

# Estimation of Breeding Values and Heterogeneous Variance in French Dairy Evaluation : First Report

*Christèle Robert-Granié and Bernard Bonaiti*

*INRA, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas, France*

---

## Abstract

A heterogeneous variance model assuming heterogeneous residual variances and known constant ratios was implemented and applied to the analysis of milk yield in the French Holstein population. The linear model for the log residual variance included a region-year fixed effect and a herd-year random effect with a within-herd autocorrelation. The estimate of this correlation coefficient reached 0.50. Residual standard deviation doubled in 20 years, whereas differences across regions were more limited and decreased over time. This model was compared with the homogeneous model and with the multiplicative model of Meuwissen *et al.* (1996). On a first approach, both heterogeneous variance models provided very similar results. Accounting for heterogeneous variances had limited consequences on AI bull ranking. However, the important change in within birth year EBV variability in the past would affect international comparisons.

---

## 1. Introduction

The traditional genetic evaluations rely upon a strong assumption of constant dispersion parameters. This assumption may be unrealistic when environmental conditions are heterogeneous or when the data span a long time period (24 years in the French evaluation). Several sources of heterogeneous variances are identified. The most important one is probably the increase in phenotypic variance as a scale effect, with the increase of production level. Within year, the phenotypic variance varies with herds, and particularly with the proportion of short completed lactations, the proportion of in-progress lactations, with production level, and with management practices. On average, heritability is known to increase with production level (Bonaiti, 1982), but its relationship with phenotypic variability is unclear.

Several evaluation systems (Wiggans & VanRaden, 1991) use a preadjustment for within herd heterogeneous phenotypic variance. However, because heterogeneity of observed data may depend on an heterogeneity of age, month of calving,

culling rate, mating or selection practices, preadjustment factors may be inadequate, and a simultaneous estimation of location and dispersion parameters is preferable.

The multiplicative model of Meuwissen *et al.* (1996) is presently the only national evaluation system with such a joint estimation of location and dispersion parameters. It relies on the assumption that all fixed and random location parameters are affected in proportionality with residual standard deviation. Moreover, it assumes constant variance ratios.

The validity of this multiplicative model, and particularly the quality of the fit, is still unknown. Its most appealing property is its computational ease because variance adjustment only affects the right-hand side of the equation system, while the left-hand side remains unchanged. In this paper, we present a more classical model where random effects are normally distributed with heterogeneous variances (Robert, 1996; Foulley, 1997). Both models are compared to each other and to the homogeneous model, on the basis of the analysis of the French Holstein data set.

## 2. Model of Analysis

The model could be written as follows

$$y_i = X_i\beta + t_1\sigma_{ei}Z_iQg + t_1\sigma_{ei}Z_iu + t_2\sigma_{ei}W_i p + e_i$$

with

- $y_i$  being the vector of performance in herd-year  $i$ ,
- $\beta$  being the fixed effects of herd-year, month of calving, age at calving, and days dry (the latter three effects being nested within year, region and parity),
- $g$  being the vector of genetic group effects
- $u$  being the vector of standardized breeding values, normally distributed  $N(\mathbf{0}, \mathbf{A})$
- $p$  being the vector of standardized permanent environmental effects, normally distributed  $N(\mathbf{0}, \mathbf{I})$
- $e_i$  being the vector of residuals in herd-year  $i$ , normally distributed  $N(\mathbf{0}, \mathbf{D}\sigma_{ei}^2)$ , where  $\mathbf{D}$  is a diagonal matrix containing the inverse of the weight of each record
- $t_1$  and  $t_2$  being two known constants function of the repeatability and the heritability.

Variances were assumed to vary according to herd, year, and region :

$$\log(\sigma_{ei(j)}^2) = \lambda_j + v_i$$

with

- $\lambda_j$  being the fixed effect of year-region  $j$  of herd-year  $i$ , and
- $v_i$  being the random effect of herd-year  $i$ .

Herd-year effects on variance were assumed to be correlated within herd assuming an autoregressive model. The correlation  $\rho$  and the variance of herd-year effects  $\sigma_v^2$  were estimated as in Meuwissen *et al.* (1996).

As REML appeared to be computationally unaffordable, a ML algorithm was used to estimate residual variances, as suggested by Foulley (1997).

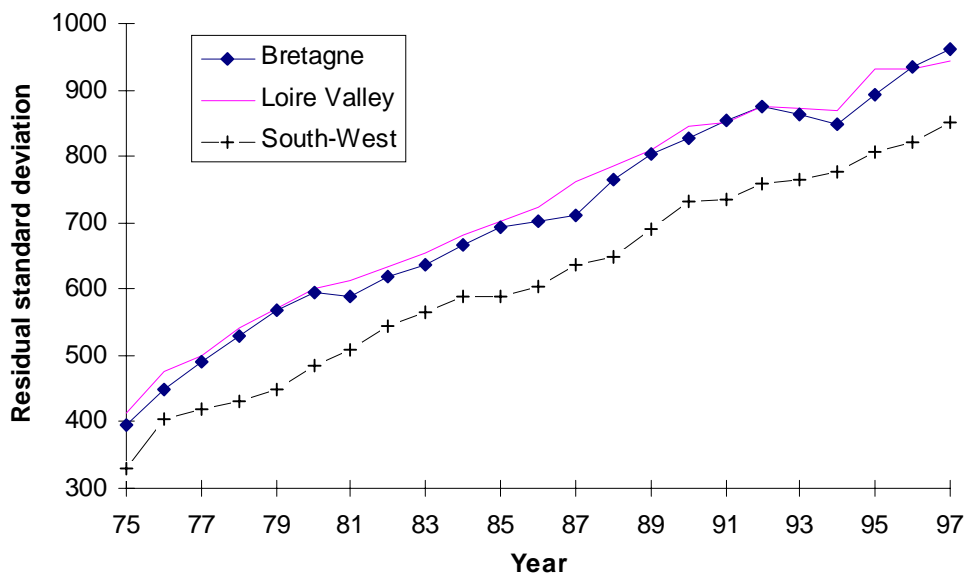


Figure 1. Evolution of residual standard deviation in three different regions.

Three models derived from the model used in the national evaluation (same fixed effects and genetic group definition, same genetic parameters) were used to analyze milk yield recorded in the French Holstein population from 1974 to 1997 :

(1) the homogeneous model,

$$y = X\beta + ZQg + Zu + Wp + e$$

the multiplicative model of Meuwissen et al. (1996),

$$y_i = (X_i\beta + Z_iQg + Z_iu + W_i p + e_i) \sigma_{ei}$$

(3) the heterogeneous model with constant ratios,

$$y_i = X_i\beta + t_1\sigma_{ei}Z_iQg + t_1\sigma_{ei}Z_iu + t_2\sigma_{ei}W_i p + e_i$$

In the multiplicative model, the  $\rho$  and  $\sigma_v^2$  coefficients are set to the estimated value issued from heterogeneous model.

In these three analyses, foreign information of imported bulls was not considered. This point is the only difference between French official evaluation and the homogeneous model.

### 3. Results

#### *Distribution of variances*

Table 1 shows the parameter estimates of the herd-year variance distribution for the heterogeneous model. The correlation estimate ( $\rho$ ) is moderate (0.49), and not as high as in Meuwissen *et al.* (1996), suggesting some variance heterogeneity within herd across year. In practice, with such a value of  $\rho$ , the estimate of each herd-year variance is mostly influenced by the data spanning a 3-year period, whereas in the Dutch situation all data strongly contributed to estimate any herd-year variance. Parameter  $\lambda$  increased with year,

resulting in the doubling of residual standard deviation in the last 20 years (Figure 1). Within year, variances varied across regions, but to a smaller extent (Figure 1).

Table 1. Estimates of the parameters of the herd-year variance distribution ( $\rho$ ,  $\sigma_v^2$ , and  $\lambda$ ).

Year	$\lambda$ (1)	$\sigma_v^2$	$\rho$
1975	11.94	0.49	0.61
1980	12.69	0.34	0.52
1985	12.98	0.29	0.49
1990	13.33	0.29	0.48
1995	13.52	0.26	0.45
Mean (2)	12.99	0.32	0.49

(1) pooled estimate over regions

(2) average over the 1975-1997 period

Models 2 and 3 give very close residual standard deviations for each year-region. Residual deviation is slightly larger (about 20 kg) with multiplicative model.

#### *Effect of heterogeneous variances on AI bull EBVs*

The evolution of EBV standard deviation with bull birth year and the estimated genetic trend are presented in Table 2. In order to facilitate comparison between models, results are transformed so that the three methods provide similar mean and standard deviation for bull born between 1987 and 1991. Whereas within-birth year genetic variability seemed to remain constant or slightly decrease over time with the homogeneous model, it clearly decreased with multiplicative and heterogeneous models. As in the first period (70-79), multiplicative and heterogeneous models give higher genetic trends (115 and 114 versus 86 kg/year), the trends are equivalent in the second period (80-90) :120, 123 and 120 kg/year.

Table 2. Evolution of AI bull EBV according to the method (milk yield in kg).

Birth year	# bulls	Mean			Standard Deviation		
		-1	-2	-3	-1	-2	-3
70	227	-1914	-2267	-2216	510	778	780
71	285	-1840	-2171	-2121	536	794	795
72	322	-1644	-1882	-1833	621	866	867
73	357	-1647	-1878	-1831	561	775	776
74	324	-1584	-1798	-1751	606	817	814
75	343	-1616	-1844	-1798	642	861	859
76	340	-1455	-1629	-1587	655	862	861
77	363	-1341	-1472	-1431	622	799	798
78	454	-1102	-1168	-1128	658	829	828
79	494	-1062	-1112	-1076	605	748	747
80	433	-1133	-1194	-1160	609	730	729
81	573	-929	-951	-925	604	700	699
82	731	-805	-814	-793	614	682	681
83	720	-648	-636	-619	627	670	668
84	332	-697	-685	-673	547	585	585
85	530	-529	-515	-506	602	633	634
86	549	-364	-355	-348	603	620	620
87	640	-205	-203	-199	576	597	597
88	613	-30	-27	-26	595	611	611
89	604	97	98	100	588	593	593
90	748	138	139	139	554	542	544
91	653	291	283	278	601	578	579

- (1) homogeneous model
- (2) multiplicative model
- (3) heterogeneous model

Table 3 shows the proof correlations observed between methods. Within birth year, these correlation are very high, showing that adjusting for heterogeneous variances have limited impact on AI bull EBVs when progeny are well distributed in many herds. However, the overall correlation between the homogeneous model on one hand, and both heterogeneous models on the other hand, shows more important reranking across years, due to changes in genetic variability and genetic trends. Both heterogeneous variance models provided virtually the same results in term of AI bull EBV.

Table 3. Proof correlations among methods (within year correlation below diagonal, overall correlation above diagonal).

	(1)	(2)	(3)
(1)	-	0.992	0.992
(2)	0.995	-	1.000
(3)	0.995	1.000	-

(1), (2), (3) see table 2.

Multiplicative and heterogeneous models give slightly lower results for US and Canadian imported bulls (Table 4). Some bull EBV substantially vary between models. Heterogeneous

model produce larger differences than multiplicative model. But these differences are not high enough to account for discrepancy between US and French EBV.

Table 4. Comparison of breeding values for progeny tested bulls and imported bulls (bulls born between 1986 and 1991).

	# bulls	Multiplicative Model (D2)			Heterogeneous Model (D3)		
		avg diff.	std. dev. diff	max. diff.	avg diff.	std. dev. diff	max. diff.
French bulls	3807	1	58	319	2	64	342
Us imported bulls	50	-27	88	203	-78	105	306
Canadian imported bulls	18	-22	55	128	-63	53	166

D2 : EBV with correction for multiplicative model minus EBV with correction for homogeneous model

D3 : EBV with correction for heterogeneous model minus EBV with correction for homogeneous model

#### 4. Conclusion

According to these preliminary results, both heterogeneous variance models provide quite similar results, at least for AI bulls. On a first approach, AI bull EBV are hardly affected by accounting for heterogeneous variances. The major consequences are expected on female EBV on one hand, and on international comparison on the other hand, due to large change in estimated genetic variability. Further investigation is needed to compare the predictive ability of each model and the fit of the data. Other models (Foulley *et al.*, 1998), will be investigated in order to study the variation of heritability and repeatability as fonction of year, region and production level.

#### References

Bonaiti, B. 1982. Interaction genotype-milieu sur la production laitière chez les bovins II Corrélation génétique entre niveaux de production et variation de l'héritabilité. *Ann. Génét. Sel. Anim.* 14, 441-452.

Foulley, J.L. 1997. ECM approach to heteroskedastic mixed models with constant variance ratios. *Genet. Sel. Evol.* 29, 297-318.

Foulley, J.L., Quaas, R.L. & Thaon d'Arnoldi, C. 1998. A link function approach to heterogeneous variance components. *Genet. Sel. Evol.* 30.

Robert, C. 1996. Etude de quelques problèmes liés à la mise en oeuvre du REML en génétique quantitative. *Thèse de Doctorat*, Université Paul Sabatier, Toulouse, France.

Meuwissen, T.H.E., De Jong, G. & Engel, B. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci.* 79, 310-316.

Wiggans, G.R. & VanRaden, P.M. 1991. Method and effect of adjustment for heterogeneous variances. *J. Dairy Sci.* 74, 4350-4357.