

Simultaneous Estimation of Genetic Correlations for Milk Yield Among 27 Holstein Populations

Hossein Jorjani

Interbull Centre

Department of Animal Breeding and Genetics

Swedish University of Agricultural Sciences

S - 750 07, Uppsala, Sweden

Hossein.Jorjani@hgen.slu.se

Abstract

Genetic correlations among bulls' estimated breeding values in different countries, which are needed for implementation of multi-trait across country evaluations, are currently estimated in sub-sets of data in the form of certain country combinations, at most 10 countries at a time. In order to better utilize the data, especially genetic ties, it is desirable to estimate correlations among all countries of Holstein populations (currently 27 populations) at one single run. However, this is computationally not plausible at the moment and therefore, a different kind of selection of sub-set of data is needed. Here, effects of selection of sub-sets of data based on bulls' statistical connectedness and effective number of proofs on estimation of genetic correlations among countries is reported. Results indicate that reasonable estimates are obtainable by relatively low number of bulls. However, there is a need to verify these results by simulated data.

Introduction

Implementation of an international breeding value evaluation of bulls with the method known as Multi-trait Across Country Evaluation (MACE), as is done in the Interbull Centre, requires good estimates of genetic correlations. However, estimation of genetic correlation between certain country combinations is problematic primarily because of lack or shortage of bulls with proofs in these countries. Currently, genetic correlations are estimated from sub-sets of data in the form of specific country combinations, using only those bulls that have multiple proofs in different countries plus full-sibs or $\frac{3}{4}$ sibs of such bulls. In case of Holstein populations, and because of large sizes of these populations, the computational demand is so high that we are limited to considering only a few countries, at most 10 countries, at a time. The end result is that the information coming from genetic ties is not fully utilized. In order to estimate all correlations in one run by including all countries Jorjani (2000) suggested that the selection of sub-set of data to be based on individual bulls' connectedness rather than countries. The aim of the present study is to investigate the effects of selection of sub-sets of bulls on estimation of genetic correlations as compared to the currently used estimates.

Material and Method

Data sent to Interbull for March 2001 test-run was used in this study. There were a total of 97825 bulls with 114571 proofs present in the data. These bulls were reported from 24 countries and evaluated as 27 different Holstein populations. However, because some other breeds (e.g. MRY and RED) are also evaluated together with Holstein, there were altogether 40 different breed-country combinations present in the data. The pedigree file contained the information for 127284 bulls. For selection of sub-sets of bulls from each population, first each bull's Effective Number of Proofs (ENP_i) was calculated according the following equations (Jorjani, 2000):

$$ENP_i = N_c * BC_i,$$

$$BC_i = 1 - \frac{(\sum_{j=1}^{N_c} |n_{ij} - \frac{\bar{n}_i}{N_c}|) / N_c}{|n_i - \frac{n_i}{N_c}| + (N_c - 1) | \frac{n_i}{N_c} |}$$
$$N_c - 1$$

where N_c is the number of populations in the evaluation and BC_i is the statistical connectedness for bull i , and n_{ij} is the number of daughters of bull i in population j . In the next step many different selection criteria were used to select sub-sets of data for estimation of genetic correlations, though only three models are presented in this paper.

Model I: A two-step selection process was used in this model. First, from all countries 100 proofs coming from bulls with the highest ENP value were selected. These bulls had to have multiple EBVs (i.e. $ENP \geq 2.00$). In practice, for the larger populations the minimum ENP was between 6.50-8.50 and for smaller countries as low as 2.00. However, the number of selected bulls did not reach 100 for the smaller populations. In the second step breed-country combinations were divided into three groups according to their size. From smaller breed-country combinations all bulls (i.e. $ENP \geq 1.00$) and from medium sized breed-country combinations all bulls with more than one EBV (i.e. $ENP > 1.01$) were added to the list of bulls selected in the first step. This model was assumed to be the base model providing enough information for unbiased estimation of variances, reasonably good estimates of covariances and utilizing genetic ties to a high degree, all within the computational restrictions of currently available hardware.

Model II: In this model a one-step selection process, similar to the first step of Model I was used.

Model III: In this model, again a one-step selection process was used. The difference with Model II was that for the larger populations the selection criterion was more strong (i.e. $ENP \geq 9.00$) and for smaller countries more relaxed (i.e. $ENP > 1.01$).

Estimation of genetic correlation for each sub-set of data comprising bulls from all 27 populations followed the normal practice at the Interbull Centre (Sigurdsson *et al.*, 1996).

Further, in order to examine the effects of time edit on estimated genetic correlation Models I and II were also used together with time edit so that only bulls born 1984 or thereafter were used in the analysis.

Results and Discussion

Actual number of EBVs reported from three small and three large populations are shown in Table 1 for bulls with different number of proofs (NP). As an example, from country 9 there are 1500 EBVs reported from bulls that have 2 proofs each and 4 EBVs from bulls that have 21 proofs each. These EBVs may be coming from bulls that have their country of first registration as Country 9, or they may be coming from other countries' bulls.

Table 1. Number of EBVs reported from three small and three large populations for bulls with different number of proofs (NP)

NP	Country					
	19	27	4	2	5	9
1	104	249	693	8697	12149	26124
2	1	5	11	797	880	1500
3	3	5	2	293	287	563
4		3	6	166	194	292
5	2	3	4	123	145	173
6		5	1	104	111	154
7	1	2	2	79	70	77
8			2	55	56	63
9	2	1	7	51	61	55
10	2	2	2	53	58	49
11		1	1	47	48	41
12			1	37	38	32
13	1		2	41	45	43
14	1		2	27	27	29
15	1	1	3	21	22	22
16	3		2	21	21	20
17	2		2	12	12	12
18			1	13	13	13
19	6		1	17	17	17
20	2			6	6	6
21	2		2	4	4	4

Expected value of ENP is NP/N_c and therefore, for a bull with totally balanced distribution of number of daughters across countries $ENP=NP$. The more unbalanced is the number of daughters in different countries, the higher becomes the difference between ENP and NP. Obviously, this puts at advantage those bulls that are simultaneously progeny tested in different countries. And from this it has been postulated that the use of ENP will lead to selection of less unbiased sub-set of bulls.

Table 2 shows basically the same kind of statistics as is shown in Table 1 with the exception that bulls are categorized according to their effective number of proofs (ENP), instead of number of proofs (NP). Because ENP is a continuous variable, they have been rounded towards zero to the nearest integer value. It can be seen that 1537 EBVs are reported from Country 9 from bulls that have an ENP value between 2.00 and 2.99.

Table 2. Number of EBVs reported from three small and three large populations for bulls with different effective number of proofs (ENP)

ENP	Country					
	19	27	4	2	5	9
1	104	250	699	8947	12565	26457
2	2	8	15	873	880	1537
3	4	9	12	312	242	533
4	3	6	4	156	154	279
5	2	1	4	111	137	174
6	6	2	4	91	90	111
7	2	1		39	51	51
8	1		3	58	65	66
9	4		3	41	45	45
10	5		3	17	16	17
11				18	18	18
12				1	1	1

Comparison of values (numbers of reported EBVs) for NP and ENP also shows that it is mainly the right tail of distribution that is affected. In other words it is postulated and the hope is that, for example, a bull with 20 proofs and a very unbalanced number of daughters is put at a very disadvantageous position compared to a bull with 10 proofs and balanced number of daughters.

Table 3. Number of bulls, ancestors without own EBV and average number of EBVs per bull in the three Models of analysis

	Model I	Model II	Model III
Bulls	1110	716	436
Ancestors	674	342	319
Proofs / bull	3.66	4.20	3.83

Table 3 shows the number of bulls, ancestors and average number of EBVs per bull in the models used. Two kinds of dilution of information can be recognized in these models. In Model I a proportionately smaller number of bulls with large

number of EBVs must be utilized to estimate covariances. In Model III we are encountered with a mixture of highly selected sub-set of bulls and a different sub-set with only one or two EBVs, however, not too many genetic ties to bind these two groups together. Model II comprising bulls with an average number of EBVs per bull equal to 4.2 poses another difficulty, that is how valuable and representative are this group of highly selected bulls for estimation of correlations.

Table 4 shows the average correlation of each of the three small and the three large populations with the other 26 Holstein populations. The first observation is that these correlations are smaller than the ones used in Interbull routine runs. That is because the guiding principle in implementation of MACE in Interbull's routine runs, as suggested by Sigurdsson *et al.* (1996) is that genetic correlations cannot be overestimated. Therefore, for each country combination the highest correlation ever obtained is used in routine runs. If we take this guiding principle for granted, then Model III yields the best results. Intuitively, one expects Model II to lead to highest, though biased, estimates of genetic correlation, because of heavy reliance on a highly selected sub-set of bulls.

Table 4. Average of genetic correlations for three small and three large populations with other 26 Holstein populations

	Country					
	19	27	4	2	5	9
Model I	0.649	0.851	0.840	0.826	0.878	0.845
Model II	0.720	0.851	0.844	0.829	0.885	0.851
Model III	0.736	0.872	0.860	0.851	0.901	0.869

Results shown in Table 4 are rather consistent across countries, however, correlations between specific country combinations, especially for smaller populations, show occasional anomalies (results not shown) in the form of very low genetic correlations or a specific pattern across models which is not consistent with the general pattern of the models investigated. This is because using a single value for minimum ENP across countries of a certain size (*i.e.* small, medium and large size) may lead to very few bulls for that specific country combination. This phenomenon may be defined as vulnerability of sub-setting process to population structure.

Weigel and Banos (1997) showed that using information coming from old bulls may lead to

biased estimates of breeding values. To see if the same can be said about genetic correlations two variants of Models II and III with time edit are also presented. Table 5 compares number of bulls, ancestors and average number of proofs for Models II and III and their equivalent time-edited Models.

Table 5. Number of bulls, ancestors without own EBV and average number of EBVs per bull in the three Models of analysis

	Model II	Model II 84	Model III	Model III 84
Bulls	716	461	436	301
Ancestors	342	266	319	246
Proofs / bull	4.20	4.27	3.83	3.91

Table 6. Average of genetic correlations for three small and three large populations with other 26 Holstein populations in Models II and III and their time-edited equivalents

	Country					
	19	27	4	2	5	9
Model II	0.720	0.851	0.844	0.829	0.885	0.851
Model II 84	0.681	0.736	0.846	0.826	0.869	0.852
Model III	0.736	0.872	0.860	0.851	0.901	0.869
Model III 84	0.685	0.768	0.858	0.840	0.885	0.862

The resulting genetic correlations from time-edited data are shown in Table 6. For the most part the changes are very small, so that one can draw the conclusion that time edit has very little effect on the estimated genetic correlations. On the other hand, we can observe again the vulnerability of the sub-setting process to population structure, as older bulls may be contributing significantly to the links between some populations, especially between some Eastern Europe populations and other Holstein populations.

As mentioned before, interpretation of the results obtained in the present study has a dependence on the suggestion put forward by Sigurdsson *et al.* (1996) that within the framework of the current methodology it is not possible to overestimate the genetic correlations. This suggestion, however, has been challenged by Klei and Weigel (1998). To resolve the issue it seems only logical to conduct a new simulation study and re-examine the results obtained by Sigurdsson *et al.* (1996) and Klei and Weigel (1998).

Acknowledgements

Financial support from USDA/NAAB is gratefully appreciated.

References

- Jorjani, H. 2000. Well-connected, informative sub-sets of data. *Interbull Bulletin* 25, 22-25
- Klei, B. & Weigel, K.A. 1998. A method to estimate correlations among traits in different countries using data on all bulls. *Interbull Bulletin* 17, 8-14.
- Sigurdsson, A., Banos, G. & Philipsson, J. 1996. Estimation of genetic (co)variance components for international evaluation of dairy bulls. *Acta Agric. Scand., Sect. A, Anim. Sci.* 46, 129-136.