\Phi}$). Similarly, VCV matrices were derived for the other random effects. Furthermore, adding the resultant variance matrices together resulted in a phenotypic VCV matrix ($\hat{\mathbf{P}}$) with (co)variance components for weights at the 10 equidistant ages.

We then obtained the variance of vec($\hat{\mathbf{G}}$) as in equation (4) and similarly for the two types of maternal effects, which in this case were all matrices of dimensions 100×100. Following equations (4), (5) and (6) we obtained the ASE of the heritability estimate for each age. Results from this example are shown in Figure 2.

In addition, a series of piecewise estimates at specific ages were obtained using the equivalent univariate model (direct and maternal genetic effects only, no permanent environmental effects fitted). Estimates were taken from day 50 onwards using a 50-day age window for each univariate estimate up to 500 days and these are also shown in Figure 2.

3. RESULTS AND DISCUSSION
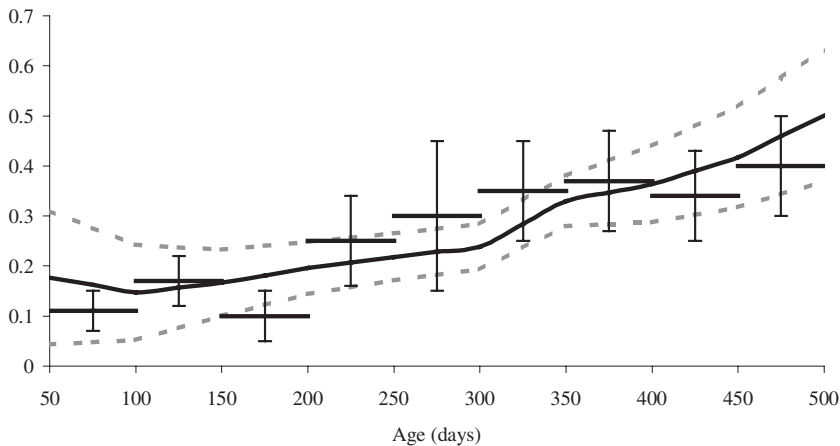
Figure 2. Heritability estimates of weight over time ± standard errors from random regression (continuous line) and univariate (discrete lines) analysis.

As can be seen in Figure 2, the ASE for heritability estimates from RR are lower (0.05−0.07) in the middle of the trajectory, and higher (0.08−0.13) at the ends of the trajectory. The pattern at the end of the trajectory is largely due the nature of polynomials, which have no constraint at the ends. This result is consistent with that of Fischer and van der Werf [1] who demonstrated problems associated with polynomials, in particular high order Legendre polynomials.

The ASE for heritabilities at given ages obtained using RR were smaller (0.04−0.13) throughout much of the trajectory than those obtained from piecewise univariate analyses of portions of the same data within 50 day age

windows throughout the 450 day trajectory (0.04−0.15). The large standard errors for specific univariate estimates were largely a function of fewer data within that age window, in particular beyond 200 days. Furthermore, the larger ASE values at the ends of the trajectory are largely due to polynomial instability, which is further exacerbated by fewer data at these ages. Instability has been shown in heritability estimates at the ends of the trajectory when using polynomials in RR, even in cases where there was more data at the ends [1].

Furthermore, the ASE obtained using the RR model were comparable or lower than those obtained from piecewise univariate analysis except at the ends of the trajectory, showing where RR estimates are more accurate. It is particularly useful to obtain standard errors of heritability at the ends of the trajectory as concerns are often raised about the robustness of such estimates due to the polynomial nature of the covariance function and justification of this concern was demonstrated in this study. Moreover, the same method can be applied to other random effects to get standard errors for their respective proportion of phenotypic variance (*e.g.* maternal heritability).

4. CONCLUSION

This study demonstrated a method for computing ASE for genetic parameter estimates derived using RR models applied to a field data set. The method produced plausible standard error values for estimates of heritability and this provides insight into the discussion of robustness and accuracy of RR estimates of heritability at specific age points.

ACKNOWLEDGEMENTS

REFERENCES

[1] Fischer T.M., van der Werf J.H.J., Effect of data structure on the estimation of genetic parameters using random regression, in: Proc. 7th World Cong. Genet. Appl. Livest. Prod., Montpellier, France, 19–23 August 2002, CD-ROM communication No. 17-08.
[2] Fischer T.M., van der Werf J.H.J., Banks R.G., Ball A.J., Description of lamb growth using random regression on field data, Livest. Prod. Sci. (2004) In press.

[3] Gilmour A.R., Cullis B.R., Welham S.J., Thompson R., ASREML reference manual, NSW Agriculture, Orange, 2002.

[4] Jamrozik J., Schaeffer L.R., Dekkers J.C.M., Genetic evaluation of dairy cattle using test day yields and random regression model, J. Dairy Sci. 80 (1997) 1217–1226.

[5] Kirkpatrick M., Lofsfold D., Bulmer M., Analysis of the inheritance, selection and evolution of growth trajectories, Genetics 124 (1990) 979–993.

[6] Kirkpatrick M., Lofsvold D., Measuring selection and constraint in the evolution of growth, Evolution 46 (1992) 954–971.

[7] Lynch M., Walsh B., Genetics and analysis of quantitative traits, Sinauer Associates, Sunderland, MA, USA, 1998.

[8] Meyer K., Estimating covariance functions for longitudinal data using a random regression model, Genet. Sel. Evol. 30 (1998) 221–240.

[9] Meyer K., Estimates of covariance functions for growth of Australian beef cattle from a large set of field data, in: Proc. 7th World Cong. Genet. Appl. Livest. Prod., Montpellier, France, 19–23 August 2002, CD-ROM Communication No. 11-01.