

Tolerance Values for Validation of Trends in Genetic Variances Over Time

W.F. Fikse¹, Z. Liu², P.G. Sullivan³

¹ Interbull Centre, Uppsala, Sweden

² VIT, Verden, Germany

³ CDN, Guelph, Ontario, Canada

Introduction

Presence of systematic trends in genetic variance estimated from bull breeding values potentially affects international comparisons and hence leads to suboptimal selection decisions. Thus, the Interbull Technical Committee initiated a working group to devise a test to validate trends in estimated genetic variances.

Sullivan (1999) devised a method to estimate of genetic variance within strata using breeding values and prediction error variance of Mendelian sampling deviations (**PEV_{MS}**). Computation of **PEV_{MS}** is not feasible for large data sets, and Fikse *et al.* (2003) outlined a procedure to approximate **PEV_{MS}** using (approximated) reliabilities (**REL**) of each animal and its parents. Depending on the quality of these approximations there can be time trends in bias of approximated **PEV_{MS}**. The criteria for validating the absence of time trends in genetic variances therefore need to be robust to the approximate nature of the procedure.

The aim of this study was to develop a procedure to obtain tolerance intervals for within-year estimates of genetic variance. The tolerance intervals need to reflect both the dependency on approximations and the limited sample size.

Estimation of Genetic Variance

Estimates of genetic variance by birth year can be obtained if Mendelian sampling deviations (**MS**) and their prediction error variance (**PEV_{MS}**) are available for all animals (Sullivan, 1999):

$$\hat{\sigma}_{u_i}^2 = \frac{1}{q_i} \sum_{k=1}^{q_i} d_k \left[\hat{m}_k^2 + PEV(\hat{m}_k) \right],$$

where q_i is the number of animals in birth year i , and d_k the inverse of the genetic variance (proportional to σ_u^2) that has not been explained by known parents. Rearranging this equation yields:

$$\hat{\sigma}_{u_i}^2 = \frac{\sum_{k=1}^{q_i} d_k \cdot \hat{m}_k^2}{q_i - \sum_{k=1}^{q_i} d_k \cdot PEV^*(\hat{m}_k)}, \quad [1]$$

where $PEV^*(\hat{m}_k)$ is **PEV_{MS}** expressed in genetic variance units.

Bootstrapping Tolerance Intervals

For the data set used in Fikse *et al.* (2003) it was observed that for more than 97% of the animals the true **PEV_{MS}** was within the interval [IPEV, uPEV], where IPEV=0.99·a**PEV_{MS}**, uPEV=1.02·a**PEV_{MS}**, and a**PEV_{MS}** is the approximated **PEV_{MS}**. Thus, a lower bound of the genetic variance was computed by accumulating IPEV for all animals and inserting this in formula [1], and the upper bound was computed similarly by accumulating uPEV for all animals. Note that **PEV_{MS}** (and uPEV) is at most d_k^{-1} times the genetic variance.

Uncertainty due to limited sample size was dealt with by bootstrapping the lower and upper bound of the estimated genetic variance. For each birth year animals were sampled with replacement, with the sample size equal to the number of animals in that birth year. IPEV and uPEV were accumulated for all animals in the bootstrap sample and inserted in formula [1] to yield a lower and upper bound of the genetic variance for the bootstrap sample. This was repeated to obtain 1000 bootstrap replicates for each birth year. The lower and upper limits

of the tolerance interval for the genetic variance were obtained as the $\alpha/2^{\text{th}}$ percentile of the lower bound samples, and the $(1-\alpha/2)^{\text{th}}$ percentile of the upper bound samples, respectively.

Illustration

Data

First lactation records on milk yield for 32,306 Guernsey cows produced between January 1990 and December 1999 from the United States were used. Pedigree records on an additional 2861 bulls and 33,710 cows were added. For more details about the data set see Fikse *et al.* (2003).

Method

Approximate reliabilities for all animals were computed using the procedure described by Misztal and Wiggans (1988; hereafter referred to as REL_m). Two different sets of REL , in addition to REL_m , were considered:

- REL_{m95} : REL_m multiplied by 0.95.
- REL_{m_v} : REL_m with variation increased by adding a random normal deviate $\sim N(0, 0.05)$. Reliabilities of parents were required to lie between zero and one, and reliabilities of animals were required to be larger than the reliability of the parent average.

The upper and lower limit of the tolerance interval was determined with the bootstrap procedure for each of these three sets of reliabilities. For comparison, the exact estimate of the genetic variance was computed according to formula [1], using the true PEV_{MS} .

Results

The exact estimate of the genetic variance within birth year was always contained in the tolerance interval obtained by bootstrapping, as expected (Figure 1). The width of the tolerance interval was fairly constant over time, with the exception of the first few years. The number of cows was smaller in those years and there were more cows without records compared with 1988 onwards.

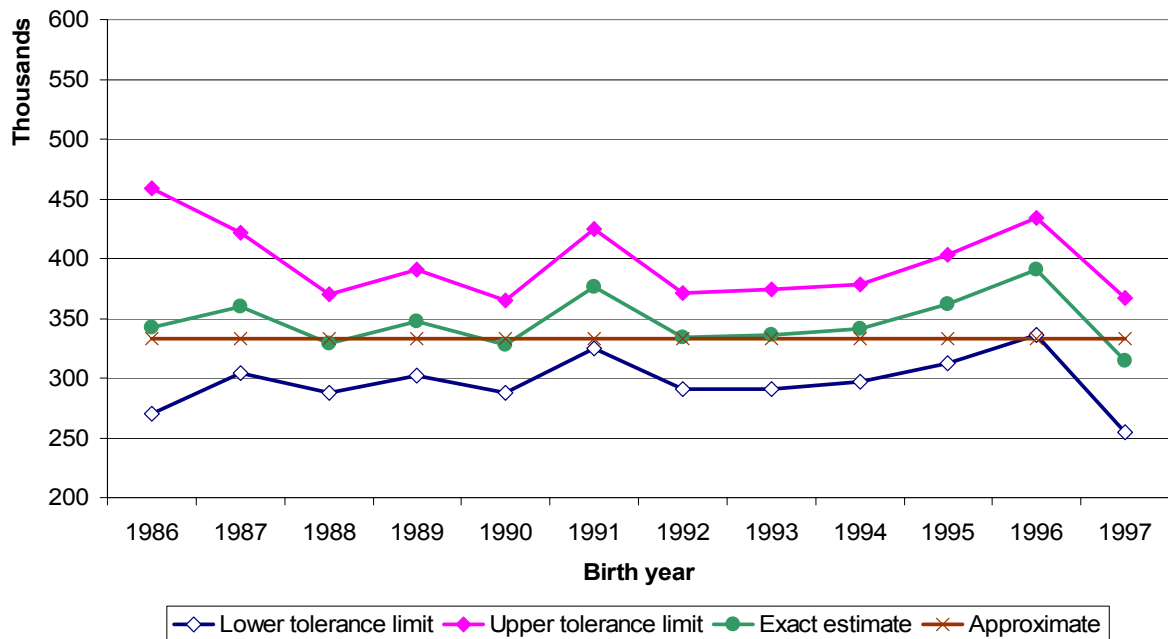


Figure 1. Exact within-year estimate, approximate across-year estimate and tolerance intervals for the genetic variance, using REL_m in the approximation of PEV_{MS} .

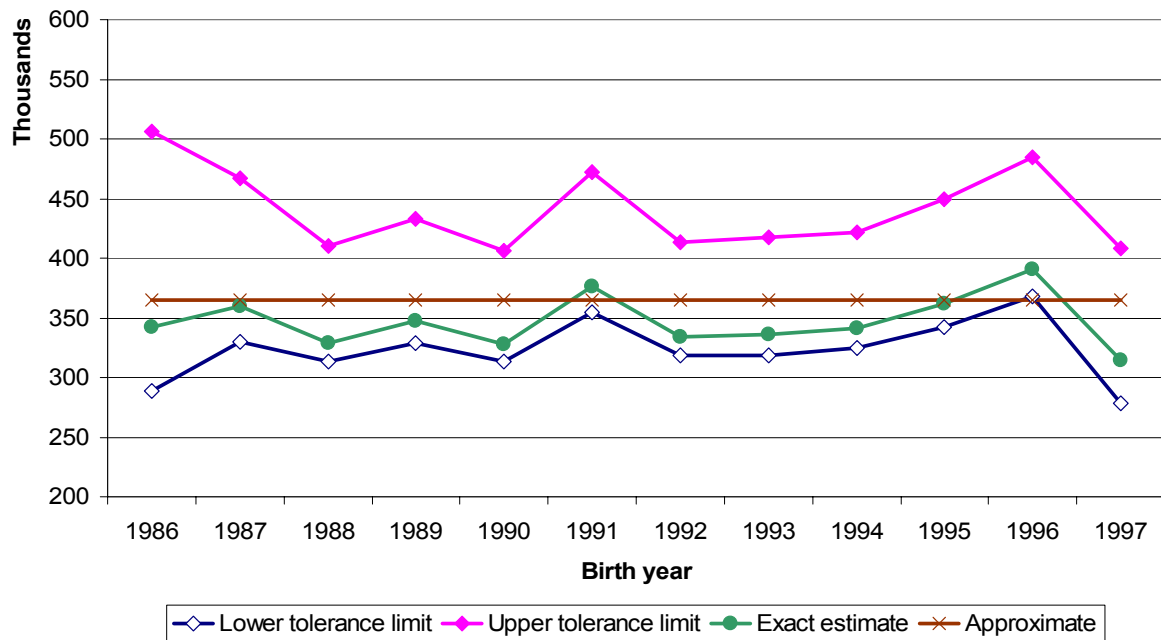


Figure 2. Exact within-year estimate, approximate across-year estimate and tolerance intervals for the genetic variance, using REL_{m95} in the approximation of PEV_{MS} .

In the absence of a trend in genetic variance over years, it is expected that for $\alpha \cdot 100\%$ of the years the across-year estimate of genetic variance is outside the tolerance interval by chance. Thus, if the across-year estimate falls outside the tolerance interval for more than

$\alpha \cdot 100\%$ of the years, the hypothesis that no trend exists would be rejected. The across-year estimate of the genetic variance was 333,000 kg^2 , which was outside the tolerance interval for 1996 (Figure 1). Thus, absence of a time trend in genetic variance is validated.

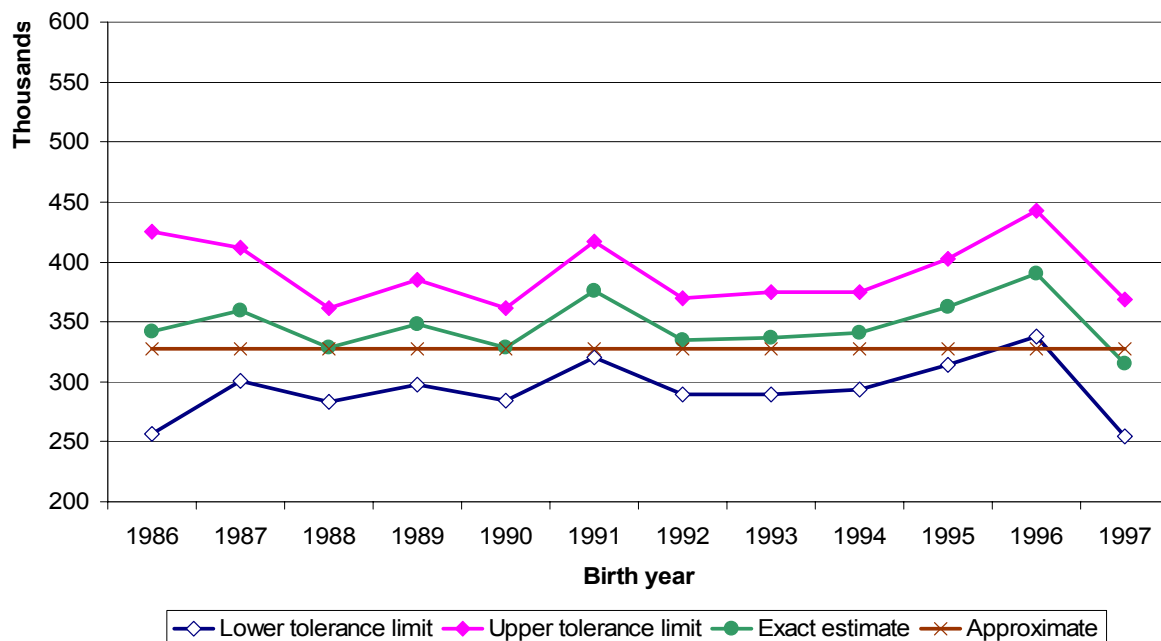


Figure 3. Exact within-year estimate, approximate across-year estimate and tolerance intervals for the genetic variance, using REL_{m-v} in the approximation of PEV_{MS} .

The same conclusion could be drawn when reliabilities with artificially introduced noise were used to approximate the genetic variance and to obtain tolerance intervals (Figures 2 and 3). The exact estimates of the within-year genetic variance were within the tolerance interval despite the artificially introduced noise in reliabilities. Regression of the reliabilities towards zero (REL_{m95}) resulted in both the approximated genetic variance and the tolerance limits being shifted upwards (Figure 2). The tolerance interval was somewhat wider as well. Adding random noise to the reliabilities did not affect the tolerance interval or the approximated across-year genetic variance estimate significantly.

Note that REL_m itself were approximated reliabilities, and that the alternatives REL_{m95} and REL_{m-v} over-accentuate the effect of the approximate nature of the procedure.

Discussion

Previous studies demonstrated the feasibility to estimate genetic variance within year and gender when good approximate reliabilities were available. The present study outlined a procedure to obtain tolerance intervals for the within-year genetic variance taking into account the approximations necessary to esti-

mate the genetic variance and the limited sample size.

For the data set utilized in this study it was observed that for more than 97% of the animals the true PEV_{MS} was within the interval $[0.99 \cdot aPEV_{MS}, 1.02 \cdot aPEV_{MS}]$, and the values 0.99 and 1.02 were subsequently assumed in the procedure to obtain tolerance intervals. These values can be modified if approximation methods perform better or worse than was the case in the present study. Widening the interval, however, will probably reduce the power of the test.

References

- Fikse, W.F., Klei, L., Liu, Z. & Sullivan, P.G. 2003. Procedure for validation of trends in genetic variance. *Interbull Bulletin* 31, 30-36.
- Harris, B. & Johnson, D. 1998. Approximate reliability of genetic evaluations under an animal model. *J. Dairy Sci.* 81, 2723-2728.
- Misztal, I. & Wiggans, G.R. 1988. Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* 71, 27-32.
- Sullivan, P.G. 1999. REML estimation of heterogeneous sire (co)variances for MACE. *Interbull Bulletin* 22, 146-148.

