



Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle

L. Rönnegård,*†¹ M. Felleki,*† W. F. Fikse,* H. A. Mulder,‡§ and E. Strandberg*

*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, PO Box 7023, 750 07 Uppsala, Sweden

†Statistics Unit, School of Technology and Business Studies, Dalarna University, 791 88 Falun, Sweden

‡Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, the Netherlands

§Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH, Wageningen, the Netherlands

ABSTRACT

Trait uniformity, or micro-environmental sensitivity, may be studied through individual differences in residual variance. These differences appear to be heritable, and the need exists, therefore, to fit models to predict breeding values explaining differences in residual variance. The aim of this paper is to estimate breeding values for micro-environmental sensitivity (vEBV) in milk yield and somatic cell score, and their associated variance components, on a large dairy cattle data set having more than 1.6 million records. Estimation of variance components, ordinary breeding values, and vEBV was performed using standard variance component estimation software (ASReml), applying the methodology for double hierarchical generalized linear models. Estimation using ASReml took less than 7 d on a Linux server. The genetic standard deviations for residual variance were 0.21 and 0.22 for somatic cell score and milk yield, respectively, which indicate moderate genetic variance for residual variance and imply that a standard deviation change in vEBV for one of these traits would alter the residual variance by 20%. This study shows that estimation of variance components, estimated breeding values and vEBV, is feasible for large dairy cattle data sets using standard variance component estimation software. The possibility to select for uniformity in Holstein dairy cattle based on these estimates is discussed.

Key words: dairy cattle, genetic heterogeneity, milk yield, somatic cell score

INTRODUCTION

Differences between animals in uniformity for a certain trait may be described in terms of differences in residual variance. For example, for some sires, con-

siderable variation in performance exists within their daughter group, whereas offspring of other sires show relatively little variation. Models for micro-environmental sensitivity include breeding values explaining differences in residual variance (also referred to as genetic heterogeneity of residual variance), and selection for increased uniformity can be performed by selecting animals based on these breeding values. Estimates of such variance-controlling breeding values are referred to as vEBV (analogous to the term vQTL proposed by Rönnegård and Valdar (2011, 2012; i.e., loci controlling residual variance).

Selecting on vEBV is important for situations in which it is economically important to have low variation between individuals, to have trait values near a certain optimum, or when individual trait values run the risk of transgressing a certain (unfavorable) threshold (Mulder et al., 2008). Furthermore, substantial levels of genetic variation for residual variance have been detected in studies from a wide range of animal species with a genetic coefficient of variation around 20% or above (Hill and Mulder, 2010). This implies that a standard deviation change in vEBV for one of these traits would alter the residual variance by 20% or more.

The review of Hill and Mulder (2010) does not report any results for dairy cattle, and even though large differences between sires for daughter group variances for milk yield in Holstein were recognized by Van Vleck (1968) and Clay et al. (1979), the genetic control of these differences has so far not been investigated. Despite the recognized importance of this question, there are strong computational limitations to estimating the vEBV and their associated variance components on large-scale dairy cattle data. Bayesian models have been proposed where the model fitting uses Markov chain Monte Carlo (MCMC) methods, but to date, these have only been applied on data sets including at most 10,000 observations (e.g., Sorensen and Waagepetersen, 2003).

A fast alternative method that has been applied on data from rabbits (Garreau et al., 2008), mice (Gutiérrez et al., 2006), chicken (Mulder et al., 2009), and

Received September 25, 2012.

Accepted December 29, 2012.

¹Corresponding author: lm@du.se

Table 1. Description of the Swedish Holstein data

Item	Data
No. of records	1,693,154
No. of animals	177,411
Years studied	2002–2009
No. of herds	1,759
No. of herd-testdays	21,570
Mean age at calving (d)	838
Somatic cell score	
Mean	2.36
Median	2.05
Variance	2.8
Milk yield (L/d)	
Mean	29.13
Median	29.20
Variance	45.5

beef cattle (Neves et al., 2011, 2012) is to fit an animal model and subsequently use the squared residuals as response in a second animal model. However, this method does not account for the uncertainty in the estimated residuals and therefore gives biased estimates of variance components because the expected variance of estimated residuals is smaller than the true residual variance.

Another method, combining speed and unbiased estimation and based on hierarchical generalized linear models (Lee and Nelder, 1996), was presented by Rönnegård et al. (2010). They showed that a model for genetic heterogeneity can be described as a double hierarchical generalized linear model (**DHGLM**; Lee and Nelder, 2006) and that it can be fitted using standard variance component estimation programs, such as ASReml (Gilmour et al., 2009).

The aim of the current study was to estimate breeding values for environmental sensitivity (vEBV) in milk yield and SCS, and their associated variance components, on a large dairy cattle data set having more than 1.6 million records. To our knowledge, this is the first study to estimate vEBV in dairy cattle. Based on the vEBV, the possibility to select for uniformity in Holstein dairy cattle is discussed.

MATERIALS AND METHODS

Data Description

Data included 1.6 million test-day records on SCS and milk yield for nearly 180,000 Swedish Holstein cows (Table 1, Figure 1). Data included information from first lactation only, and each cow had on average 9.5 recorded test-days. The number of sires having daughter records was 762. Pedigree was traced back such that sires of all cows with records had at least 2 generations of male ancestors known. A standard logarithmic trans-

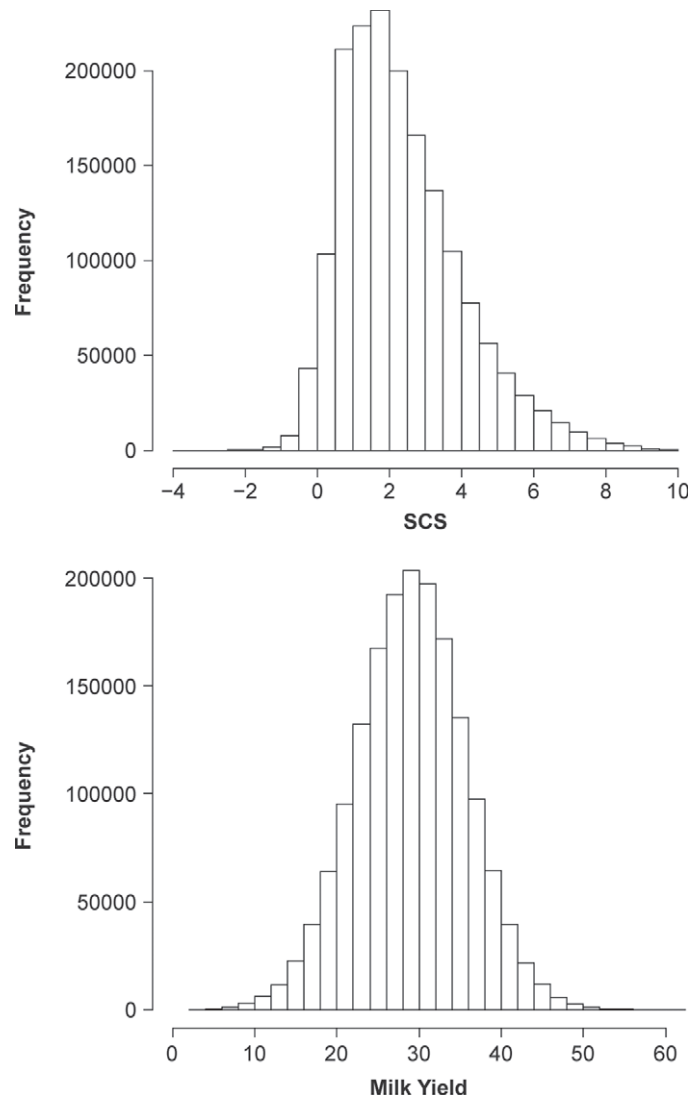


Figure 1. Histograms for the studied traits somatic cell scores and milk yield (L/d).

formation of SCC were used to produce the trait values for SCS [$SCS = \log_2 (SCC/100,000) + 3$].

Statistical Model

The fitted model consists of 2 parts, the mean and the residual variance. The model describing the mean includes the response variable y (either SCS or milk yield), fixed effects β , a random animal effect a , and a random permanent environmental effect p :

$$y = \mathbf{X}\beta + \mathbf{Z}a + \mathbf{W}p + e,$$

where \mathbf{X} , \mathbf{Z} , and \mathbf{W} are design matrices for the fixed and random effects. The animal effects are $a \sim N(0, \mathbf{A}\sigma_a^2)$,

where \mathbf{A} is the additive relationship matrix, and the permanent environmental effects are $p \sim N(0, \mathbf{I}\sigma_p^2)$, where \mathbf{I} is the identity matrix.

The residuals e were assumed to be independently normally distributed but with different variances for each observation. The model for the residual variance is

$$V(e) = \exp(\mathbf{X}_v\beta_v + \mathbf{Z}a_v + \mathbf{W}p_v),$$

where β_v are the fixed effects in the model for the residual variance including the intercept term μ_v , and a_v and p_v are the animal and permanent environmental effects, respectively, in the model for the residual variance. We assume $a_v \sim N(0, \mathbf{A}\sigma_{a_v}^2)$ and $p_v \sim N(0, \mathbf{I}\sigma_{p_v}^2)$. In the current paper, we compute the breeding values for the mean \hat{a} (EBV) and the residual variance \hat{a}_v (vEBV), assuming no specific dependency between a and a_v [i.e., $\text{cov}(a, a_v) = 0$]. Following the modeling by Sorensen and Waagepetersen (2003), it is also assumed that $\text{cov}(p, p_v) = 0$. Because there are repeated measurements within lactation, the vEBV will capture inherited differences in the variation of phenotypic values around the fitted lactation curve in the mean part of the model (as illustrated in Figure 2).

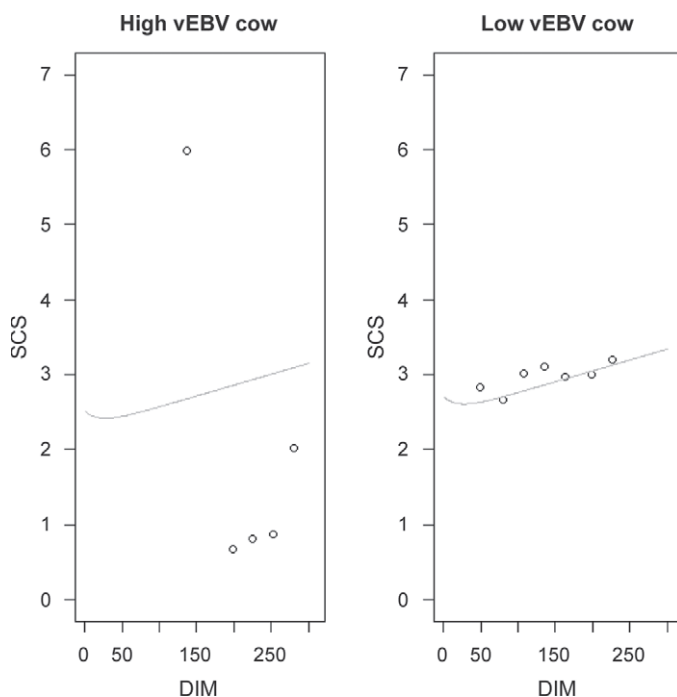


Figure 2. Illustration of the trait measured for EBV for environmental sensitivity (vEBV) in SCS. A cow with high vEBV will have a large variation in measurements around the fitted mean curve (gray). The figure shows observed values for 2 example cows that were found to have high and low vEBV in the subsequent analysis.

The method of interlinked generalized linear models to fit DHGLM by Rönnegård et al. (2010) was used for estimation of variance components and breeding values. The estimation method iterates between several rounds of linear and generalized linear mixed models in AS-Reml (Gilmour et al., 2009) by fitting a weighted linear animal model for the mean part and fitting the adjusted squared residuals from the mean model using a generalized linear animal model (with gamma distribution and log link function as in Rönnegård et al., 2010) for the residual variance part. Predicted observations are obtained for the latter model and used as new weights for the mean model. At convergence, the 2 models give variance components estimates (VCE) $\hat{\sigma}_a^2$, $\hat{\sigma}_p^2$, $\hat{\sigma}_{a_v}^2$, and $\hat{\sigma}_{p_v}^2$, as well as EBV \hat{a} and \hat{a}_v .

The following fixed effects were considered: herd-testday (**htd**), year-season of calving (**ys**), age at calving (**AgeC**), and DIM. Four seasons were defined: January to March, April to June, July to September, and October to December. Adjacent herd-testdays were merged to ensure at least 20 observations within each htd class, using the algorithm by Crump et al. (1997), which resulted in 21,570 htd (Table 1).

Fixed effects included in the mean model were htd, ys, AgeC, (AgeC)², (AgeC)³, DIM, $\exp(-0.05 \times \text{DIM})$, where the modeling of DIM follows Wilmlink (1987). Fixed effects included in the residual variance model were herd, ys, AgeC, (AgeC)², DIM, and (DIM)². Different fixed effects in the mean and residual variance model were chosen to ensure computational feasibility and at the same time capture the most important effects. Here, the modeling of DIM in the residual variance model follows Jaffrezic et al. (2000), which was important because of expected differences in residual variance along the lactation curve, especially as differences between sires' daughter group averages for DIM could be expected. A fixed regression test-day model for the mean was applied, but a DHGLM including random regression could easily be specified. Such modeling would require careful consideration about the order of the random regression polynomial function for the permanent environment effect, because residual heterogeneity could be absorbed by the random regression permanent environment effect; for example, as reported by Ødegård et al. (2003). This approach of accommodating for residual heterogeneity over the course of lactation appears to be expensive, in the sense that increasing the polynomial for the permanent environmental effect from 3 to 4 implies that 4 additional covariance components need to be estimated and a large number of BLUPs. Applying a DHGLM and a second-order fixed polynomial as in the present study requires estimation of only 2 additional regression coef-

Table 2. Estimated parameters for milk yield and SCS in the exponential heterogeneity model (SE in parentheses)

Item	Variance components ¹	Milk yield	SCS
Mean model	σ_a^2	8.78 (0.21)	0.28 (0.011)
	σ_p^2	12.40 (0.14)	1.03 (0.0085)
Residual variance model	$\sigma_{a_v}^2$	0.049 (0.0034)	0.046 (0.0038)
	$\sigma_{p_v}^2$	0.37 (0.0031)	0.61 (0.0040)

¹ σ_a^2 = variance for the additive genetic effects; σ_p^2 = variance for the permanent environmental effects; $\sigma_{a_v}^2$ = variance for the additive genetic effects in the model for the residual variance; $\sigma_{p_v}^2$ = variance for the permanent environmental effects in the model for the residual variance.

ficients in the variance model, and moreover, allows for other sources of heterogeneity to be modeled.

The estimation procedure was implemented in ASReml (Gilmour et al., 2009) and convergence was assessed after 20 iterations. The variance component estimates changed by less than 10^{-3} between the last ASReml runs, and changes in variance components between runs were much smaller than the standard errors on the variance components. The estimation took 7 d per trait on a Linux server.

Two additional analyses were performed to assess the sensitivity of model specifications. The values of SCS were not perfectly normally distributed (Figure 1; skewness 0.89, kurtosis 0.87), which might affect the variance component estimates (Yang et al., 2011); therefore, Box-Cox transformed $[(y^\lambda - 1)/\lambda]$ SCC were used as response in an additional analysis. The sensitivity of specified fixed effects was also analyzed. Here, the potential bias in $\hat{\sigma}_{a_v}^2$ for misspecification of fixed effects was investigated by deleting herd effect from the residual variance part of the model in an additional analysis on both traits.

RESULTS

Estimates for Milk Yield and SCS

The VCE for milk yield were $\hat{\sigma}_a^2 = 8.78$, $\hat{\sigma}_p^2 = 12.40$, $\hat{\sigma}_{a_v}^2 = 0.049$, and $\hat{\sigma}_{p_v}^2 = 0.37$, with small standard errors reported by ASReml. The ratio of estimates to standard errors was greater than 12 for all variance components, indicating accurate parameter estimates (Table 2). The average fitted residual variance was 9.4, and as a reference, estimated residual variance from a linear mixed model (with constant residual variance) was 10.5, and the variance components for the genetic and permanent environmental effects were $\hat{\sigma}_a^2 = 8.2$ and $\hat{\sigma}_p^2 = 11.9$. The VCE for SCS were $\hat{\sigma}_a^2 = 0.28$, $\hat{\sigma}_p^2 = 1.03$, $\hat{\sigma}_{a_v}^2 = 0.046$, and $\hat{\sigma}_{p_v}^2 = 0.61$. The average fitted residual variance was 1.16, and as a reference, estimated residu-

al variance from a linear mixed model was 1.35 and with variance component estimates $\hat{\sigma}_a^2 = 0.27$ and $\hat{\sigma}_p^2 = 0.96$.

Further Interpretation of Estimates

For milk yield, the estimate of the genetic variance in the residual variance was $\hat{\sigma}_{a_v}^2 = 0.049$, which gives a standard deviation of $\hat{\sigma}_{a_v} = 0.22$. Hence, increasing the average vEBV (\bar{a}_v) by 1 $\hat{\sigma}_{a_v}$ increases the residual variance by around 20%, because the model for the residual variance is multiplicative and $\exp(\bar{a}_v + 0.22) = 1.2 \cdot \exp(\bar{a}_v)$. Similarly, decreasing vEBV by 1 $\hat{\sigma}_{a_v}$ would decrease the residual variance by 20%. Decreasing vEBV for SCS with 1 SD would also decrease the residual variance by around 20% for that trait.

When looking at the EBV and vEBV for the 100 sires having daughter groups with the most records, we observed a clear relationship between EBV and daughter group means, and between the vEBV and the daughter group averages of within-cow variances, as expected, for both traits (Figures 3 and 4). The latter relationship is expected because a sire having high vEBV will tend to have daughters with a large variation of trait values within a lactation (as illustrated in Figure 2).

Although we did not fit a correlation between the breeding values a and a_v , a strong positive product-moment correlation between the EBV and vEBV of all 762 sires having daughter records was found for both traits (0.60 and 0.38; Table 3). Hence, the residual variance tends to increase with increased EBV, for both traits. The between-trait correlations were also positive with product-moment correlation values between breeding values between 0.10 and 0.28 (Table 3). A similar pattern in correlations between EBV and vEBV were also found using Calo-type genetic correlations (Calo et al., 1973), which accounts for differences in reliabilities between breeding values (Table 3, lower diagonal values).

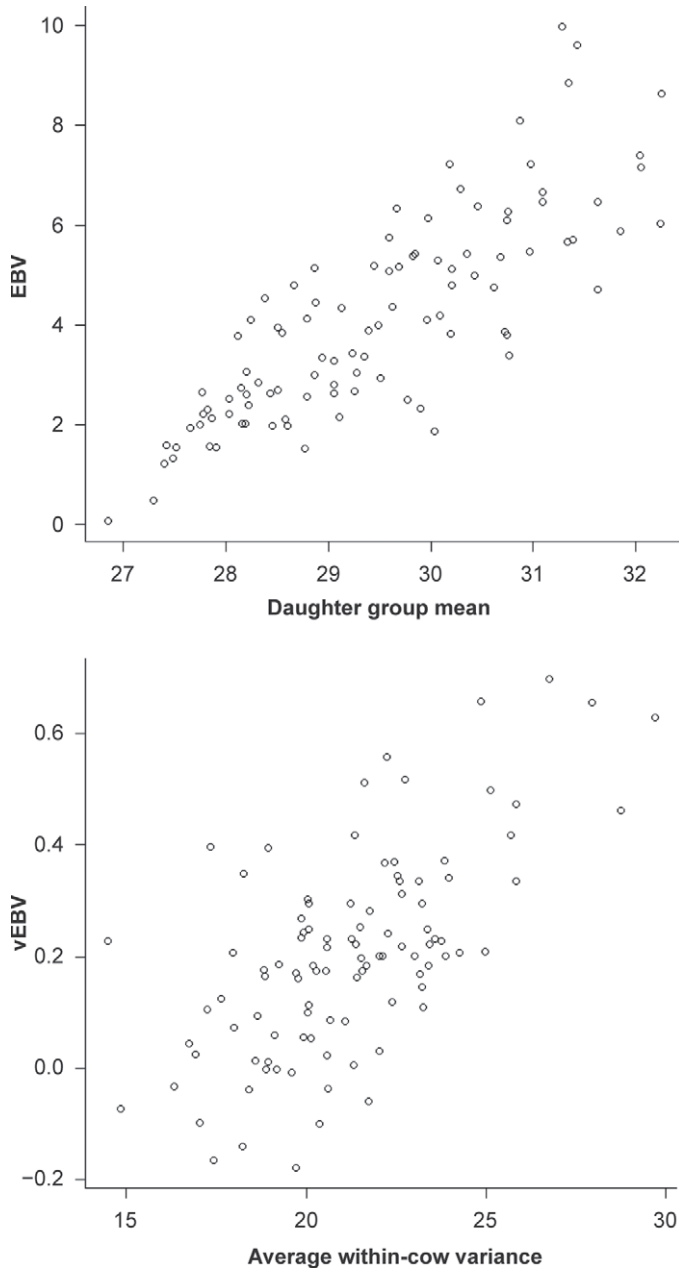


Figure 3. Estimated breeding values and EBV for environmental sensitivity (vEBV) for milk yield plotted against daughter group means and daughter group averages of within-cow variances, respectively, for the 100 sires having the largest daughter groups.

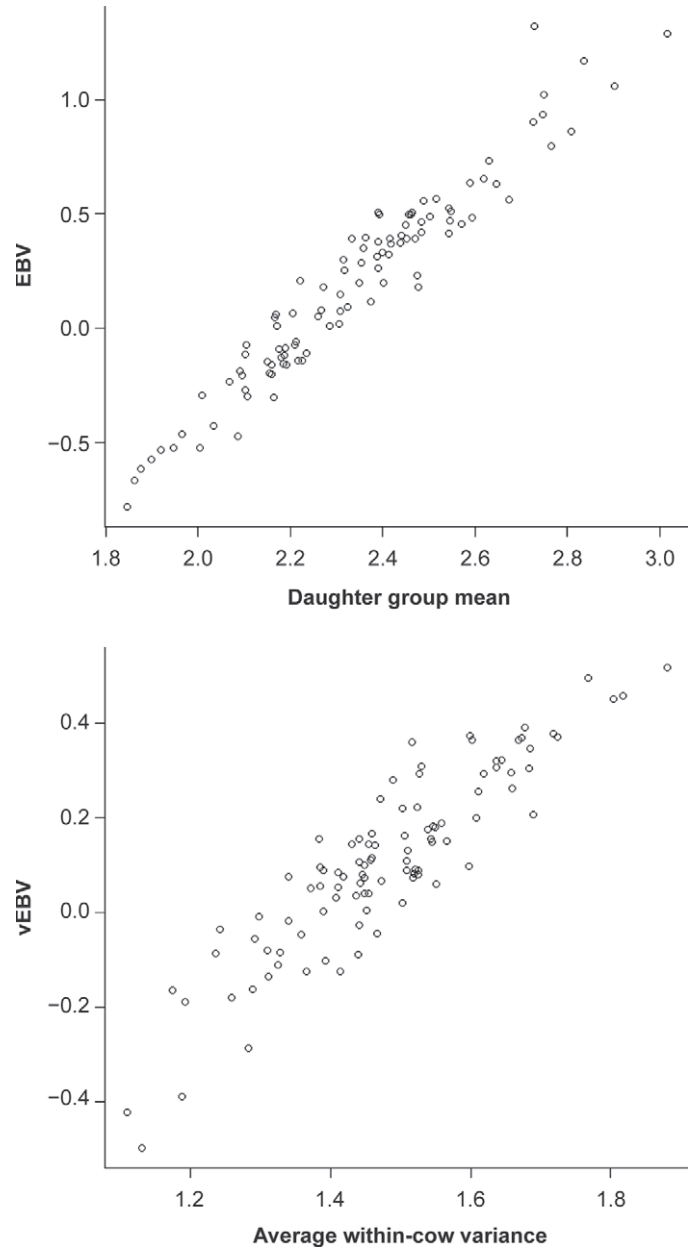


Figure 4. Estimated breeding values and EBV for environmental sensitivity (vEBV) for SCS plotted against daughter group means and daughter group averages of within-cow variances, respectively, for the 100 sires having the largest daughter groups.

The reliability for EBV of sires increased with increasing number of records of the daughter groups per sire, which is also the case for vEBV although these reliabilities are smaller (Figures 5 and 6), because they are more difficult to estimate than the ordinary EBV basically because variances are harder to estimate than means, typically requiring 5 times as many observations to achieve comparable precision (cf. “Tukey’s rule

of 5” in Lee and Nelder, 2006). The reliabilities were obtained from prediction error variances and were not corrected for inbreeding.

The estimated fixed effects of DIM on the residual variance were large, with substantially higher variance in the beginning and the end of the lactation period, for both milk yield and SCS (Figure 7). For milk yield, the residual variance was around 5 times greater at the end

Table 3. Correlations between EBV and EBV for environmental sensitivity (vEBV) for milk yield (MY) and SCS for sires, with Pearson product-moment correlations given above the diagonal and Calo-type genetic correlations given below the diagonal

Item	MY		SCS	
	EBV	vEBV	EBV	vEBV
MY				
EBV		0.598	0.096	0.202
vEBV	0.737		0.215	0.280
SCS				
EBV	0.109	0.274		0.380
vEBV	0.261	0.403	0.508	

compared with the middle of the lactation period. The effect of age at calving was quite small with less than 20% change between extremes.

Sensitivity Analysis and Model Assessment

Applying Box-Cox transformation to the original SCC, the estimate of the genetic variance in the residual variance was $\hat{\sigma}_{a_v}^2 = 0.043$, which is close to the original estimate. We therefore concluded that the estimate was not sensitive to the slight skewness of SCS and a possible need for further transformation. The estimated Box-Cox transformation parameter was $\hat{\lambda} = -0.3$ (SE = 0.001, i.e., significantly $\neq 0$, where $\lambda = 0$ would correspond to the log-transformation used to calculate the SCS).

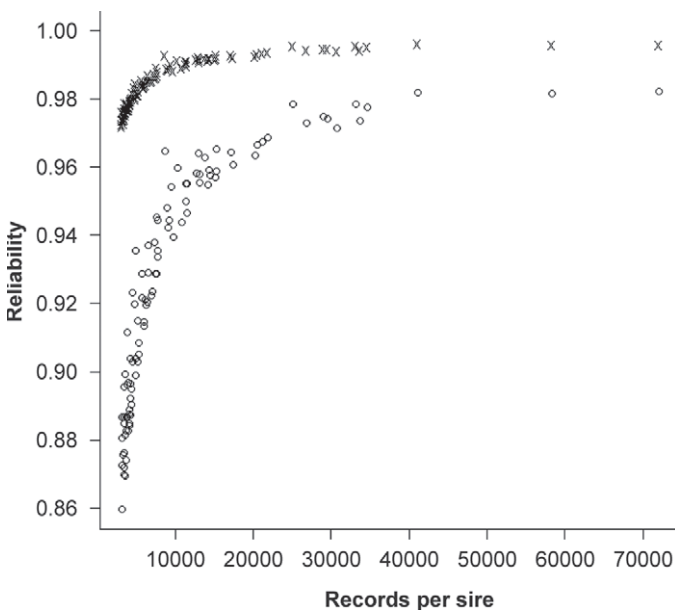


Figure 5. Reliabilities for EBV (x) and EBV for environmental sensitivity (vEBV; o) for milk yield in relation to number of records per sire.

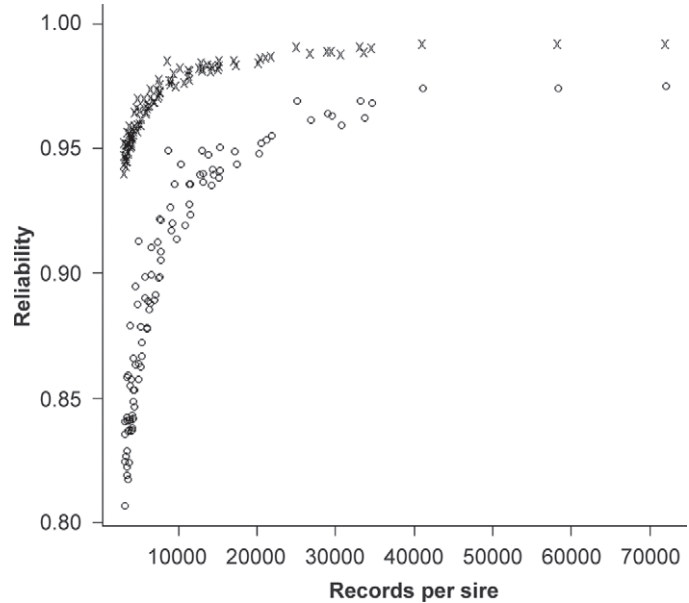


Figure 6. Reliabilities for EBV (x) and EBV for environmental sensitivity (vEBV; o) for SCS in relation to number of records per sire.

The estimates were, however, sensitive to exclusion of herd effects in the residual variance part of the model, producing substantially higher values of $\hat{\sigma}_{a_v}^2$. For SCS, these estimates were $\hat{\sigma}_a^2 = 0.28$, $\hat{\sigma}_p^2 = 1.02$, $\hat{\sigma}_{a_v}^2 = 0.20$, and $\hat{\sigma}_{p_v}^2 = 0.58$, and for milk yield were $\hat{\sigma}_a^2 = 8.79$, $\hat{\sigma}_p^2 = 12.42$, $\hat{\sigma}_{a_v}^2 = 0.27$, and $\hat{\sigma}_{p_v}^2 = 0.30$.

DISCUSSION

In this study, we showed that fitting a model for genetic heterogeneity is possible for large dairy data sets using standard VCE software. The estimated values of the standard deviation for vEBV for milk yield and SCS was around 0.2 ($\hat{\sigma}_{a_v} \approx 0.20$). Hence, a 1-unit SD change in vEBV would cause a 20% change in the residual variance, which is a substantial amount. It should also be noted that considerable residual variation is explained by the fixed effects in the residual variance model (Figure 7). Furthermore, individual cows show different levels of (nonheritable) variation in phenotypic values around the lactation curve, because the estimates of $\sigma_{p_v}^2$ were large and significantly different from zero (Table 2).

The vEBV captures the inherited variation of the phenotypic values around the fitted mean curve within a lactation (Figure 2), which causes differences between sires for daughter group variances. The difference in daughter group variances as observed by Van Vleck (1968) and Clay et al. (1979) might be caused by segre-

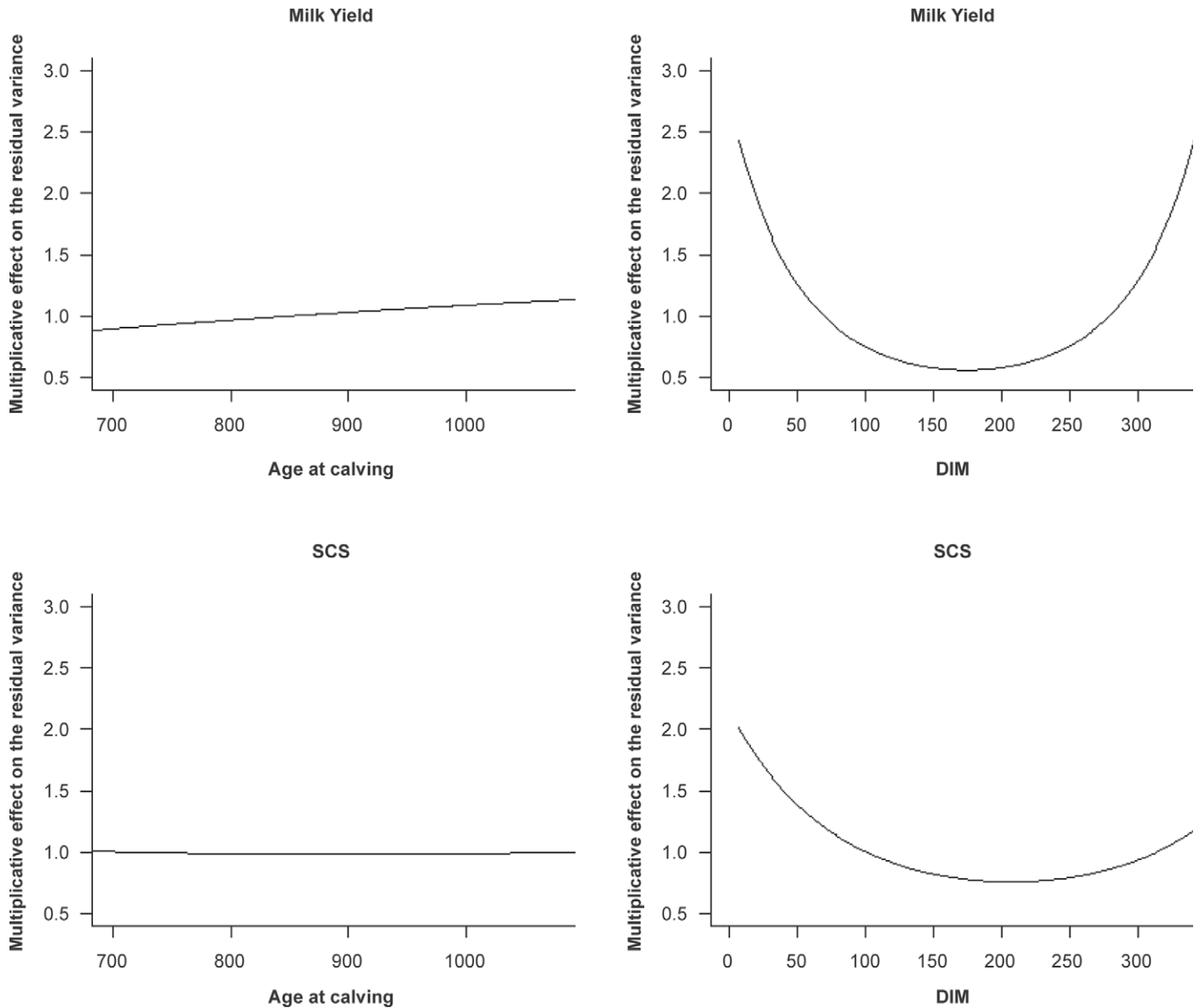


Figure 7. Multiplicative fixed effects of age at calving (d) and DIM on the residual variance for the traits milk yield and SCS.

gation of major genes controlling the trait and selection for reduced vEBV might favor sires homozygous for such a major gene. This scenario is not very likely, however, because not many major genes affecting SCS and milk yield have been found and, as illustrated by Figure 2, selection for reduced vEBV aims at decreasing the variation of phenotypic values around the fitted mean curve of a cow and should not directly affect segregating major genes (i.e., genes having a major effect on the mean). Segregating major genes would cause large differences of lactation means within some daughter groups, whereas a reduction of vEBV would merely reduce the variation around individual lactation curves.

Because the vEBV capture inherited variation around the fitted mean curve, they may also be affected by genetic variation in persistency. It would be of interest, therefore, to study how a model including random regression for DIM in the mean part would influence the vEBV, and vice versa. This is not within the scope of the current study but investigation in this area is proposed for future applications.

Modeling

The method of interlinked generalized linear models to fit DHGLM assumes modeling of the residual

variance using a log-link function that ensures positive values of the predicted residual variance (i.e., the exponential model in Hill and Mulder, 2010). The estimated fixed and random effects in the residual variance part of the model are therefore multiplicative, and the relative effect on the residual variance does not depend on the magnitude of the average residual variance. Hence, relative effects can easily be compared between traits and species.

The model fitted in this study did not include a correlation between the random animal effects in the mean and residual variance parts of the model, which is a parameter of interest (see e.g., Sorensen and Waagepetersen, 2003), because a positive correlation, for instance, would imply that selection on high EBV would also give high vEBV and thereby increase the residual variance. However, we did find positive sample correlations between EBV and vEBV (Table 3), which might be interpreted as a scale effect that could be removed by transformation of the phenotypic values. Scale effects would occur when the variance increases with the mean; that is, the coefficient of variation is rather constant. Yang et al. (2011) pointed out that estimates in a genetic heterogeneity model will be sensitive to model assumptions of normally distributed residuals. A remedy could be to fit a Box-Cox parameter together with all other parameters in the model. Yang et al. (2011) showed that this Box-Cox transformation substantially affected the genetic correlation between EBV and vEBV. A true genetic correlation between EBV and vEBV would generate a small amount of skewness in the data, when all other effects are normally distributed (Mulder et al., 2007). Therefore, in real data, the question is whether the skewness is truly due to a correlation between EBV and vEBV or due to a scale effect. We found that even though the distribution of SCS was slightly skewed, this had very little effect on $\hat{\sigma}_{a_v}^2$ and conclude that it is not very likely that the moderate correlation between EBV and vEBV for SCS could be explained by a pure scale effect and that there is no need for further transformation of the data.

The sensitivity analysis, in which herd effects were left out from the residual variance model, showed that careful modeling of fixed and random effects for residual variance is important to obtain unbiased estimates of variance components. Leaving out herd effects from the residual variance part of the model resulted in a 5-fold increase in $\hat{\sigma}_{a_v}^2$. Heterogeneity of variance among herds is very common in dairy cattle (e.g., Brotherstone and Hill, 1986; Weigel and Gianola, 1993; Meuwissen et al., 1996). The suggested DHGLM can be nicely used to account for heterogeneity of residual variance among herds or other environmental effects.

The variance component estimates for the mean part of the model (i.e., $\hat{\sigma}_a^2$ and $\hat{\sigma}_p^2$) were slightly greater under the DHGLM compared with a homoscedastic linear mixed model for both SCS and milk yield. In Rönnegård et al. (2010), the VCE for the mean part of the model were slightly lower under a DHGLM for pig litter size compared with a homoscedastic model. The differences were quite small and we could not conclude in which direction heteroscedasticity affects variance component estimates in general.

Possibilities for Genetic Improvement

The estimated values of the standard deviation for vEBV for milk yield and SCS ($\hat{\sigma}_{a_v} \approx 0.20$) were in the lower range of previous estimates in other species reported by Hill and Mulder (2010; “genetic coefficient of variation”), but suggest that the genetic variance in residual variance is large relative to the residual variance itself, indicating that residual variance could be changed by selection. The Mulder-Hill heritability for the residual variance (i.e., the regression of a_v on squared phenotype values; Mulder et al., 2007) was around 0.005 for both traits, which is low compared with previous estimates in other species (Hill and Mulder, 2010). This value shows that high numbers of records per sire are required to obtain accurate vEBV to achieve substantial genetic response in residual variance (Mulder et al., 2007), which is also evident from the reliabilities of vEBV (Figures 5 and 6).

Based on the genetic variance in residual variance and when having large amounts of records per sire, it seems possible to select for reduced residual variance. Mulder et al. (2008) argued that selection for vEBV is only relevant if the profit function is nonlinear. Reported profit functions for milk yield are typically linear, suggesting that selection on vEBV may not have priority, but this ignores the fact that uniformity of production and product quality has a value of its own. For example, less variation between the highest- and lowest-producing cow in a herd will ease adapting diet composition to production level. In addition, low residual heterogeneity indicates that cows are less sensitive to small disturbances in the production environment, and is as such a measure of uniformity. In summary, it appears to be beneficial to reduce the residual heterogeneity for milk yield by breeding and selection, but an estimate of the economic value of vEBV is needed to properly weight this trait in the breeding goal. Deriving economic values for vEBV may need bio-economic modeling (e.g., Groen, 1988) rather than profit equations, which might be too simplified. Assuming that economic value of vEBV is negative (i.e., favoring uniformity), it

may be difficult to reduce residual variance or vEBV because of the unfavorable positive correlation between EBV and vEBV for milk yield (Table 3).

For SCS, the positive product-moment correlation (calculated indirectly from the EBV and vEBV) is favorable, because selection would aim to reduce both the level and variation of SCS. Selection for reduced vEBV is motivated by its relation with incidence of clinical and subclinical mastitis; Urioste et al. (2012) reported genetic correlations between the raw SD (a proxy of vEBV) and clinical and subclinical mastitis of 0.8 and 0.9, respectively, for the same population. Selection for EBV of SCS has an economic value in its own right because payment systems for milk often include a penalty for bulk tank SCS that are too high (e.g., Veerkamp et al., 1998). In addition, there are some concerns that the base level of SCC should not be reduced (Shook and Schutz, 1994; De Haas, 2003). This could be translated to mean that SCC levels are ideally between a lower and an upper bound (i.e., an optimum range). Based on this principle of an optimum range between a lower and upper bound, economic values were derived for the EBV for mean SCS and for vEBV using the equations for differential profit with thresholds given in Mulder et al. (2008). For simplicity, the lower bound for SCC was assumed to be 40 and the economic values for EBV and vEBV as a function of the upper bound for SCC were calculated (Figure 8) for the studied Swedish Holstein population. This reflects different price penalty systems in which the penalty depends on the upper threshold applied. Figure 8 shows that the economic value for vEBV is substantially higher, in absolute terms, than that for the EBV when the upper threshold is >350. Here, selection for lower variance of SCS is more important than decreasing the mean of SCS, because the current population mean is already within the desired optimum range. The economic value of EBV is, however, quite sensitive to the thresholds and the population mean.

Future Developments

Model selection tools and more formal testing of VCE than Wald-type of tests using standard errors should be developed, which is possible using Lee and Nelder's h-likelihood framework for DHGLM (Lee and Nelder, 1996; Lee et al., 2006). Such tools would facilitate the possibility for selection between models including both random regression for DIM in the mean model and genetic heterogeneity in the residual variance. Developing a model including multiple traits should be theoretically straightforward but might be computationally challenging. The model applied in our current study does not include a parameter of the correlation between random

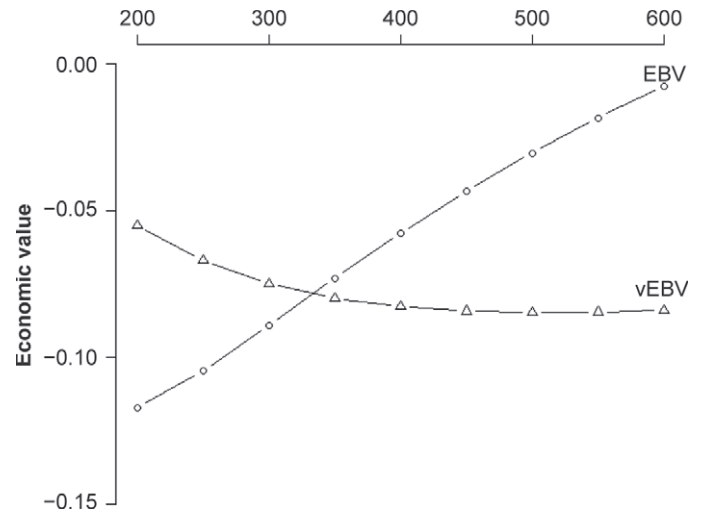


Figure 8. Economic values for EBV and vEBV for environmental sensitivity (vEBV) in the Swedish Holstein population as a function of the upper threshold in selection for optimized values of SCC (lower threshold = 40).

effects in the mean and residual variance parts of the model. Such an extension should also be possible using the h-likelihood and would give further insight to the analysis of EBV and vEBV. Furthermore, detection of loci affecting the variance (see Rönnegård and Valdar, 2012) would substantially improve our understanding of vEBV for SCS and milk yield. A study investigating the relationship between vEBV for SCS and clinical mastitis would also give insights for using EBV and vEBV for SCS to select against mastitis.

CONCLUSIONS

We estimated EBV and vEBV of milk yield and SCS in the Swedish Holstein population based on data from 2002 to 2009. The study shows that this is possible using standard VCE software (e.g., ASReml). A change in vEBV of 1 genetic SD for one of these traits would alter the residual variance by 20% in the Swedish Holstein population. Hence, the residual variance could be changed by selection in dairy cattle for traits such as milk yield and SCS, but estimates of both EBV and vEBV requires large data sets. Furthermore, an investigation of economic value showed that vEBV can be more important than EBV to consider when selecting for a trait within an optimum range in Holstein.

ACKNOWLEDGMENTS

The authors are grateful to the Swedish Dairy Association (Vattugatan, Stockholm) and farmers for providing the data. This project was financed by the

RobustMilk project, which is financially supported by the European Commission under the Seventh Research Framework Programme (Grant Agreement KBBE-211708). The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the Commission or its services.

REFERENCES

- Brotherstone, S., and W. G. Hill. 1986. Heterogeneity of variance amongst herds for milk production. *Anim. Prod.* 42:297–303.
- Calo, L. L., R. E. McDowell, L. D. VanVleck, and P. D. Miller. 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. *J. Anim. Sci.* 37:676–682.
- Clay, J. S., W. E. Vinson, and J. M. White. 1979. Heterogeneity of daughter variances of sires for milk yield. *J. Dairy Sci.* 62:985–989.
- Crump, R. E., C. S. Haley, R. Thompson, and J. Mercer. 1997. Assigning pedigree beef performance records to contemporary groups taking account of within-herd calving patterns. *Anim. Sci.* 65:193–198.
- De Haas, Y. 2003. Somatic cell count patterns: Improvement of udder health by genetics and management. PhD Thesis. Wageningen Univ., Wageningen, the Netherlands.
- Garreau, H., G. Bolet, C. Larzul, C. Robert-Granie, G. Saleil, M. San-Cristobal, and L. Bodin. 2008. Results of four generations of canalizing selection for rabbit birth weight. *Livest. Sci.* 119:55–62.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0. VSN International Ltd., Hemel Hempstead, UK.
- Groen, A. F. 1988. Derivation of economic values in cattle breeding: A model at farm level. *Agric. Syst.* 27:195–213.
- Gutiérrez, J. P., B. Nieto, P. Piqueras, N. Ibáñez, and C. Salgado. 2006. Genetic parameters for canalisation analysis of litter size and litter weight traits at birth in mice. *Genet. Sel. Evol.* 38:445–462.
- Hill, W. G., and H. A. Mulder. 2010. Genetic analysis of environmental variation. *Genet. Res. (Camb.)* 92:381–395.
- Jaffrezic, F., I. M. S. White, R. Thompson, and W. G. Hill. 2000. A link function approach to model heterogeneity of residual variances over time in lactation curve analyses. *J. Dairy Sci.* 83:1089–1093.
- Lee, Y., and J. A. Nelder. 1996. Hierarchical generalized linear models (with Discussion). *J. R. Stat. Soc. B* 58:619–678.
- Lee, Y., and J. A. Nelder. 2006. Double hierarchical generalized linear models. *Appl. Stat.* 55:139–185.
- Lee, Y., J. A. Nelder, and Y. Pawitan. 2006. Generalized linear models with random effects: Unified analysis via h-likelihood. Chapman & Hall/CRC, London, UK.
- Meuwissen, T. H. E., G. De Jong, and B. Engel. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci.* 79:310–316.
- Mulder, H. A., P. Bijma, and W. G. Hill. 2007. Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. *Genetics* 175:1895–1910.
- Mulder, H. A., P. Bijma, and W. G. Hill. 2008. Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genet. Sel. Evol.* 40:37–59.
- Mulder, H. A., W. G. Hill, A. Vereijken, and R. F. Veerkamp. 2009. Estimation of genetic variation in residual variance in female and male broilers. *Animal* 3:1673–1680.
- Neves, H. H. R., R. Carvalheiro, and S. A. Queiroz. 2012. Genetic variability of residual variance of weight traits in Nelore beef cattle. *Genet. Sel. Evol.* 44:19.
- Neves, H. H. R., R. Carvalheiro, V. M. Roso, and S. A. Queiroz. 2011. Genetic variability of residual variance of production traits in Nelore beef cattle. *Livest. Sci.* 142:164–169.
- Ødegård, J., J. Jensen, G. Klemetsdal, P. Madsen, and B. Heringstad. 2003. Genetic analysis of somatic cell score in Norwegian cattle using random regression test-day models. *J. Dairy Sci.* 86:4103–4114.
- Rönnegård, L., M. Felleki, F. Fikse, H. A. Mulder, and E. Strandberg. 2010. Genetic heterogeneity of residual variance—Estimation of variance components using double hierarchical generalized linear models. *Genet. Sel. Evol.* 42:8.
- Rönnegård, L., and W. Valdar. 2011. Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics* 188:435–447.
- Rönnegård, L., and W. Valdar. 2012. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet.* 13:63.
- Shook, G. E., and M. M. Schutz. 1994. Selection on somatic cell score to improve resistance to mastitis in the United States. *J. Dairy Sci.* 77:648–658.
- Sorensen, D., and R. Waagepetersen. 2003. Normal linear models with genetically structured residual variance heterogeneity: A case study. *Genet. Res.* 82:207–222.
- Urioste, J. I., J. Franzén, J. J. Windig, and E. Strandberg. 2012. Genetic relationships among mastitis and alternative somatic cell count traits in the first three lactations of Swedish Holsteins. *J. Dairy Sci.* 95:3428–3434.
- Van Vleck, L. D. 1968. Variation of milk records within paternal-sib groups. *J. Dairy Sci.* 51:1465–1470.
- Veerkamp, R. F., A. W. Stott, W. G. Hill, and S. Brotherstone. 1998. The economic value of somatic cell count payment schemes for UK dairy cattle breeding programmes. *Anim. Sci.* 66:293–298.
- Weigel, K. A., and D. Gianola. 1993. A computationally simple Bayesian method for estimation of heterogeneous within-herd phenotypic variances. *J. Dairy Sci.* 76:1455–1465.
- Wilmink, J. B. M. 1987. Adjustment of test day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16:335–348.
- Yang, Y., O. F. Christensen, and D. Sorensen. 2011. Analysis of a genetically structured variance heterogeneity model using the Box-Cox transformation. *Genet. Res. (Camb.)* 93:33–46.