The structure of the residual error variance of test day milk yield in random regression models

V.E. Olori[12], W.G. Hill[1], S. Brotherstone[1]

1.   *Institute of Cell Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK.*

2.   *Irish Cattle Breeding Federation, Shinagh House, Bandon, Co Cork, Ireland*

Abstract

The effect of assuming a constant residual (measurement) error structure in the analysis of test day milk records with a random regression model was investigated. Additive genetic and permanent environmental covariance components were each modelled with a 3rd order orthogonal polynomial regression while measurement error was allowed to be either constant in all or some stages of lactation, or to vary from stage to stage. Coefficients of the covariance functions and residual variance components were jointly estimated by restricted maximum likelihood using the DFREML software package.

There were significant increases (P<0.05) in the log-likelihood which suggests improvement in the fit of the derived covariance function as the constraint on residual error variance was removed. There was, however, little difference in estimates of the other variance components. Residual error was high in early lactation and declined rapidly to a stable value in mid/late lactation, so with a constant error assumption, residual variance was underestimated and heritability overestimated in early lactation.

1. Introduction

Estimation of variance components for test day milk records by restricted maximum likelihood (REML) with a random regression model (Meyer and Hill, 1997), entails the derivation of covariance functions for additive genetic and permanent environmental variance components. Maximisation of the likelihood in the iterative process is undertaken with respect to coefficients of the covariance function(s) and estimates of the residual variance components (Meyer 1998a). Constraining residual variance by assuming it is constant for all or some stages of lactation reduces the number of parameters to be estimated and hence the dimension of the likelihood search. Under such constraints, estimates of residual variance in each stage of lactation depend on which lactation stages (weeks) are assumed to have the same measurement error (Olori et al., 1998).

The joint estimation of residual variances and coefficients of the covariance functions for the other components may affect the decomposition of phenotypic variance if the fit of the covariance function depends on the constraint on residual variances, if they are not modelled by a continuous function. This study investigates what effect the constraint placed on the variation of residual variance across lactation stage has on estimating covariance functions for the other components of variance and hence the resulting parameters.

2. Material and Methods

First lactation records of daily milk yield between weeks 4 and 40 of 488 Holstein Friesian cows in one herd were summarised to obtain 37 weekly averages of daily milk yield per cow. Additive genetic and permanent environmental covariance components were

each modelled with covariance functions using a third order (quadratic) orthogonal polynomial regression (on the Legendre scale) as a sub-model in a random regression animal model. Residual variance was assumed to be either constant throughout lactation (ME=1) or to vary only between 4 (ME = 4) or 10 (ME=10) lactation stages obtained by grouping records in different weeks of lactation. Alternatively, residual variance was assumed to be different for yield in every week of lactation (ME =37).

The fixed effect of lactation stage was modelled with the same order polynomial regression as in the random part of the model. Other fixed effects included in the model were gestation stage (9 monthly classes, including 0 for records obtained before the onset of gestation), year/season of production (29 classes comprising two or three calendar month seasons within year), and the age of the cow at calving (11 classes comprising cows aged between 23 and 33 months of age).

Coefficients of the covariance functions and residual variance components were jointly estimated while additive genetic and permanent environmental (co)variances were subsequently derived from the respective covariance functions using the DXMRR option of the DFREML package (Meyer, 1998b). Fit of models was based on a χ^2 test of the log-likelihoods (Kirkpatrick *et al.*, 1990).

3. Results and Discussion

Significant changes (P<0.05) in the maximum log-likelihood were observed as the number of ME classes specified in the model increased. This implies that the fit of the derived covariance function improved as the constraint on the variation of residual variance across lactation stage was reduced from total to none. Removing the constraint also increased the number of estimated parameters from 13 to 49. The number of parameters to be estimated and the changes in log-likelihood relative to the model with ME=1 are given in Table 1.

The increase in fit may be attributed to a more accurate estimation of the residual component of variance as more independent estimates, at different stages of lactation, were allowed with the removal of the constraint on variation especially in the early stages of lactation. When residual variance (V_E) was assumed to vary across lactation stage (ME=37), its estimate was highest ($12kg^2$) for yield in the fourth week of lactation (first week in the trajectory) and declined rapidly to about $2kg^2$ in lactation week 8. It subsequently varied

Table 1. Changes in maximum log-likelihood for models with variable error structures relative to the model with a constant error structure.

| Number of Measurement Error classes	Number of parameters estimated	Change in log-likelihood
1	13	-
4	16	251*
10	33	296*
37	49	440*

* Significant change (P<0.05).

at random between 2 and $4kg^2$ for the remaining stages of lactation. Conversely, when residual variance was assumed constant for all stages of lactation (ME=1), the estimate of V_E was $3.16kg^2$. This value was approximately equivalent to the mean of the estimates for each of the 37 weeks. Figure 1 shows a plot of residual variance estimates at each stage of lactation for all models.

Analysis of variance showed no significant difference (P>0.05) in estimates of additive genetic (V_A) and permanent environmental (V_{PE}) variances between models with different ME classes. Estimates of V_A and V_{PE} from the different models are plotted in Figure 2, which shows little difference in any of the curves as ME was increased from 1 to 37. This suggests that the number of ME classes assumed did not cause the estimates at each stage of lactation to vary significantly.

Figure 3 shows a plot of the heritability estimates at different stages of lactation for each model. Again there was generally no significant difference in the estimates between the models (P<0.05). Heritability estimates in lactation weeks 4 and 5 were high (0.32 and 0.33 respectively) when residual variance was assumed constant at all stages of lactation (ME=1). Corresponding estimates were about 0.24 for both weeks when measurement error was assumed to be the same for yield in lactation weeks 4 and 5 but different from the rest (ME=4). The h^2 estimates were about 0.21 and 0.26 respectively when measurement error was assumed to vary between the two weeks (models with ME=10 and 37).
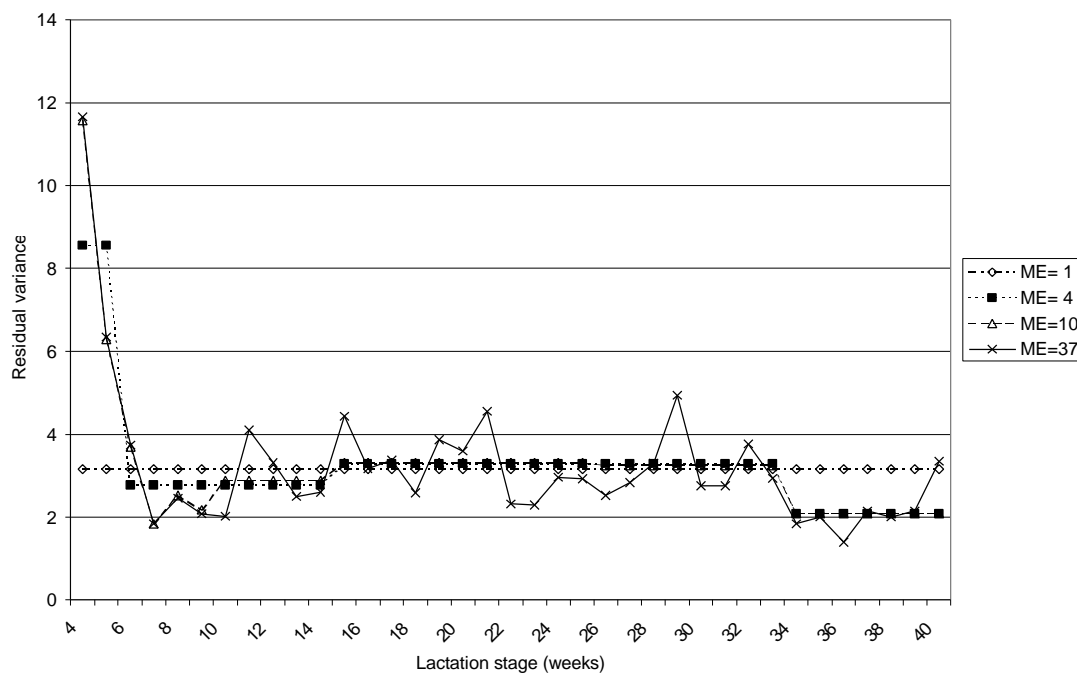
Figure 1. Residual variance of milk yield estimated by random regression models with different measurement error (ME) classes.
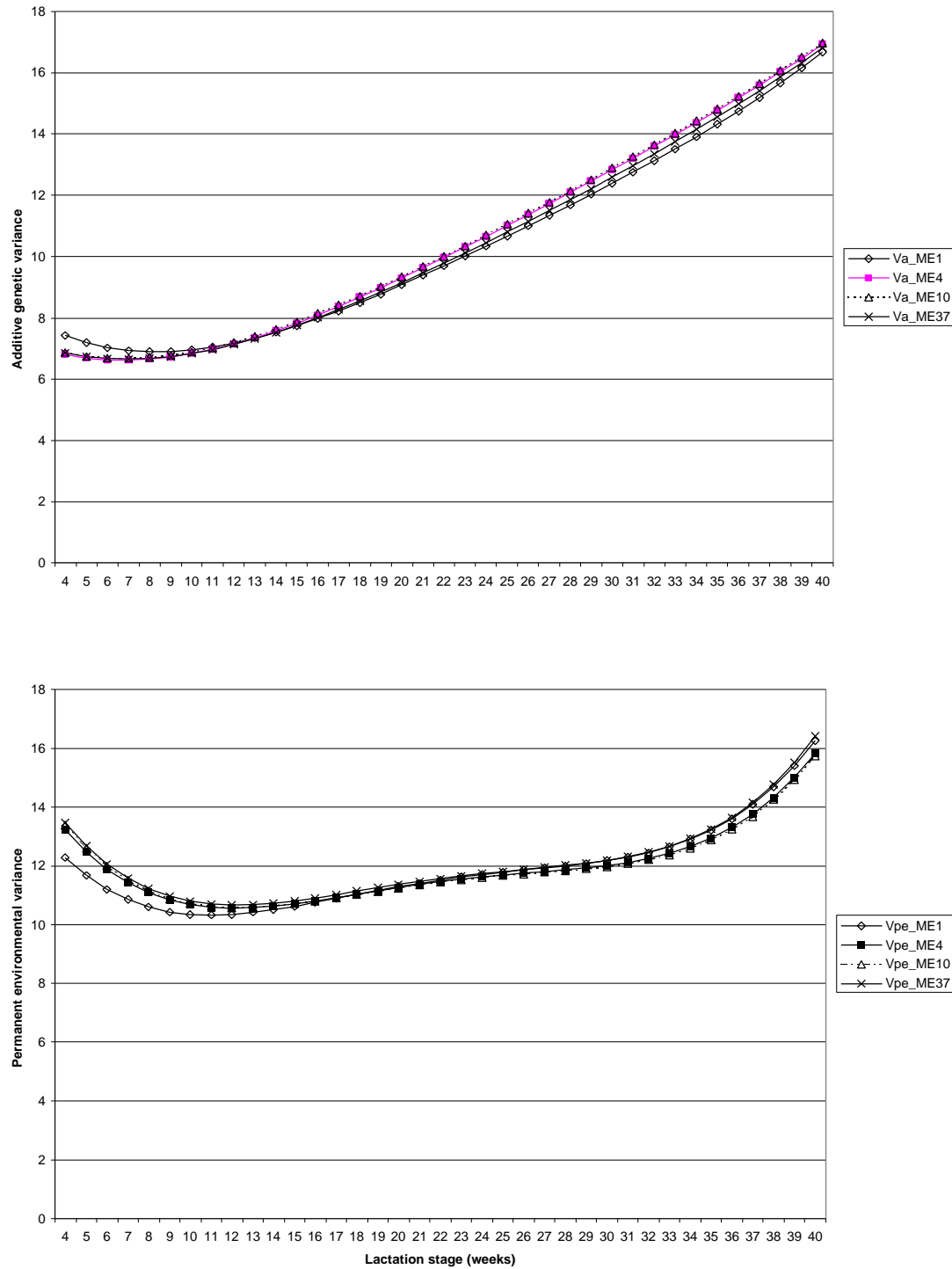
Figure 2. Additive genetic (Top) and permanent environmental variance of milk yield estimated by random regression models with different measurement error (ME) classes.

Estimates of individual residual variance for each stage of lactation (Model ME=37), compared with the others suggest that V_E in early lactation is highly variable while V_E in mid lactation is reasonably constant. A constant error assumption therefore causes residual variance in early lactation to be underestimated. This can have consequences on the estimates of total variance and hence the heritability of yield at this stage of lactation. This may partly explain the inconsistent of h^2 in early and late stages of lactation previously reported (Jamrozik et al., 1997; Kettunen et al., 1998).

The result of this study suggests the need to model residual error of test day records with a function. This will allow residual variance of yield at each stage of lactation to be estimated individually without increasing the number of parameters. Further more the assumption of zero covariance between MEs at different stages of lactation may not be correct. If so, specifying a covariance function to model residual error will facilitate the estimation of residual variance and covariances of milk yield at different stages of lactation.

4. Conclusions

Residual component of variance is more accurately estimated when independent estimates at each stage of lactation are made. For test day milk yield, residual error is high and highly variable in early lactation, low and relatively constant in mid /late lactation. Assuming a constant residual error structure in random regression models biases estimated of residual variance especially in early lactation but has no significant effect on the other components of variance.
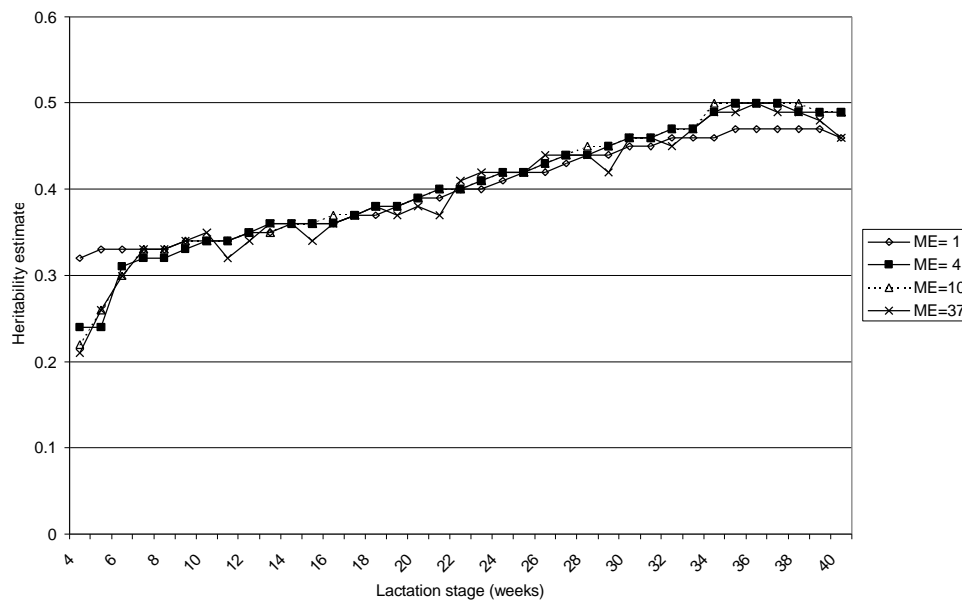
Figure 3. Heritability of milk yield estimated by random regression models with different measurement error (ME) classes.

References

Jamrozik, J. and Schaeffer, L. R. 1997. Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holsteins. *Journal of Dairy Science* **80**:762-770.

Kettunen, A., Mantysaari, E. A., Stranden, I. and Poso, J. 1998. Estimation of genetic parameters for first lactation test day production using random regression models. *Proceedings of the 6th world congress on genetic applied to livestock production, Armidale, Australia* **23**: 307-400.

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

Meyer, K. 1998a. Estimating covariance functions for longitudinal data using a random regression model. *Genetic, Selection, Evolution* **30**:221-240.

Meyer, K. 1998b**.** "DXMRR" A program to estimate covariance functions for longitudinal data by restricted maximum likelihood. *Proceedings of the 6th world congress on genetic applied to livestock production, Armidale, Australia* **27**: 465-466.

Meyer, K. and Hill, W. G. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal 'repeated' records by restricted maximum likelihood. *Livestock Production Science* **47**: 185-200.

Olori, V.E., Hill, W.G., McGuirk, B.J. and Brotherstone, S. 1998. Estimating variance components for test day milk records by restricted maximum likelihood with random regression animal model. **Accepted,** *Livestock Production Science*.