J. Dairy Sci. 96:2627–2636
http://dx.doi.org/10.3168/jds.2012-6198

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, considerable 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 Waagepeterson, 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.
[1] Corresponding author: lrn@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 transformation of SCC were used to produce the trait values for SCS [SCS = $\log_2$ (SCC/100,000) + 3].

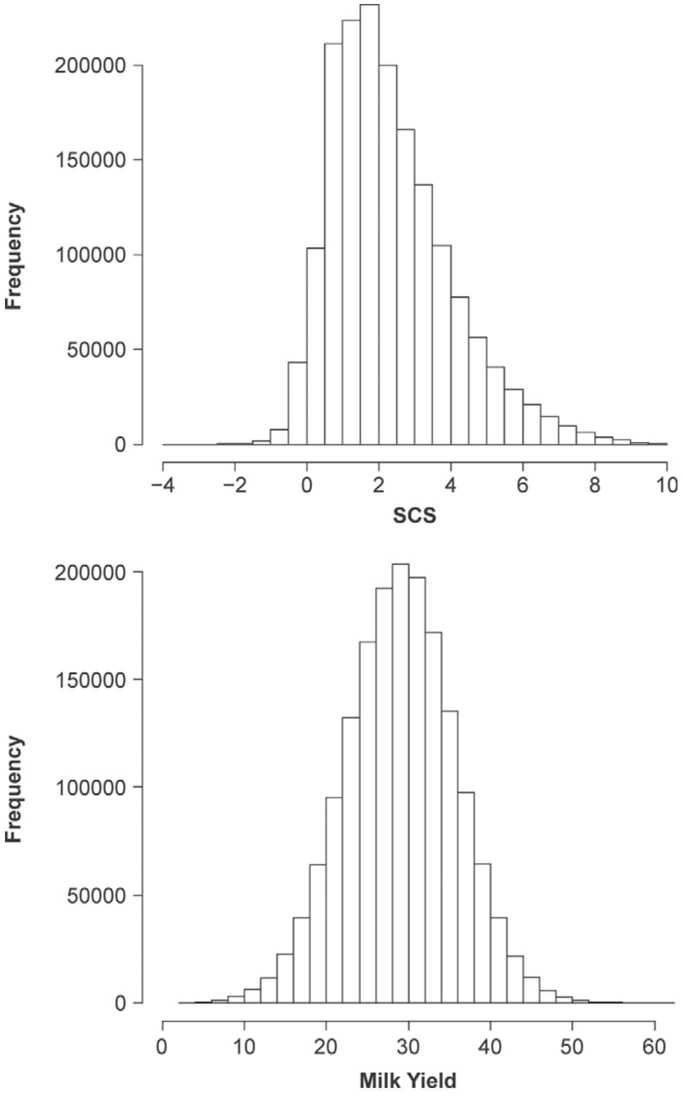

Figure 1. Histograms for the studied traits somatic cell scores and milk yield (L/d).

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