

Illustration of a Trend Validation Test for Longevity Evaluations

Vincent Ducrocq,

Station de Génétique Quantitative et Appliquée, Département de Génétique Animale, Institut National de la Recherche Agronomique, 78352 Jouy en Josas, France
email: vincent.ducrocq@dga.jouy.inra.fr

1. Introduction

In international evaluations, it has been repeatedly found and shown that the ranking of top bulls is quite sensitive to national genetic trends. When the national genetic trends are incorrect, biased evaluations are expected (Bonaiti *et al.*, 1993). To detect inconsistencies, trend validation tests have been proposed, in particular for production traits (Boichard *et al.*, 1995; Interbull, 2004). Interbull requires that national data must pass these tests before being included in an international evaluation. For production traits, the first test (Method I) compares the genetic trend based on the first lactation only with the one obtained using all lactations. The second one (Method II) verifies that the regression coefficient of daughter yield deviations (DYDs) defined within bull x year combinations on the year of use of the sire is not significantly different from 0. The third one (Method III) compares national proofs in successive evaluation runs.

In the case of traits described by nonlinear models, validation of national genetic trend is substantially more complex than for linear models. This is particularly the situation for length of productive life, usually called 'longevity', when evaluated using a survival analysis approach. For longevity, Method I is not applicable. Method III is difficult to implement and interpret because new information in the most recent run does not only come from new young daughters but also from a better knowledge of the actual length of productive life of the old ones. Finally, Method II requires the proper definition of DDs (Daughter deviations) and EDCs (Equivalent Daughter Contribution – which represents the amount of information, i.e., the weight associated to each DD).

We recently proposed a general approach to compute the DDs and EDCs to be used in

Method II for genetic trend validation (Ducrocq *et al.*, 2003). This paper presents its actual implementation for the French genetic evaluation on longevity. The test revealed a large overestimation of the trend in the Holstein breed. We will describe how a survival analysis model leading to a valid genetic trend was constructed.

2. Method

2.1. DD and EDC calculation

Given the complexity of survival analysis models, national genetic evaluations are based on sire-maternal grand-sire models. The DDs and EDCs for longevity are a straightforward by-product of the computation of approximate animal model solutions, as described in Ducrocq (2001) and Ducrocq *et al.* (2001, 2003). This is very briefly outlined here.

Assuming a Weibull proportional hazards animal model, the hazard function $h(t)$ of a cow f is described as:

$$h(t; \mathbf{x}_f(t)) = \rho t^{\rho-1} \exp\{\mathbf{x}_f(t)' \mathbf{b} + a_f\} \quad [1]$$

where $\mathbf{b}$ is a vector of possibly time dependent environmental effects and a_f is the additive genetic value of cow f .

Let cum_f be the cumulative hazard function evaluated at censoring time for a censored cow f ($\delta_f = 0$) or at failure time for an uncensored cow ($\delta_f = 1$). To get an estimate of $\mathbf{b}$ and $\mathbf{a} = \{a_f\}$ in [1], it is necessary to maximize the logarithm of the joint posterior density of all parameters (Ducrocq and Casella, 1996). At the joint mode, the first derivative of the joint posterior density with respect to each a_f is 0, which implies the following equality:

$$\delta_f - \text{cum}_f - \frac{1}{\sigma_a^2} \left(\mathbf{A}^{-1} \mathbf{a} \right)_f = 0 \quad [2]$$

where $\mathbf{A}$ is the relationship matrix and σ_a^2 the additive genetic variance. With the fixed ($\hat{\mathbf{b}}$) and sire effects ($\hat{a}_{\text{sire}}$) estimates obtained from the sire-maternal grand sire model, it is easy to solve [2] for the missing genetic part ($a_f - \frac{1}{2} a_{\text{sire}} - \frac{1}{4} a_{\text{mgs}}$). This is used to reconstruct an approximate estimate $\hat{a}_f$ of the cow genetic effect. Having $\hat{a}_f$, rewrite [2] as:

$$w_f \hat{a}_f + \frac{1}{\sigma_a^2} \left(\mathbf{A}^{-1} \mathbf{a} \right)_f = w_f y_f^* \quad [3]$$

$$\text{with } w_f = \text{cum}_f \quad [4]$$

$$\text{and } y_f^* = \frac{\delta_f}{\text{cum}_f} + \hat{a}_f - 1 \quad [5]$$

Equation [3] is identical to a typical equation of a univariate BLUP animal model, after correction for fixed effects. This suggests a definition of EDCs and DDs required to apply Method II based on the sum of the w_f 's and y_f^* 's over sires and years.

2.2. Validation test

Method II of Boichard *et al.* (1995) relies on the model:

$$DD_{ij} = s_i + t_j + e_{ij} \quad [6]$$

with $\text{var}(e_{ij}) = 1 / \text{EDC}_{ij}$, DD_{ij} is the sum of the cows' pseudo-records y_f^* over all daughters of sire i , having a first calving on year of use m ($j=1, \dots$) of their sire (i.e., usually $j=1$ or 2 for daughters of a young bull, $j>4$ for second crop daughters of national bulls); EDC_{ij} is the sum of the weights w_f 's over the same animals; s_i is the effect of sire i , which was treated as random in Boichard *et al.* (1995) but which will be treated as random here; t_j is the effect of year of use of any sire. When the estimated genetic trend is correct, the expectation of the year of use effect t_j is 0 . Note that, in a survival analysis model, larger sire values correspond to

higher risk of culling on their daughters. Therefore the interpretation of the year of use solutions must be done with care: if, for example, the estimate of t_j decreases with time, this means that environmental effects on culling in more recent years are overestimated and that sire effects on culling are underestimated, or equivalently, that the genetic trend on longevity is over-optimistic.

A variation of this model is:

$$DD_{ij} = s_i + t_{jk} + e_{ij} \quad [7]$$

where t_{jk} is the effect of year of use for sires with first daughters born in a given period k of, say, 2 to 5 consecutive years. This way, it is possible to visualize problems that are more specific to older or more recent bulls, who have substantially different proportions of censored daughters.

Interbull (2004) recommendation for production traits is based on a modification of model [6]:

$$DD_{ij} = s_i + j \beta + e_{ij} \quad [8]$$

i.e., the effect of year of use is replaced by a regression on year of use. The genetic trend is accepted as valid when the absolute value of the estimated regression coefficient β is less than 1% of the genetic standard deviation σ_a for the trait, or equivalently, less than 2% of the sire standard deviation σ_s . The latter measure will be used here. Admittedly, the 1% or 2% limit is quite arbitrary. For the reasons exposed before, a negative β reflects an over-optimistic genetic trend.

3. Application

3.1. Initial data and model

The French routine genetic evaluation of dairy bulls on production-adjusted length of productive life of their daughters has been done twice a year since June 1997, with a release in June and October. The model is described in details in Ducrocq (1999). It includes all cows milk recorded with a calving (up to 10) after December 1, 1984 and daughters of AI bulls

with at least 10 daughters. Length of productive life is the number of days from first calving to the last test date before culling, or for censored cows, to the last test date before either the end of study period, the date of sale to another herd or the end of milk recording in the herd. The model is a Weibull proportional hazard model similar to model [1] with a Weibull baseline hazard function with fixed shape parameter $p=2$ and, on the exponential part in [1]:

- a set of time-independent covariates: sire and maternal grand sire (randomly distributed with a multivariate normal distribution), age at first calving class.

- a set of time-dependent covariates: fixed year-season (2 seasons), stage of lactation, variation in herd size class, within herd-year deviation classes for milk yield (as an interaction with year-season), protein content and fat content and a random, log-gamma distributed herd-year-season with parameter $\gamma=4$). The stage of lactation effect is defined within lactation number (lactations 1, 2, 3, 4, 5, 6+) and with changes t' days after the current calving ($t'=30, 60, 150, 240$ and date when dried). For details, see Ducrocq (1999).

Since 2002, the evaluation has been performed using an improved version of the Survival Kit (Ducrocq and Sölkner, 1998) which uses a full Newton–Raphson approach to maximise the joint posterior density of the parameters, instead of a quasi-Newton approach. This change considerably decreased computational time (divided by >10 in Holstein, with a stricter convergence criterion), at the expense of a large increase in memory requirement. This version also includes the computation of approximate cow solutions as well as the cows' pseudo-records and weights defined in [4] and [5]. The dataset used was the one for the official October 2003 run for the Holstein breed, with about 9.5 million cows and about 25000 sires.

3.2. Validation of the EDC and DD formulae

EDCs and DDs were computed as described above. They were included in a univariate BLUP sire model with a sire variance equal to the one used in the Weibull evaluation and a residual variance of 1. The sire solutions from the Weibull and BLUP models were compared.

Restricting the comparison to sires with at least 10 uncensored daughters, the correlation between these sire solutions was above 0.99. The correlation between EDCs and number of uncensored daughters for the same period was also larger than 0.99. This reflects once again that the reliability of longevity proofs in survival analysis models is a function of the number of uncensored daughters, not of the total number of daughters (Ducrocq, 1999; Yazdi *et al.*, 2002).

3.3. Trend validation with the current model

The estimates of the year of use effects are reported in figure 1 (under “old” model). These estimates present a strong negative trend, with a decrease of about $0.9 \sigma_s$ over 10 years. Using model [8], the regression coefficient estimate is equal to $-0.101 \sigma_s$, well outside the Interbull limits. This corresponds to a strong underestimation of the genetic trend on culling, i.e, a strong overestimation of the genetic trend in published longevity breeding values. This trend is reported on figure 2. A very optimistic increase of $+0.9 \sigma_a$ was found over the last 10 years. The application of model [7] further illustrates the inconsistency of the evaluation (figure 3). The younger the bulls are, the steeper the decrease.

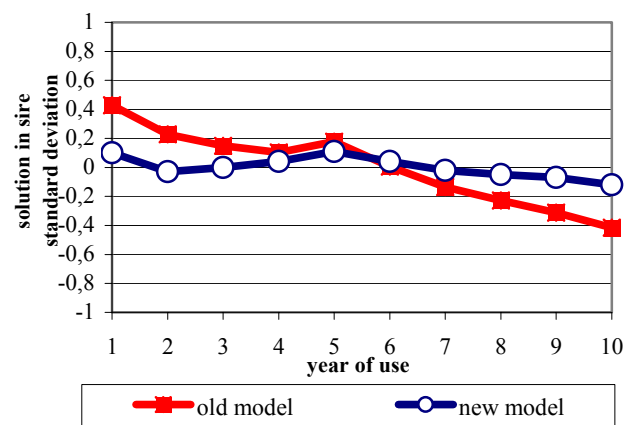


Figure 1. Estimates of the year of use effect using model [6] for the current (“old”) and the improved (“new”) evaluation models.

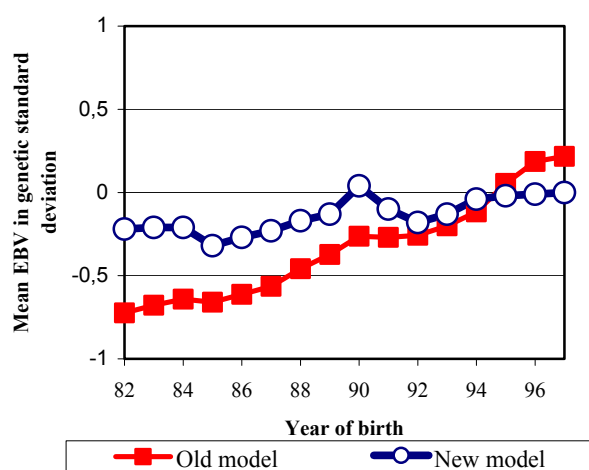


Figure 2. Genetic trends for the current (“old”) and the improved (“new”) evaluation models (positive values are favourable).

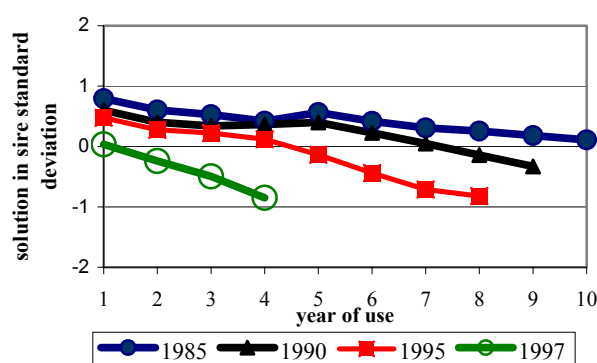


Figure 3. Estimates of the year of use effect using model [6], for the current (“old”) evaluation model, distinguishing bulls with first daughters calving before 1987 (class “1985”), between 1988 and 1992 (“1990”), between 1993 and 1996 (“1995”) and after 1996 (“1997”).

3.4. An improved model

A new model was looked for to obtain a more adequate genetic trend. First, a number of interactions between the factors already included in the current model were tested. None reduced the genetic trend. Some increased it. Then, the Weibull shape parameter ρ was re-estimated maintaining the current factors. Surprisingly, a large value was obtained (3.8 instead of 2), resulting in a very unrealistic negative genetic trend: the trend validation test showed fast increasing estimates of the year of

use effect, mirroring figure 3. This indicated that the definition of the baseline hazard function played a critical role on the estimated genetic trend.

Therefore, a different model was defined, following the work of Roxström *et al.* (2003) and Ducrocq (2002). In this model, the baseline hazard function is defined within lactation:

$$h(t; \mathbf{w}(t)) = h_c(\tau) \exp\{ \mathbf{w}(t)' \mathbf{b} \} \quad [9]$$

where the baseline $h_c(\cdot)$ is no longer a function of t , the number of days since first calving, but of τ the number of days since the most recent calving. This allows a much better description of the change in probability of being culled within lactation. $h_c(\cdot)$ was chosen to be a piecewise Weibull hazard function with a different parameter ρ for different combinations c of stages of lactation (with 4 stages: 0-270 days, 270-380, >380, dry period) $\times$ lactation number (1,2, 3+) $\times$ year-season. A few changes were also included in the fixed effect parts. These changes lead to results that passed the trend validation test for the March 2004 Interbull test run, but at the expense of tremendous computational difficulties: the Survival Kit had to be appropriately modified, first to allow a simple implementation of model [9], but mainly to accommodate the important numerical problems due to the recurrent existence of a non-positive definite Hessian matrix during the initial iteration steps of the maximisation. Even with a full Newton-Raphson approach, model [9] was much more time-consuming than the current model.

After the March 2004 test, simplifications or improvements were investigated. First, the trait definition was slightly changed: as one doubtful assumption of the current model is that the sire effect on risk of culling is constant over all the life of his daughters, it was decided to restrict the data to the first 5 lactations of a cow, i.e., to consider that a cow was censored at the end of her 5th lactation if she had a sixth calving. Then, the piecewise Weibull hazard was defined with stages of lactation as above but within year (not year-season) and lactation (1, 2, ..5). At the same time, only data since 1988 were considered (instead of 1984). These three changes alleviated the computational difficulties.

For the fixed effects in the model, the effect of within herd-year deviation class were defined within year-season (4 seasons) for milk and within year for fat and protein content. A stage of lactation x lactation number x class of milk production effect was included because it was found that the effect of low production on culling is different in early and late lactation, and in first and later lactations. A year of dam effect was added to correct for the genetic part not accounted for in the sire-maternal-sire model. A region (8) x year effect was included to account for regional differences. Finally, a herd-year effect replaced the herd-year-season effect which was poorly estimated in small herds with often only 0 or 1 cow culled in one season. The results for all these factors will be presented in more details at the next EAAP meeting in Bled (Slovenia).

The genetic trend obtained was only slightly positive with a variation of about $0.12 \sigma_a$ over the last ten years (figure 3). The trend validation test was much better (see figure 1). In fact, the regression coefficient estimate for model [8] still indicated a slight overestimation ($\beta = -0.017 \sigma_s$). The graph similar to figure 3 (model [7], with the same periods) did not reveal any clear discrepancy. Minor variations of the model are under study to get an even better trend validation test but, in any case, the new model will be used for the June 2004 evaluation.

5. Discussion

An obvious question is why the choice of a more flexible baseline hazard function had such an impact on genetic trend. A possible explanation follows: for many years, average daily milk production has been continuously increasing. For obvious economical reasons, farmers have delayed the moment when a cow, for example, a sterile cow, is culled because, with her increased production, she is profitable for a longer time. As an illustration, figure 4 shows the increase with time of the average number of days in lactations when a first lactation cow is culled, in the Holstein data set.

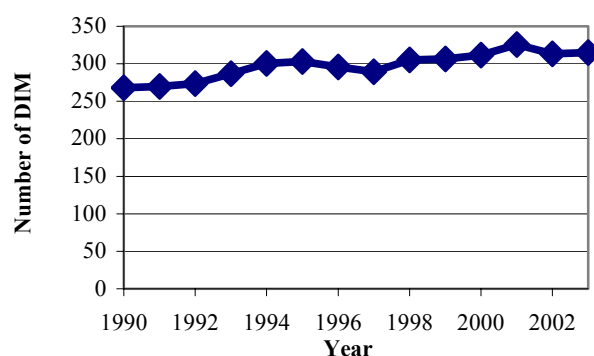


Figure 4. Number of days in milk (DIM) when culled for first lactation Holstein cows in France.

The baseline hazard function in the Weibull model represents the global culling pattern of the population under study. The choice of a unique baseline may have prevented a proper consideration of the observed delayed culling within lactation, due to a globally higher milk production. As a consequence, younger sires with on average higher production EBVs may have been inadequately considered as improving the longevity of their daughters.

References

- Boichard, D., Bonaiti, B., Barbat, A. & Mattalia, S. 1995. *J. Dairy Sci.* 78, 431-437.
- Bonaiti, B., Boichard, D., Barbat, A. & Mattalia, S. 1993. *Interbull Bulletin* 8.
- Ducrocq, V. & Casella, G. 1996. *Genet. Sel. Evol.* 28, 505-529.
- Ducrocq, V. & Sölkner, J. 1998. The Survival Kit - V3.0 : a package for large analyses of survival data. *6th World Cong. Genet. Appl. Livest. Prod.*, Armidale, N. S. Wales, Australia, 27, 447-448.
- Ducrocq, V. 1999. *Interbull Bulletin* 20, 60-67.
- Ducrocq, V. 2001. *Interbull Bulletin* 27, 147-152.
- Ducrocq, V. 2002. A piecewise Weibull mixed model for the analysis of length of productive life of dairy cows. *7th World Cong. Genet. Appl. Livest. Prod.*, Montpellier, France, August 19-23, 2002, Communication n° 20-04.
- Ducrocq, V., Delaunay, I., Boichard, D. & Mattalia, S. 2003. *Interbull Bulletin* 30, 101-11.

- Ducrocq, V., Boichard, D., Barbat, A. & Larroque, H. 2001. Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. *52nd EAAP Annual Meeting*. Budapest, Hungary, 7.
- Interbull, 2004. Interbull trend validation procedure http://www.interbull.slu.se/service_documentation/General/framesida-general.htm
- Roxström, A., Ducrocq, V. & Strandberg, E. 2003. *Genet. Sel. Evol.* 35, 305-318.
- Yazdi, M.H., Visscher, P.M., Ducrocq, V. & Thompson, R. 2002. *J. Dairy Sci.* 85, 1563-1577.