

Heterogeneity of variance and its implications on dairy cattle breeding

T.K. Muasya^{1,3#}, E.D. Ilatsia¹, T.M. Magothe^{2,3} and A.K. Kahi³

¹Animal Genetic Resources Group, National Animal Husbandry Research Centre, Kenya Agricultural Research Institute, P.O. Box 25, 20117 Naivasha, Kenya

²Livestock Recording Centre, Ministry of Livestock and Fisheries Development, P.O. Box 257, 20117 Naivasha, Kenya

³Animal Breeding and Genetics Group, Department of Animal Sciences, Egerton University, P.O. Box 536, 20115 Egerton, Kenya

Abstract

Milk yield data ($n = 12307$) from 116 Holstein-Friesian herds were grouped into three production environments based on mean and standard deviation of herd 305-day milk yield and evaluated for within herd variation using univariate animal model procedures. Variance components were estimated by derivative free REML algorithm, and significance tests done using the F_{\max} procedure. Phenotypic, additive genetic and residual variances were heterogeneous across production environments. The estimates of variances in kg^2 were 708349.9, 966853.7 and 2229338.9; 112266.9, 209949.9 and 689942.1; and 512660.0, 734854.5 and 1317808.3 for phenotypic, additive genetic and residual variances, respectively for production environments 1, 2 and 3. The heritability estimates were 0.15 ± 0.04 , 0.22 ± 0.05 and 0.31 ± 0.03 , respectively. The magnitude of estimated breeding values for sires and their ranking was influenced by production environment. Each production environment had a unique set of superior sires. The more variable production environments contributed more bull-dams at various selection intensities. Effects of heterogeneous variances need to be accounted for in genetic evaluation for Holstein-Friesian in Kenya.

Keywords: Variance components, heterogeneity, milk yield, Holstein-Friesian

[#]Corresponding author. E-mail: muasyakt@yahoo.com

Introduction

Most animal genetic evaluations assume that animal performance data came from the same population with constant variation among herds and no genotype by herd interaction. Dairy cattle evaluation using Best Linear Unbiased Predictions (BLUP) requires appropriate variance components to provide solutions. Use of BLUP assumes independence of genetic and environmental variances from the mean and that they are homogenous across herds or environments, and that the genetic correlation between genetic values in different environmental variance groups is unity (Meyer, 1989).

Different dairy cattle production environments have unique constraints (Stotz, 1983) which lead to heterogeneity of parameters and variance ratios (e.g. heritability) (Costa *et al.*, 2000; Olukoye & Mosi, 2002) and may be related to genotype by environment interaction and feeding management (Brotherstone & Hill, 1986). Heteroscedasticity across production environments reduces the accuracy of predicted breeding values relative to the population mean and can lead to favouring high performers from more variable herds over high performers from low-variance herds, causing a reduction in response to selection (Hill, 1984). These biases in evaluations accumulate over time as dams and daughters tend to express records in the same herds or environments and may become more severe as intensity of selection increases (Vinson, 1987).

Clustering of herds into production environments where climate and management differ (Weigel & Rekaya, 2000), helps to increase the precision of genetic parameter estimates and also facilitates borderless evaluations. In Kenya's Holstein-Friesian cattle population, non-genetic factors are an important cause of heterogeneity of variance at phenotypic level (Olukoye & Mosi, 2002). However, response to selection is also influenced by selection intensity and heritability apart from phenotypic standard deviation (Falconer & Mackay, 1996), and therefore genetic and residual variances should be investigated for heteroscedasticity. Dairy cattle evaluation using BLUP is just being implemented in Kenya (Ojango & Pollot, 2001; Magothe *et al.*, 2006) and the usefulness of the evaluations will depend on how well its assumptions match the data. The objective of the study was to test existence of heterogeneity of variance of

milk yield as influenced by herd production environment in Kenya and determine its effect on sire and bull dam selection.

Materials and Methods

Milk yield data were obtained from the Dairy Recording Services of Kenya (DRSK) on 116 dairy cattle herds participating in performance recording. DRSK is a private and voluntary farmer funded organisation which was started by breed societies and carries out all official milk recording and butterfat testing and makes available records to the Livestock Recording Centre (LRC) for data analyses. The results are then submitted to the Central Artificial Insemination Station for contract mating schemes which test and select sires for the national breeding programme. A total of 12628 records were initially available before cleaning and editing. Records included in the analyses were of cows that calved between 1985 and 2005 and had completed the current lactation by the time of the analysis. A record was included in the final dataset if it also contained information on pedigree, season and year of calving, parity and herd of each cow. After cleaning and editing, 12307 records spanning 22 years and in four seasons and 12 parities were available for analyses. The herds were kept in high (Agro-ecological zone III and below) and medium potential areas (agro-ecological zone IV). The animals were kept under different production systems where grazing was on natural pastures or fodder was grown, harvested and fed to the animals. The 305-day yield was derived by cumulating yield up to 305 days, or by linear regression for cows with lactations shorter than 305 days but at least 150 days.

The mean and standard deviation of 305-day milk yield for each herd were obtained and used as the basis of clustering the herds. The two variables provided a measure of production intensity in each production environment (Naya *et al.*, 2002; Raffrenato *et al.*, 2003).

Preliminary analyses using PROC GLM of SAS (2002) were carried out to determine the important non-genetic factors for inclusion in the final model. The following fixed effects model was used for the exploratory analyses:

$$y_{ijklm} = \mu + H_i + P_j + SC_k + YC_l + e_{ijkl} \quad 1$$

where, y_{ijklm} = 305-day milk yield, μ = underlying mean, H = fixed herd effect with $I=1, 2, \dots, 116$, P_j = fixed effect of parity with $j = 1, 2, \dots, 12$, SC_k = fixed effect of season of calving with $k = 1, 2, 3, 4$, YC_l = fixed effect of year of calving with $l = 1, 2, \dots, 22$ and e_{ijkl} = random residual error distributed as $N(0, I\sigma_e^2)$.

Identification of different production environments was performed with cluster analysis using herd production level and standard deviation of herd 305-day milk yield. The herds were grouped by hierarchical cluster analysis using PROC Cluster of SAS (2002), with minimum variances within group. Derivation of the appropriate number of clusters was based on the pseudo F-statistic.

Genetic characterization of the clusters was done by estimating phenotypic, additive genetic, residual variances and heritability using a univariate animal model. Variance components were estimated using the DFREML package (Meyer, 1998) and the F_{max} procedure was used to test for homogenous variances. The following mixed model was fitted:

$$y_{ijklm} = \mu + H_i + P_j + SC_k + YC_l + a_m + e_{ijklm} \quad 2$$

where, a_m = random additive genetic effects distributed as $N(0, A\sigma_A^2)$, where A is the numerator relationship matrix and σ_A^2 is the additive genetic variance. Only fixed effects that were significant as defined in equation 1 were included in this model.

Results and Discussion

Three clusters were formed and for each the mean 305-day milk yield (**MY**), average standard deviation (**SDMY**), number of herds and the resultant number of records are shown in Table 1. There were a total of 693 sires in the data set of which 405, 218 and 377 had progeny records in clusters 1, 2 and 3, respectively, while only 71 sires had records in all clusters. The herd clusters differed ($P < 0.05$) for the two descriptive variables.

The differences in clusters in their original variables show that milk yield varies with the level of management and environment (Costa *et al.*, 2000) and this could be due to general feeding and genotype by

environment correlation and methods of feeding concentrates (Brotherstone & Hill, 1986). A more effective management would elicit higher performance and greater variability within a production environment (Boldman & Freeman, 1990; Naya *et al.*, 2002; Raffrenato *et al.*, 2003).

Table 1 Mean and standard deviation for 305-day milk yield by cluster

Production environment	No. of records	MY, kg	SDMY
Cluster 1	3689	3155.6 ^a	1057 ^a
Cluster 2	2474	4424.5 ^b	1283 ^b
Cluster 3	6144	5750.6 ^c	1500 ^c
Overall	12307	4826.8	1908

^{a,b,c} rows within column with common superscripts do not differ ($P > 0.05$)

MY - 305-day milk yield

SDMY - standard deviation of 305-day milk yield

Phenotypic, additive genetic, residual variances and heritability estimates are presented in Table 2 for each production environment. Estimates of phenotypic, additive genetic and residual variances increased with herd production environment, as did heritability estimates. The F_{\max} statistical test revealed that all the variances (phenotypic, additive genetic, residual variances) were different ($P < 0.05$). The results in Table 2 clearly indicate that there is heteroscedasticity of variance components for milk yield.

Table 2 Estimates of variance components and heritability estimates (\pm s.e) for milk yield in different production environments

Production environment	Phenotypic variance (kg ²)	Additive genetic variance (kg ²)	Residual variance (kg ²)	Heritability
Cluster 1	708349.9 ^a	112266.9 ^a	512660.0 ^a	0.15 \pm 0.04
Cluster 2	966853.7 ^b	209949.9 ^b	734854.5 ^b	0.22 \pm 0.05
Cluster 3	2229338.9 ^c	689942.1 ^c	1317808.3 ^c	0.31 \pm 0.03
Overall	1843122.8	593356.3	1089928.9	0.32 \pm 0.02

^{a,b,c} rows within column with common superscripts do not differ ($P > 0.05$)

The effect of heterogeneous variances on estimated sire breeding values (estimated progeny differences) is shown in Figure 1. The Estimated Progeny Differences for all sires changed positively across the clusters by a magnitude of 3.2% while the breeding values of sires common to all the three clusters changed positively by a magnitude of 37.3% (Fig. 1). The magnitude of breeding value estimates for common sires was different and increased with cluster variance (Fig. 1).

The 10 best sires for each cluster are given in Table 3. When ranked by order of superiority, no sire was common to all clusters. The results indicate that each production environment requires a different set of sires if selection were to be based on improved milk yield.

The proportion of cows selected as bull-dams from the different production environments is presented in Table 4. There were a total of 3403 cows with positive estimated breeding values from the dataset analysed. Cluster 3, which was most variable (Table 3), accounted for 51% of all cows with positive estimated genetic merit. Clusters 1 and 2 contributed only 28% and 21% of cows to this group, respectively

(Table 4). As selection intensity increased the number of cows selected from the more variable cluster (cluster 3) increased at the expense of those from the less variable clusters.

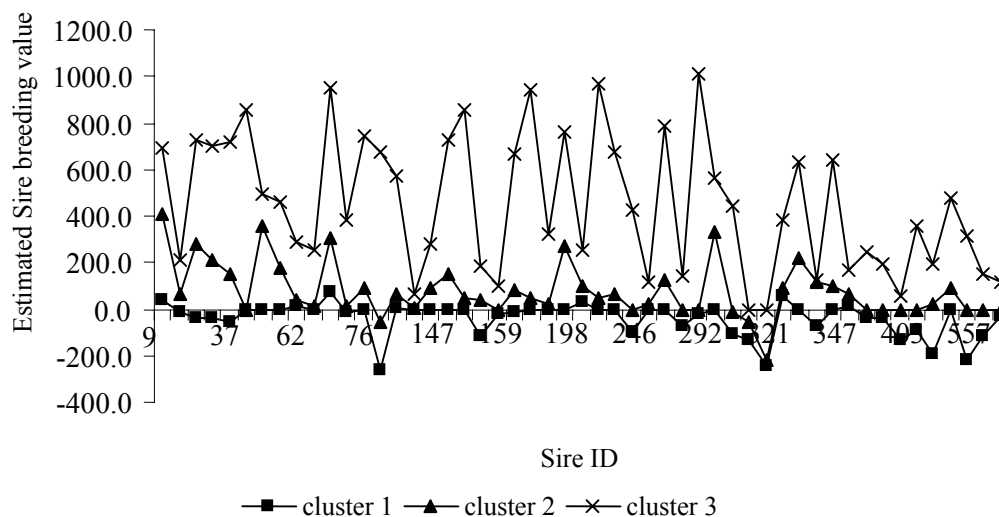


Figure 1 Estimated breeding values of common sires by cluster

Table 3 Ten best sires in each production environment in order of superiority

Cluster 1		Cluster 2		Cluster 3	
Sire ID	EBV	Sire ID	EBV	Sire ID	EBV
11	672.4	46	1167.9	139	2304.6
200	656.3	42	1070.6	410	1919.8
256	557.1	73	758.7	293	1739.6
439	517.1	37	623.0	296	1666.1
528	502.4	336	555.9	31	1653.0
106	499.3	84	477.9	65	1648.9
485	498.7	9	457.1	284	1570.0
563	429.6	347	441.3	310	1563.1
95	379.1	24	430.7	104	1445.9
371	349.1	489	412.1	411	1404.2

EBV - estimated breeding value

Response to selection is a function of selection intensity, heritability and phenotypic standard deviation. The change in variance components across the production environments, especially heritability has been shown to reduce the accuracy of predicted breeding values relative to the population mean and therefore affects response to selection (Hill, 1984). The current study has shown that number of individuals selected from each cluster would differ as long as the heritability estimates are different. High performers in less variable herds are likely to be undervalued while high performers in highly variable herds would be overvalued (Hill 1984). Such biases influence the selection of Bull-dams and superior sires in a breeding

programme such as Kenya's, and would cause a reduction in response to selection (Hill, 1984) and/or genotype by environment interaction. These biases in evaluations accumulate over time as dams and daughters tend to express records in the same herds or environments (Vinson, 1987). While different sires were shown to be suitable for different production environments, meeting each cluster's sire needs would be expensive to implement (Lin & Togashi, 2002). The production environments may not be large enough to sustain sufficient genetic diversity and selection intensity, and inbreeding levels might increase over time.

Table 4 Proportion of cows selected from each production environment at varying selection intensities (3, 6, 15 and 29%)

Cluster	Selection intensity				
	Overall	Top 100	Top 200	Top 500	Top 1000
1	954 (28%)	2 (2%)	5 (3%)	29 (6%)	92 (9%)
2	723 (21%)	5 (5%)	8 (4%)	39 (8%)	128 (13%)
3	1726 (51%)	93 (93%)	187 (94%)	432 (86%)	780 (78%)
Total	3403	100	200	500	1000

Conclusions

Variance components for Holstein-Friesian herds in Kenya for milk yield are influenced by herd production level. Heterogeneous variance components could be accounted for by scaling of observations within herd or environment standard deviations, treating random effects as separate traits in different environments or by use of multiplicative mixed-model equations in genetic evaluations to avoid selection biases that arise from favouring high performers in more variable herds.

Acknowledgements

The authors wish to acknowledge the LRC and DRSK for provision of data and the Kenya Agricultural Research Institute and Egerton University for provision of computing facilities.

References

- Brotherstone, S. & Hill, W.G., 1986. Heterogeneity of variance among herds for milk production. *Anim. Prod.* 42, 297-303.
- Boldman, K.G. & Freeman, A.E., 1990. Adjustments for heterogeneity of variances by herd level in dairy cows and sire evaluation. *J. Dairy Sci.* 73, 503-512.
- Costa, C.N., Blake, R.W., Pollak, E.J., Oltenacu, P.A., Quaas, R.L. & Searle, S.R., 2000. Genetic analysis of Holstein cattle populations in Brazil and the United States. *J. Dairy Sci.* 83, 2963-2974.
- Falconer, D.S. & Mackay, T.F.C., 1996. "Introduction to quantitative genetics". Fourth edition, Longman, Harlow, UK.
- Graser, H-U., Smith, S.P. & Tier, B., 1987. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.* 64, 1362-1370.
- Hill, W.G., 1984. On selection among groups with heterogeneous variance. *Anim. Prod.* 39, 473-477.
- Lin, C.Y. & Togashi, K., 2002. Genetic improvement in the presence of genotype by environment interaction. *Anim. Sci. J.* 73, 3-11.
- Magothe, T.M., Ilatsia, E.D., Wasike, C.B., Migose, S.A. & Kahi, A.K., 2006. Genetic evaluation of milk yield of *Bos taurus* dairy breeds in Kenya. In: Proceedings of the 10th KARI scientific conference, Nairobi, Kenya (in press).
- Meyer, K., 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using derivative free algorithms. *Genet Sel. Evol.* 21, 318-324.
- Meyer, K., 1998. DPREML. Version 3a, user notes, Animal breeding and Genetics Unit, University of New England, Armidale, NSW, Australia.

- Naya, H., Urioste, J.I. & Franco, J., 2002. Identification of production environments and presence of G x E interactions in Uruguay using Holstein herds records. In Proc. 7th World Congress on Genetics Applied to Livestock Production, August 19-23, 2002, Montpellier, France.
- Ojango, J.M.K. & Pollot, K., 2001. Genetics of milk yield and fertility traits in Holstein-Friesian cattle on large scale Kenyan farms. *J. Anim. Sci.* 79, 1742-1750.
- Olukoye, G.A. & Mosi, R.O., 2002. Non-genetic causes of heterogeneity of variance in milk yield among Holstein-Friesian herds in Kenya. *The Kenya Veterinarian* 25, 18-23.
- Raffrenato, E., Blake, R.W., Oltenacu, P.A., Carvalheira, J. & Licitra, G., 2003. Genotype by environment interaction for yield and somatic cell score *J. Dairy Sci.* 86, 2470–2479.
- SAS, 2002. The Statistical Analysis System, Version 8 for windows. SAS Institute Inc., Cary NC, USA.
- Stotz, D., 1983. Production techniques and economics of smallholder production systems in Kenya. In: *Farm Management Handbook of Kenya, Vol. IV*, Nairobi, Kenya.
- Vinson, W.E., 1987. Potential bias in genetic evaluations from difference in variation within herds. *J. Dairy Sci.* 70, 2450-2455.
- Weigel, K.A. & Rekaya, R., 2000. A multiple-trait herd cluster model for international dairy sire evaluation. *J. Dairy Sci.* 83, 815–821.