# Experience with Correction for Heterogeneous Variances in Milk Production Traits in the Netherlands

**_G. De Jong_**
*Dutch Cattle Syndicate (NRS)*
*P.O. Box 454, 6800 AL Arnhem,The Netherlands*
*E-mail: jong@nrs.nl*

## Introduction

In national genetic evaluations for milk production traits lactation records or testday records are used which are collected in a time span of 10 to 30 years. In this period a lot of changes occur. The production level of the cow population increases over time due to improvement of management and increase of genetic potential. This increase of production level also affects the variation found in the lactation records; an increase of the mean of a trait goes together with an increase of variation of the trait.

In a regular genetic evaluation the mean is adjusted for via several fixed effects (eg. month of calving, age, parity etc.). Usually when estimating breeding values for a trait, this trait is considered to have as constant variance across time and across herds. In real data the variance of a trait can, besides production level, be affected by herd (effect of management, housing, feeding), region, parity, lactation length etc.Therefore one should also adjust for the heterogenity of variance. This avoids for example the effect of having more chance of getting high cowindexes in a high yielding herd than in a low yielding herd. Bull dams are mainly found in high yielding herds with high phenotypic variance. Further the pedigree indexes of young AI test bulls appeared to be overestimated with about a quarter of a genetic standard deviation, which is mainly due to an overestimate of the dam=s index.

This paper shows the method applied in the Dutch genetic evaluation to correct for heterogenous variance and the effects on estimated breeding values for different groups of animals.

## Method

In the method applied in the Netherlands, the correction for heterogenity occurs during the breeding value estimation procedure. Simultaneously the mixed model for breeding values and the model heterogeous variance are solved.

The animal model as used for the national genetic evaluation is as follows:

$$Y_{ijklm} = HYS_i + CM_j + PERM_{kl} + A_k + error$$

where:

| | | |
|---|---|---|
| $Y_{ijklm}$ | : | 305-days lactation for milk, fat or protein yield, with precorrection for:<br>- age at calving;<br>- days open during the lactation;<br>- heterosis and recombination; |
| $HYS_i$ | : | herd-year-season-parity class i, in which the lactation is compared with lactations of herdmates; |
| $Cm_j$ | : | calving month*year j, when the lactation was started; |
| $PERM_{kl}$ | : | permanent environment effect kl of cow k in herd l (repeatability=0.55); |
| $A_k$ | : | breeding value of cow k (heritability=0.35). The family relationship matrix is added, following the procedures of an animal model; |
| error | : | error term of observation $Y_{ijklm}$ |

For the correction for heterogeneous variance data are assumed to be homogeneous within strata and heterogeneous across strata: a herd-year class represents a stratum here. Records in herd-year i are represented by:

$$Y_i \;=\; (X_i b + Z_i u + e_i)\;\exp(\gamma_i/2)$$

where

b = fixed effect vector
u = vector of breeding value
$e_i$ = vector of environmental effects, and
$X_i$ and $Z_i$ = design matrices for fixed effects and breeding value

The $\gamma_i$ can further be described as:

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

where scalar $S_{1i}$ = 1 and, and row vector $S_{2i}$ has a 1 at position i for herd-year i and 0 elsewhere. The herd-year effects are assumed to be correlated within hers according to an autoregressive model (Wade, 1993). For example, if a herd contains 4 herd-years, the variance of the herd-year effects pertaining to that herd is:

$$\sigma^2_{hy} \quad \begin{vmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{vmatrix}$$

where ρ = correlation between consecutive years within a herd. Further full description of the model can be found in Meuwissen et al. (1996).

## Results

Extra variation can occur in a herd when animals orginating from two or more breeds are present in a herd, which is common in the Dutch situation in which Friesians, Holsteins and Maas-Rijn-Yssel cattle are kept in the same herds. The method of Meuwissen considers such breed effects, whereas regular usde methods as described by Hill (1984) and Wiggans (1991) do not. If estimation of heterogeneous variance does not consider breed effects, the estimates of variances within herd-yaers-parities are inflated. Subsequent correction for heterogeneous variance thus unjustly reduces differences between breeds or genetic groups. This reduction applies to superior animals that are imported into populations. Because the correction for heterogeneous variance is too large, animals may be underestimated and have a reduced probability of selection. The same underestimation may apply to their crossbred offspring and, to a lesser extent, to animals from superior genetic groups.

Another effect which is taken care by the method is the reduction in variance due to selection. Further covariances due to genetic relationships are recognized as correction for heterogeneous variances and breeding value estimation are done simultaneously.

A test to check how the correction method works is to compare parent average (PA) of progeny tested bulls with the realized breeding value (RBV) of these bulls. When comparing the result of no correction for hetergeneous variance and applying the correction method (table 1) it is obvious that the average PA is higher than the average realized breeding values. The difference between average PA and average realized breeding value is smaller (40%). The major correction in the PA is the breeding value of the dam. The dams= breeding values appear to be overestimated while sires of bulls do not show differences for both methods, with or without correction for heterogeneous variances. On average dams are estimated 60 to 80 kg milk higher without correction for heterogeneous variance than with this correction.

The average breeding value of progeny tested bull in the Netherlands (having a first crop of daughters) are 20 kg milk higher with correction than without correction (table 2). Import bulls (with only second crop daughters in the BV) have, on average, a 20 kg lower breeding value for milk yield with correction compared to without correction. The absolute difference (about 55 kg milk) and standard deviation of differences (about 70 kg milk) are higher for import bulls than for progeny tested bulls. Progeny tested bulls showed an absolute difference of breeding values of 30 kg milk and a standard deviation of 30 kg milk for both methods. The changes for import bull are twice as large than for progeny tested bulls, although differences are relatively small.

The introduction of correction for heterogeneous variances has more influence on the estimated breeding values for cows than for bulls. Table 3 show the change in breeding value for milk yield and Inet (fl) for cows born in 1992, classified per

Inet class. It is obvious that cows having a high Inet based on the system without correction for heterogeneous variances showed by far the largest change. Bull dams are spread in more herds and are not only found in herds with a high production level after applying the correction for heterogeneous variance.

In the Dutch data set a clear relationship between herd production level and herd variance is found. When the production level increases higher herd variance is estimated. Further larger herds tended to have a higher variance. Regional differences are also found as shown in Table 4.

The standard deviation of breeding values for milk yield for tested bulls across year of birth is more stable when correction for heterogeneous of variance is applied than without (Table 5). Milk yield trait as measured in old data is made more comparable with the milk yield trait as measured in later years, since not only adjustment is made for the mean but also for variance. Therefore breeding values of older animals are made more comparable with breeding values of younger animals.

## Conclusions

The method for correction for heterogeneous variances as described by Meuwissen (1996) is a method which corrects observations for difference in herd variances, while taking into account genetic effects on the phenotypic herd variance. Further this method is rather flexible in modelling; easily different effects which affect the variance can be added to the model. Further the method is applicable for large data sets.

After appliying the correction for heterogeneous variances small effect are found on estimated breeding values for bulls. The correction has more effect on estimated breeding values for cows. The parent average is a better predictor for a test bulls´ estimated breeding value (based on daughters) when correction for heterogeneous variances is applied than without the correction. Breeding values of older animals are made more comparable with breeding values of younger animals as not only an adjustment is made for the mean but also for variance.

## References


Hill, W.G., Edwards, M.R., Ahmed, M.-K.A. and Thompson, R. 1983. Heritability of milk yield and composition at different levels and variability of production. *Anim. Prod. 36,* 59.

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

Wade, K.M. and Quaas, R.L. 1993. Solutions to a system of equations involving a first-order autoregressive process. *J. Dairy Sci. 76,* 3026.

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

Table 1. Realized breeding values minus parent average for milk yield for progeny tested bull per year of birth with and without correction for heterogeneous variances.

| Year of birth | number of bulls | without correction | with correction |
|---|---|---|---|
| 81 | 87 | -79 | -59 |
| 82 | 140 | -119 | -89 |
| 83 | 153 | -104 | -69 |
| 84 | 226 | -161 | -115 |
| 85 | 206 | -101 | -62 |
| 86 | 200 | -159 | -109 |
| 87 | 167 | -114 | -47 |
| 88 | 158 | -121 | -54 |
| 89 | 105 | -168 | -100 |
| 90 | 78 | -98 | -38 |
| total | 1520 | -126 (100%) | -78 (60%) |

*bulls have at least 75% Holstein genes and their dams realized a lactation in the Netherlands. The base for the correction is the average variance for all lactations in the data set.

Table 2. Comparison of breeding values for milk yield for progeny tested bulls and import bulls. Difference = BV with correction for heterogeneous variances minus BV without correction (=diff).

| Progeny tested bulls | | | | | | Import bulls | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Year of birth | avg. diff. | avg. abs. diff. | stand. dev. diff | max. diff | n | year of birth | avg. diff. | avg. abs. diff. | stand.dev. diff. | max. diff. | n |
| 85 | 19.3 | 29 | 29.9 | 179 | 371 | 81 | -12.8 | 55 | 74.5 | 202 | 73 |
| 86 | 22.1 | 30 | 30.2 | 153 | 333 | 82 | -12.0 | 49 | 70.0 | 249 | 82 |
| 87 | 24.6 | 33 | 24.1 | 183 | 323 | 83 | -20.9 | 56 | 68.0 | 248 | 61 |
| 88 | 23.1 | 32 | 31.0 | 114 | 384 | 84 | -16.4 | 58 | 76.6 | 244 | 45 |
| 89 | 19.8 | 31 | 34.4 | 157 | 412 | 85 | -18.8 | 56 | 72.1 | 221 | 44 |
| 90 | 18.2 | 27 | 27.9 | 88 | 290 | 86 | -24.3 | 56 | 78.4 | 283 | 53 |

* the base for the correction for heterogeneous variance is determined by cows born in 1990.

Table 3. Difference between breeding values for milk yield for cows born in 1992. Difference = BV with correction for heterogeneous variance minus BV without correction (=diff).

| Inet-<br>class | n | diff kg milk<br>avg | <br>abs | diff Inet<br>avg | <br>abs | <br>std |
|---|---|---|---|---|---|---|
| -50/-1 | 23212 | 4,7 | 24,0 | 2,1 | 8.8 | 12 |
| 0/49 | 28528 | 4.5 | 23.5 | 2.3 | 8.2 | 11 |
| 50/99 | 28638 | 3.5 | 24.4 | 2.3 | 8.6 | 12 |
| 100/149 | 23473 | 1.0 | 26.1 | 1.4 | 9.4 | 13 |
| 150/199 | 15361 | -2.3 | 29.3 | 0.0 | 11.1 | 15 |
| 200/249 | 8000 | -10.1 | 34.4 | -3.0 | 13.9 | 18 |
| 250/299 | 3229 | -21.5 | 43.2 | -8.2 | 18.1 | 22 |
| 300/349 | 1029 | -43.0 | 58.0 | -17.9 | 24.6 | 25 |
| 350/399 | 249 | -71.8 | 79.5 | -29.0 | 32.8 | 26 |
| 400/449 | 51 | -122.4 | 122.4 | -48.5 | 48.9 | 25 |
| 450/499 | 18 | -121.5 | 121.5 | -56.7 | 56.7 | 33 |
| 500/549 | 3 | -234.3 | 234.3 | -76.3 | 76.3 | 25 |

* the base for teh correction for heterogeneous variance is determinde by cows born in 1990.

Table 4. Regional effect on correction factors for heterogeneous variance. LSQ solutions, corrected for herd production level and herdsize.

| Region | correction factor |
|---|---|
| Friesland | -1.9 |
| Groningen | -0.1 |
| Drenthe | -1.7 |
| Overijssel | 0.7 |
| Gelderland | 0.4 |
| South Holland/Utrecht/Zeeland | -1.5 |
| Norht Holland | -1.8 |
| Branbant/Limburg | 5.9 |

Table 5. Standard deviation of breeding values for milk yield for tested bulls per year of birth, estimated with and without correction for heterogeneous variance.

| year of birth | n | without correction | with correction |
|---|---|---|---|
| 72 | 106 | 401 | 457 |
| 73 | 193 | 347 | 422 |
| 74 | 209 | 376 | 451 |
| 75 | 239 | 399 | 490 |
| 76 | 233 | 494 | 555 |
| 77 | 276 | 418 | 510 |
| 78 | 274 | 438 | 511 |
| 79 | 301 | 473 | 537 |
| 80 | 320 | 471 | 565 |
| 81 | 329 | 448 | 520 |
| 82 | 369 | 485 | 532 |
| 83 | 307 | 477 | 520 |
| 84 | 335 | 505 | 540 |
| 85 | 337 | 462 | 480 |
| 86 | 290 | 526 | 530 |
| 87 | 279 | 505 | 507 |
| 88 | 346 | 527 | 518 |
| 89 | 371 | 539 | 529 |
| 90 | 325 | 488 | 482 |
| 91 | 265 | 481 | 471 |

* the base for teh correction for heterogeneous variance is determined by cows born in 1990.