

Simultaneous Estimation of Genetic Parameters and Heterogeneous Variances for Production Traits of Holstein Cattle in Japan

K. Hagiya^a, Y. Atagi^a, T. Shirai^a, M. Suzuki^b

^aNational Livestock Breeding Center, Fukushima 961-8511, Japan

^bDepartment of Food Production Science, Obihiro University of Agriculture and Veterinary Medicine, Obihiro 080-8555, Japan

Introduction

The existence of heterogeneous phenotypic variances for production and type traits of dairy cows due to genetic, permanent environmental and other environmental factors has been well documented (e.g. Everett *et al.*, 1982 ; Kawahara *et al.*, 2001). Heterogeneity in herd year variances could cause bias in breeding value estimates, and several methods have been investigated to adjust it.

Everett *et al.* (1982) suggested cows in herds with greater variance had a higher chance to have elite status than those in herds with lower variance. Everett and Keown (1984) proposed a method to adjust heterogeneous variances across contemporary groups applying a log transformation. They reported that a log transformation of the data could remove large part of a relationship between mean yields and variance in the herd; however, the log transformation leads to negative correlations between herd variance and herd production level.

Wiggans and VanRaden (1991) developed a procedure to account for the differences in phenotypic variances for milk production across time, region, herd and parity. Gianola *et al.* (1992) described a procedure to estimate herd variances using empirical Bayes Method. Weigel and Gianola (1993) reported a computational method based on Bayesian procedure to adjust variances within and across subclass. Meuwissen *et al.* (1996) suggested that it is important to consider covariances across genetic relationships and reduction of variances due to selection, and they presented a method to estimate breeding values accounting for phenotypic heterogeneous variance. Their method seems to be more appropriate at the view of theoretical prediction and has been accepted in some national genetic evaluation.

The purpose of this study is to estimate heritabilities, repeatabilities and heterogeneous variances of herd-year-parity-milking frequency (HYPF) subclass simultaneously for milk, fat and protein yields of Holstein cattle in Japan by Meuwissen's method applying Method R.

Materials and Methods

Data

Data were 5,552,004 lactation records from 1st to 5th calving on 2,066,992 Holstein cows between January 1985 and August 2004. The pedigree information containing 2,742,992 animals was also used.

Models

Heterogeneous variances were assumed in HYPF and corrected applying the autoregressive model (Meuwissen *et al.*, 1996). Genetic parameters were estimated applying a following model:

$$y_{ijklm} = (HYPF_i + RMY_j + A_k + u_l + pe_l + e_{ijklm}) \exp(\gamma_i / 2)$$

where y_{ijklm} is milk, fat and protein yields of the lactation record for cow l in $HYPF_i$, region-month at calving-year j , and age group k ; $HYPF_i$ is the fixed effects of herd-year-parity-milking frequency i ; RMY_j is the fixed effects of region-calving month-year j ; A_k is the fixed effects of age group k ; u_l and pe_l are the additive genetic effects and permanent environmental effects of cow l ; and e_{ijklm} is random residuals associated with each record. Missing parents were grouped as phantom parents, according to birth year, origin and selection paths. Heritabilities and

repeatabilities were assumed homogeneous across HYPF. Heterogeneous variances were considered in the multiplicative mixed model by Kachman and Evrett (1993). The autoregressive model for γ_i containing fixed and random effects were:

$$\gamma_i = S_{1i}\beta_1 + S_{2i}\beta_2$$

where β_1 is the overall mean ; β_2 is the random effects of each HYPF for heterogeneous variances ; and S_{1i} and S_{2i} are known incidence matrices. The effects of HYPF were correlated within herd among consecutive years. For example, if the block diagonal matrix H contains 2 herd-parity-milking frequencies (HPF) and 3 consecutive years. Then,

$$H = \begin{bmatrix} 1 & \rho & \rho^2 & 0 & 0 & 0 \\ \rho & 1 & \rho & 0 & 0 & 0 \\ \rho^2 & \rho & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho & \rho^2 \\ 0 & 0 & 0 & \rho & 1 & \rho \\ 0 & 0 & 0 & \rho^2 & \rho & 1 \end{bmatrix}$$

where ρ is correlation factor among consecutive years within the HPFs. The inverse of matrix H was obtained by the algorithm outlined by Wade and Quaas (1993). The right hand side of the mixed model equation for γ_i is represented by

$$\begin{bmatrix} S'_1S_1 & S'_1S_2 \\ S'_2S_1 & S'_2S_2 + \frac{1}{\sigma_{hypf}^2}H^{-1} \end{bmatrix}$$

The inverse of above matrix is obtained within herds because the random part of the coefficient matrix is tridiagonal and thus easily absorbed the fixed part of the matrix.

Estimation of parameters to account for heterogeneous variance

The correlation ρ and σ_{hypf}^2 for heterogeneous variance were estimated as follows :

$$\begin{aligned} \hat{\sigma}_{hypf(i)}^2 &= (\hat{\beta}'_{2i}\hat{\beta}_{2i} + trace(C_{i,i}))/c_{i,i}, \\ Cov(hypf_i, hypf_{i+1}) &= (\hat{\beta}'_{2i}\hat{\beta}_{2(i+1)} + trace(C_{i,i+1}))/c_{i,i+1}, \\ \hat{\rho}_{(i)} &= Cov(hypf_i, hypf_{i+1})/\sqrt{\hat{\sigma}_{hypf(i)}^2\hat{\sigma}_{hypf(i+1)}^2}. \end{aligned}$$

where $C_{i,i}$ ($C_{i,i+1}$) is elements of year i in the inverse of the mixed model equation for γ_i ; and $c_{i,i}$ ($c_{i,i+1}$) is the number of the HYPF in year i ($i+1$). The estimates $\hat{\sigma}_{hypf}^2$ and $\hat{\rho}$ are averages of $\hat{\sigma}_{hypf(i)}^2$ and $\hat{\rho}_{(i)}$ over the years.

Method R

Heritabilities and repeatabilities were simultaneously estimated using Method R. The R_v was represented by

$$R_v = \frac{\hat{u}'A^{-1}\hat{u}_p}{\hat{u}'_pA^{-1}\hat{u}_p}$$

where R_v is the linear regression of solutions estimated from all records ($\hat{u}'$) on solutions estimated from partial records ($\hat{u}'_p$). A^{-1} is the inverse of the additive genetic relationship matrix. If assumed variance ratios for the variance of the genetic effect or permanent environmental effect are true, estimated R_v is equal to 1. This method does not require the inversion of the coefficient matrix. Therefore, Method R permits the estimation of genetic parameters for large data files. When the R_v value was not close to 1, the new estimates of variance ratios were obtained by the algorithm of Misztal *et al.* (1997). Reverter *et al.* (1994) reported that the algorithm worked best when half of the records were randomly selected as a partial record. In this study, the estimates were obtained from three different partial records (subset 1 to 3) which were sampled randomly. The estimates of genetic parameters with correction of heterogeneous variances were compared with the estimates from the same data without the correction.

Iteration scheme

Following iteration scheme was empirically found to give fast convergence, and used in this study:

Step 1. Perform rounds to estimate breeding values, but $\hat{\beta}_1$ and $\hat{\beta}_2$ are not estimated until the conversion criterion $\sum_i (\hat{u}_i - \hat{u}_{i-1})^2 / \sum_i \hat{u}_i^2 < 10^{-5}$, where $\hat{u}_i$ and $\hat{u}_{i-1}$ denote estimated breeding values at the current and previous round of iteration.

Step 2. Update the estimates $\hat{\sigma}_{hypf}^2$ and $\hat{\rho}$ when $(\hat{\sigma}_{hypf(j)}^2 - \hat{\sigma}_{hypf(j-1)}^2) / \hat{\sigma}_{hypf(j)}^2 < 10^{-4}$, where $\hat{\sigma}_{hypf(j)}^2$ and $\hat{\sigma}_{hypf(j-1)}^2$ are the estimates of σ_{hypf}^2 at current and previous rounds of iteration.

Step 3. Update the variance ratios when a difference between current and previous R_v becomes less than 10^{-5} . If the R_v is close to 1 ($0.9998 < R_v < 1.0002$), stop iteration ; otherwise, go to step 1.

Results and Discussion

Estimates of genetic parameters

Calculation of β_1 and β_2 had started around 200 rounds of iteration. The convergence occurred within 2000 rounds for genetic parameters. The estimates of heritability and repeatability for milk, fat and protein yields are shown in table 1. The estimates of heritability or repeatability were not greatly different across the three subsets. The parameter estimates without correction of heterogeneous variances estimated from the three subsets were on average 0.307, 0.289 and 0.255 for heritabilities, and 0.501, 0.484 and 0.470 for repeatabilities of milk, fat, and protein yields, respectively. The estimates of heritabilities and repeatabilities were slightly lower than those estimates using records of Japan from 1975 to 1999 of Pereira *et al.* (2001). When heterogeneous variances of HYPF were corrected, heritability and repeatability estimates were higher than those without

considering heterogeneous variances. However, the ratios of permanent environmental and residual variances on phenotypic variances decreased when heterogeneous variance were considered (Table 2). The results suggest that the increasing in estimated heritabilities after considering heterogeneous variances was induced by the decreasing of permanent environmental and residual variances.

Estimates of the parameters for heterogeneous variances

The estimates of autoregressive correlation ($\hat{\rho}$) and variances of HYPF ($\hat{\sigma}_{hypf}^2$) are shown in Table 3. The estimates $\hat{\rho}$ were close to 1, suggesting that the variances within HYPF could be considered as a homogeneous. The estimates $\hat{\sigma}_{hypf}^2$ had increased gradually across years. Further study is required to account for the difference in breeding value calculation. Incomplete records in 2004 could result in relatively higher estimates of σ_{hypf}^2 in this year. Differences in the estimates $\hat{\rho}$ and $\hat{\sigma}_{hypf}^2$ were small ($< 10^{-3}$) among three subsets for Method R. The estimates $\hat{\rho}$ and $\hat{\sigma}_{hypf}^2$ for fat were relatively higher than those of milk and protein. The coefficient of variation of the heterogeneous variances was obtained from mean $\hat{\sigma}_{hypf}^2$ of 0.091 for milk yield:

$$\sqrt{(\exp(\hat{\sigma}_{hypf}^2) - 1)} = \sqrt{(\exp(0.091) - 1)} = 31\%$$

The estimates was slightly lower than those of 33% for the data from the Netherlands in which Friesians, Holsteins and Maas-Rijn-Yssel cattle were in the same herds (Meuwissen *et al.*, 1996). On the other hand, most herds in Japan do not contain two or more breeds. When some breeds are in one herd, the differences among breeds might increase the variance of a production trait in the herd. This could be the reason for slightly lower estimates.

Conclusions

When heterogeneous variances in HYPF were corrected, the estimates of heritabilities and

repeatabilities were higher than those without considering heterogeneous variances. On the other hand, the ratios of permanent environmental variances as well as residual variances on phenotypic variances decreased. The estimated σ_{hypf}^2 had increased gradually across years. The estimates $\hat{\rho}$ and $\hat{\sigma}_{hypf}^2$ of fat were relatively higher than those of milk and protein. Further study is required to account for the heterogeneous variance efficiently in routine genetic evaluations.

Acknowledgements

The authors wish to thank Dr. Shogo Tsuruta at the University of Georgia for his valuable suggestions. We also thank Dr. Yoshinori Terawaki at the Rakuno Gakuen University for his helpful comments. Supply of data by Mr. Mitsuo Aihara in The Livestock Improvement Association Japan and Ms. Yuko Kawamura in The Holstein Cattle Association of Japan is gratefully acknowledged.

References

- Everett, R.W., Keown, J.F. & Taylor, J.F. 1982. The problem of heterogeneous within herd error variances when identifying elite cows. *J. Dairy Sci.* 65 (Suppl. 1), 100.
- Everett, R.W. & Keown, J.F. 1984. Mixed model sire evaluation with dairy cattle-experience and genetic gain. *J. Anim. Sci.* 59, 529-541.
- Gianola, D., Foulley, J.L., Fernando, R.L., Henderson, C.R. & Weigel, K.A. 1992. Estimation of heterogeneous variances using empirical Bayes methods. *J. Dairy Sci.* 75, 2805-2823.
- Kachman, S.D. & Everett, R.W. 1993. A multiplicative mixed model when the variances are heterogeneous. *J. Dairy Sci.* 76, 859-867.
- Kawahara, T., Suzuki, M., Atsumi, T., Hagiya, K. & Gotoh, Y. 2001. The Corrections for Heterogeneous Phenotypic Variance due to Herd, Round, Parity and Classifier on Type Records of Japanese Holstein. *Jpn. J. Zootech. Sci.* 72(9), 303-312.
- Misztal, I. 1997. Estimation of variance components with large-scale dominance models. *J. Dairy Sci.* 80, 965-974.
- Meuwissen, T.H.E., De Jong, G. & Engel, B. 1996. Joint estimation of breeding values of heterogeneous variance of large data files. *J. Dairy Sci.* 79, 310-316.
- Pereira, J.A.C., Suzuki, M., Hagiya, K., Yoshizawa, T., Tsuruta, S. & Misztal, I. 2001. Method R estimates of heritability and repeatability for milk, fat and protein yields of Japanese Holstein. *Jpn. J. Zootech. Sci.* 72, 372-377.
- Reverter, A., Golden, B.L., Bourdon, R.M. & Brinsk, J.S. 1994. Method R variance components procedure : application on the simple breeding value model. *J. Anim. Sci.* 72, 2247-2253.
- Van der Welf, J.H.J., Meuwissen, T.H.E. & de Hong, G. 1994. Effects of correction for heterogeneity of variance on bias and accuracy of breeding value estimation in Dutch dairy cattle. *J. Dairy Sci.* 77, 3174-3184.
- Wade, K.M. & Quaas, R.L. 1993. Solutions to a system of equations involving a first-order autoregressive process. *J. Dairy Sci.* 76, 3026-3032.
- Weigel, K.A. & Gianola, D. 1993. A computationally simple Bayes method for estimation of heterogeneous within-herd phenotypic variance. *J. Dairy Sci.* 76, 1455-1465.
- Wiggans, G.R. & VanRaden, P.M. 1991. Method and effect of adjustment for heterogeneous variance. *J. Dairy Sci.* 74, 4350-4357.

Table1. Heritability (h^2) and repeatability (r) estimates.

Trait	Subse	1		2		3		Mean (SD)	
		h ²	r	h ²	r	h ²	r	h ²	r
No correction of heterogeneous variance									
Milk		0,306	0,499	0,308	0,501	0,306	0,499	0,307 (0,002)	0,500 (0,001)
Fat		0,289	0,483	0,290	0,484	0,289	0,483	0,289 (0,001)	0,484 (0,001)
Protein		0,255	0,470	0,255	0,470	0,255	0,473	0,255 (0,000)	0,471 (0,001)
Correction of heterogeneous variance									
Milk		0,323	0,509	0,324	0,511	0,323	0,509	0,323 (0,001)	0,510 (0,001)
Fat		0,304	0,493	0,306	0,495	0,304	0,493	0,305 (0,001)	0,493 (0,001)
Protein		0,265	0,476	0,266	0,477	0,266	0,479	0,266 (0,001)	0,477 (0,001)

Table2. Ratios of genetic, permanent environmental and residual variance on phenotypic variance.

Trait	Ratio of each variance on phenotypic variance		
	Genetic	Permanent environmental	Residual
No correction of heterogeneous variance			
Milk	0.307	0.193	0.501
Fat	0.289	0.195	0.516
Protein	0.255	0.216	0.529
Correction of heterogeneous variance			
Milk	0.323	0.186	0.490
Fat	0.305	0.189	0.507
Protein	0.266	0.212	0.523

Table3. Estimates of autoregressive correlation and variance of herd-year-parity-milking frequency (HYPF).

Year	Milk			Fat			Protein		
	Number of HYPF	$\hat{\rho}$	$\hat{\sigma}_{hypr}^2$	Number of HYPF	$\hat{\rho}$	$\hat{\sigma}_{hypr}^2$	Number of HYPF	$\hat{\rho}$	$\hat{\sigma}_{hypr}^2$
1985	26,521	0.938	0.045	26,521	0.946	0.057	4,818	0.929	0.057
1986	26,856	0.952	0.048	26,856	0.959	0.060	6,292	0.899	0.061
1987	26,892	0.960	0.051	26,892	0.966	0.065	14,017	0.915	0.054
1988	27,817	0.964	0.055	27,817	0.969	0.069	22,387	0.940	0.055
1989	28,334	0.967	0.059	28,334	0.972	0.073	26,579	0.953	0.058
1990	28,117	0.971	0.063	28,117	0.974	0.079	27,177	0.960	0.062
1991	27,312	0.974	0.069	27,312	0.977	0.085	26,849	0.965	0.067
1992	26,498	0.976	0.074	26,498	0.979	0.090	26,241	0.970	0.073
1993	25,477	0.978	0.079	25,477	0.981	0.097	25,343	0.972	0.080
1994	23,794	0.981	0.086	23,794	0.983	0.104	23,736	0.976	0.087
1995	22,800	0.982	0.092	22,800	0.984	0.111	22,774	0.978	0.095
1996	21,828	0.984	0.100	21,828	0.985	0.119	21,817	0.980	0.103
1997	20,930	0.985	0.107	20,930	0.986	0.126	20,922	0.981	0.111
1998	20,236	0.986	0.113	20,236	0.987	0.133	20,232	0.983	0.118
1999	19,653	0.987	0.119	19,653	0.988	0.141	19,653	0.984	0.126
2000	19,241	0.988	0.126	19,241	0.989	0.148	19,239	0.985	0.135
2001	18,842	0.988	0.130	18,842	0.990	0.155	18,842	0.986	0.141
2002	18,774	0.988	0.131	18,774	0.990	0.157	18,774	0.987	0.143
2003	18,565	0.983	0.128	18,565	0.986	0.155	18,565	0.982	0.140
2004	12,443		0.145	12,443		0.176	12,443		0.156
Mean	23,047	0.975	0.091	23,047	0.978	0.110	19,835	0.964	0.096