

Investigations into the Use of a Random Regression Reduced Animal Model for Variance Component Estimation and Breeding Value Prediction

R. Mrode¹, I.M.S. White² and S. Brotherstone²

¹Scottish Agricultural College, Bush Estate, Penicuik, EH26 0PH, UK

²Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, West Mains Road, Edinburgh, EH9 3JT, UK

Introduction

The reduced animal model (RAM) was developed by Quaas and Pollak (1980) and allows equations to be set up only for parents in the MME and solutions for non parents obtained by back solving. The advance in computing power has made the model less attractive for the prediction of breeding values. However considering the current trend towards the use of random regression (RR) models for genetic evaluation with several equations fitted per animal, the benefits of the RAM might be more pronounced. Moreover, the estimation of variance components is generally more computationally demanding especially with the RR model and the tendency has been to fit a sire model or use Gibbs sampling. White *et al* (2006) extended the RAM for the estimation of variance components in pig data, achieving a 30% reduction in computing time and a 70% reduction in memory relative to an animal model. However, this has not been extended to RR models. The purpose of this paper is to briefly outline a random regression reduced animal model (RRRAM) for breeding value prediction and its application to the estimation of variance components.

Materials and Method

Breeding value prediction

Let the RR model for a parent be of the form:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{\Phi u} + \mathbf{\Phi p} + \mathbf{e} \quad [1]$$

where $\mathbf{y}$ is the vector of test day records or body weights at different ages, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{u}$ and $\mathbf{p}$ are vectors of RR coefficients for random animal and permanent environmental (pe) effects respectively. The matrix $\mathbf{\Phi}$ contains covariates such as Legendre polynomials for animal and pe effects and $\mathbf{X}$ is the incidence matrix for fixed effects. For the i^{th} parent, let $\mathbf{r}$ be a row vector with 1 as the only non-zero element in the i^{th} position. Then the second term can be written as $\mathbf{Zu}_{\text{par}}$ where $\mathbf{u}_{\text{par}}$ is the vector of RR coefficients for parents and $\mathbf{Z} = \mathbf{r} \otimes \mathbf{\Phi}$, the Kronecker product of $\mathbf{r}$ and $\mathbf{\Phi}$. Thus for a parent, we have:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu}_{\text{par}} + \mathbf{\Phi p} + \mathbf{e} \quad [2]$$

It is assumed that the variances of $\mathbf{u}$ and $\mathbf{p}$ are $\mathbf{G}$ and $\mathbf{P}$ respectively.

For non-parents, equation [1] can be expressed as:

$$\mathbf{y} = \mathbf{Xb} + 0.5(\mathbf{\Phi u}_s + \mathbf{\Phi u}_d) + \mathbf{\Phi m} + \mathbf{\Phi p} + \mathbf{e}$$

where s and d are the sire and dam of the animal respectively and $\mathbf{m}$ is a vector of coefficients for Mendelian sampling (MS) effects. Again the second term is $\mathbf{Zu}_{\text{par}}$, with $\mathbf{Z} = \mathbf{r} \otimes \mathbf{\Phi}$, where now $\mathbf{r}$ is a row vector of zeros with 0.5 in positions corresponding to the sire and dam of the animal. Thus for a non-parent we have:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu}_{\text{par}} + \mathbf{\Phi p}^* + \mathbf{e} \quad [3]$$

where $\Phi\mathbf{p}^* = \Phi\mathbf{m} + \Phi\mathbf{p}$ and $\mathbf{p}^*$ has variance equal to $k(1-F)\mathbf{G} + \mathbf{P}$ with k equals $\frac{1}{2}$ or $\frac{3}{4}$ or 1 if both, one or no parents are known respectively and F being the average inbreeding for both parents. The above implies that the same order of polynomial is fitted for both animal and pe effects.

The application of RRRAM therefore involves applying equation [2] for parents and [3] for non-parents.

$$\begin{pmatrix} \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + k^{-1}\mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\Phi \\ \Phi'\mathbf{R}^{-1}\mathbf{Z} & \Phi'\mathbf{R}^{-1}\Phi + \mathbf{P}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{m}}_i \\ \hat{\mathbf{p}}_i \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}_c \\ \Phi'\mathbf{R}^{-1}\mathbf{y}_c \end{pmatrix}$$

where $\mathbf{R}^{-1}$ is the inverse of the residual variance, $\mathbf{y}_c = (\mathbf{y}_c = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - 0.5(\hat{\mathbf{u}}_s + \hat{\mathbf{u}}_d))$ and k is as defined in the explanation for equation [3].

$$\begin{pmatrix} \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + k^{-1}\mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\Phi \\ \Phi'\mathbf{R}^{-1}\mathbf{Z} & \Phi'\mathbf{R}^{-1}\Phi + \mathbf{P}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_i \\ \hat{\mathbf{p}}_i \end{pmatrix} = \begin{pmatrix} \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + k\mathbf{G}(\hat{\mathbf{u}}_s + \hat{\mathbf{u}}_d) \\ \Phi'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

where now $\mathbf{y}_c = (\mathbf{y}_c = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}})$.

Thus the order of the equations for each animal equals the sum of the orders of polynomials fitted for animal and pe effects.

Application to a small data set

Table 1 gives the test day fat yields of five cows in a herd with details of herd-test day (htd), days in milk and pedigree (DATA1). The data was analysed with a RR model fitting htd effects, a fixed lactation curve using Legendre polynomials of order 4, random animal and permanent environmental (pe) effects with Legendre polynomials of order 2. Genetic parameters used in the analyses were given by Mrode (2005) in example 7.2.1.

The data was re-analysed using RRRAM.

Back solving for solutions of non-parents:

From [3] the RR coefficients for pe for non-parents include estimates of Mendelian sampling. Thus the RR coefficients for animal and pe effects must be solved for. This could be achieved in two ways. Firstly solutions for MS effects ($\hat{\mathbf{m}}$) and pe can be solved for and then RR coefficients for an animal obtained as the sum of $\hat{\mathbf{m}}$ plus parent average. From the mixed model equations for [1], the equations for $\hat{\mathbf{m}}_i$ and $\hat{\mathbf{p}}_i$ are:

Alternatively, $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{p}}_i$ for an animal i could be obtained directly using the following equations:

Variance component estimation

The application of RRRAM for the estimation of variance components also involves the use of equations [2] where $\mathbf{Z} = \mathbf{r} \otimes \Phi$ and $\mathbf{r}$ takes one of two forms depending on whether the animal is a parent or non-parent. However, an additional term is fitted for MS effects for non-parents. The variance for this extra term was constrained to be $k(1-F)\mathbf{G}$. Equation [2] suggests a possible way to set data for the RRRAM. For parents, identify the animal but set sire and dam as 'missing'; for non parents, set the animal identity as missing but give the sire and dam. The RRRAM was implemented using ASReml (Gilmour *et al.*, 2002). The RRRAM can be analysed by any software for variance components that allows for missing values in design matrices. When the number of parents with records is a small proportion of the total number of

animals, a simpler analysis can be implemented ignoring parents with records. This can be referred as the approximate RRRAM (White *et al.*, 2006) and it often captures most of the information in the data set.

Data analysis

Variance components were estimated using 40371 body weight records on 2264 Suffolk lambs collected from 1985 to 1994 (DATA2). Lambs were weighed between 7 and 24 times over a period of 150 days of age. The number of sires and dams with progeny in the data set were 138 and 716 respectively. The fixed effects consisted of year-sex and fixed growth curves nested within year-sex. The fixed curves were fitted using polynomials of order 2 for age. Random animal and permanent environmental effects were fitted using polynomials of order 2. Six classes of residual variances were fitted for weights recorded at days 2 to 25, 26 to 50, 51 to 75, 76 to 100, 101 to 125 and 126 to 159. The data set were analysed using RR and RRRAM. The pedigree for RR model and RRRAM consisted of 2807 and 1123 animals respectively.

Results and Discussion

Breed value prediction

The solutions for animal and pe effects using DATA1 for breeding value prediction with both models are shown in Table 2. As expected RRRAM gave the same solution as the RR model but there was an 11% reduction in the number of equations to be solved with the RRRAM. Note that the solutions for non-parents for the permanent environment in Table 2 are for p^* ($p_e + \text{estimates of MS}$). The appropriate solutions for p_e and animal effects for non parents were obtained by back solving (not shown) and they were the same as obtained by the RR model. The benefits of RRAM in terms of a reduction in the number of equations to be solved will be maximized in multivariate situations and in species such as pigs and

chickens where smaller proportions of progeny become parents.

Variance component

The parameters estimated from DATA2 using both model are given in Table 3. The estimates are the same for both models. The small size of the data set has not permitted reliable time differences to be recorded as both analyses run in about 40 seconds. In addition, the current methodology of fitting Mendelian sampling term for non-parents in the RRRAM has not resulted in reduction in number of equations. A possible strategy yet to be investigated is defining in ASReml, the p_e variance for non-parents, $k(1-F)G + P$ in terms of two parameters, G and P , to be estimated separately. The first is constrained to be equal to G and the second to the p_e variance of parents.

Conclusion

We demonstrated that it is feasible to analyse longitudinal data with a RRRAM both for breeding value prediction and variance components. This could result in substantial reduction in number of equations to be solved in multivariate situations and in species where small proportions of progeny becoming parents. The use of RRRAM for variance components has great potential and on going work will investigate better options of fitting the model.

References

- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. & Thompson, R. 2002. *ASReml User Guide* Release 1.0, VSN international Ltd, Hemel Hempstead, UK
- Mrode R.A. 2005. *Linear Models for the prediction of animal breeding values*, 2nd Edition, CABI, Wallingford, UK.
- Quaas, R.L. & Pollak, E.J. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* 51, 1277-1287

White, I, M.S., Roehe, R., Knap, P.W. & Brotherstone, S. 2006. Variance components for survival of piglets at

farrowing using a reduced animal model. *Gen. Sel. Evol.* In press.

Table 1. Test day fat yields (TDY) for some cows in a herd.

ANIMALS										PEDIGREE			
	4		5		6		7		8				
DIM	HTD	TDY	HTD	TDY	HTD	TDY	HTD	TDY	HTD	TDY	Cow	Sire	dam
4	1	17.0	1	23.0	6	10.4	4	22.8	1	22.2	4	1	2
38	2	18.6	2	21.0	7	12.3	5	22.4	2	20.0	5	3	2
72	3	24.0	3	18.0	8	13.2	6	21.4	3	21.0	6	1	5
106	4	20.0	4	17.0	9	11.6	7	18.8	4	23.0	7	3	6
140	5	20.0	5	16.2	10	8.4	8	18.3	5	16.8	8	1	4
174	6	15.6	6	14.0			9	16.2	6	11.0			
208	7	16.0	7	14.2			10	15.0	7	13.0			
242	8	13.0	8	13.4					8	17.0			
276	9	8.2	9	11.8					9	13.0			
310	10	8.0	10	11.4					10	12.6			

DIM = days in milk, HTD = herd-test-day

Table 2. Regressions coefficients for animal and permanent environmental effects.

Regression coefficients						
Animal effects						
	RR model			RRRAM		
1	-0.0193	0.0655	-0.0409	-0.0193	0.0655	-0.0409
2	-0.3181	-0.0394	0.0944	-0.3181	-0.0394	0.0944
3	0.3374	-0.0261	-0.0535	0.3374	-0.0261	-0.0535
4	-0.0926	0.0125	-0.1111	-0.0926	0.0125	-0.1111
5	-0.3845	-0.0716	0.2527	-0.3845	-0.0716	0.2527
6	-0.1480	0.0706	0.0507	-0.1480	0.0706	0.0507
7*	0.8263	0.0350	-0.2871			
8*	-0.2053	0.0315	0.0761			
Permanent environmental (pe) effects						
	RR model			RRRAM		
4	-0.5040	-0.3697	-1.5643	-0.5040	-0.3697	-1.5643
5	-1.0864	0.1529	1.0735	-1.0864	0.1529	1.0735
6	-2.0489	0.9857	-0.0310	-2.0489	0.9857	-0.0310
7*	3.6055	-1.0385	-0.4547	4.3371	-1.0258	-0.7405
8*	0.0337	0.2697	0.9765	-0.1156	0.2621	1.1287

* Regression coefficients for animal 7 and 8 under the reduced RR model are obtained by back solving. Regression coefficients for pe shown for 7 and 8 include Mendelian sampling effects.

Table 3. Variance components estimates.

Term	RR model		RRAM	
	Animal effect	Pe effect	Animal effect	Pe effect
a_0a_0	6.494	16.27	6.494	16.27
a_0a_1	4.517	8.082	4.517	8.082
a_0a_2	-1.896	-7.009	-1.896	-7.009
a_1a_1	4.680	6.947	4.680	6.947
a_1a_2	-0.326	-1.763	-0.326	-1.763
a_2a_2	1.645	5.708	1.645	5.708

