

3.0 METHODS OF TRANSFORMING PROOFS

3.1 Desirable properties

In order to analytically compare various methods of transforming proofs from country A (exporting country) to those of country B (importing country), a list of properties that were considered desirable had to be established. These properties should be considered as minimum standards for any new method which might be proposed.

The method should

1. Give unbiased estimates of both the intercept, a , and the slope, b .
2. Consider the difference in reliabilities of proofs from each country.
3. Allow for a possible genetic correlation between true values in each country of less than 1.
4. Minimize the variance of differences between transformed proofs and true values in country B.

From these properties a method can be derived which meets all of these conditions. The following notation and definitions shall be used to derive this method and to describe other methods.

Let

P_{ij} = the proof of the j^{th} bull in country i .

S_{ij} = the true merit (TA) of the j^{th} bull in country i .

$w_{ij} = r_{PS}^2$, squared correlation between P_{ij} and S_{ij} , assumed to be equal to $n_{ij} / (n_{ij} + k_i)$, where

n_{ij} = the number of effective daughters of the j^{th} bull in country i , and

k_i = the ratio of $\sigma_{ci}^2 / \sigma_{si}^2$ in country i , used to calculate P_{ij} .

Y_{ij} = the daughter average of the j^{th} bull in country i .

Hence,

$$P_{ij} = G_i + w_{ij} (Y_{ij} - u_i)$$

where

G_i = the genetic base constant to which all bulls in country i are compared, and

u_i = the average daughter average of all bulls in country i .

Usually, u_i and G_i are unknown to most people buying semen. The parameter a in the regressions should estimate $G_B - G_A$ in an unbiased manner. Note that this is totally unrelated to genetic differences between bulls in countries A and B.

All P_{ij} are assumed to be unbiased, and differences in P_{ij} should reflect differences among bulls in additive genetic merit within a country. Thus, any non-additive genotype by environment interactions (such as heterosis) are assumed to be negligible, and any additive genotype by environment interactions are assumed to contribute to a genetic correlation that is less than one.

3.2 Description of methods

3.2.1 Method 1 (true beta)

If true genetic values of bulls were known, then the appropriate regression coefficient would be

$\text{beta} = \text{Cov}(S_B, S_A) / V(S_A)$
and
 $\text{alpha} = \text{base difference between } S_B \text{ and } S_A \text{ for all bulls.}$

In most practical situations alpha and beta are unknown. The variance of prediction error would be

$$\begin{aligned} V(S_{Bj} - S_{Bj}) &= V(\text{beta } S_{Aj} - S_{Bj}) \\ &= \text{beta}^2 V(S_A) + V(S_B) - 2 \text{beta } \text{Cov}(S_A, S_B) \\ &= \frac{\text{Cov}(S_A, S_B)^2}{V(S_A)} + V(S_B) - 2 \frac{\text{Cov}(S_A, S_B)^2}{V(S_A)} \\ &= V(S_B) - \frac{\text{Cov}(S_A, S_B)^2}{V(S_A)} \\ &= (1 - r_g^2) V(S_B) \end{aligned}$$

Thus, if $r_g = 1$, then the smallest prediction error variance would be zero.

In the following simulation study, the values of alpha and beta were known. Thus method 1 was

$$P_{Bj} = \text{alpha} + \text{beta } P_{Aj}$$

The prediction error variance is

$$V(\text{beta } P_{Aj} - S_{Bj}) = \text{beta}^2 V(P_{Aj}) + V(S_B) - 2 \text{beta } \text{Cov}(P_{Aj}, S_B)$$

If

$$V(P_{Aj}) = w_{Aj} V(S_A)$$

and

$$\text{Cov}(P_{Aj}, S_B) = w_{Aj} \text{Cov}(S_A, S_B),$$

then

$$V(\text{beta } P_{Aj} - S_{Bj}) = (1 - r_g^2 w_{Aj}) V(S_B) \quad (1)$$

Thus, (1) is the smallest possible prediction error variance of this method. If the average w_{Aj} is .65 and $r_g = 1$, then the prediction error variance would be .35 $V(S_B)$ for that group of bulls. Note that if $r_g = .9$ and $w_{Aj} = .65$, then the prediction error variance increases to .4735 $V(S_B)$.

3.2.2 Method 2 (ordinary least squares)

The usual regression approach uses the model

$$P_{Bj} = a + b_2 P_{Aj} + e_j$$

and estimates of a and b_2 are calculated by least squares.

The estimator of b_2 is

$$\begin{aligned} b_2 &= \text{Cov}(P_{Aj}, P_{Bj}) / V(P_{Aj}) \\ &= w_{Aj} w_{Bj} \text{Cov}(S_A, S_B) / w_{Aj} V(S_A) \\ &= w_{Bj} \text{beta} \end{aligned}$$

This assumes that $V(w_A) = V(w_B) = 0$. Deviations from this assumption are in most practical situations negligible. Still, b_2 is a biased estimator of beta . The prediction error variance from using b_2 would be

$$V(b_2 P_{Aj} - S_{Bj}) = (1 - r_g^2 w_{Aj} (2 w_{Bj} - w_{Bj}^2)) V(S_B)$$

The quantity, $(2 w_{Bj} - w_{Bj}^2)$ is less than or equal to one in all cases, and consequently, the prediction error variance using b_2 is always greater than or equal to the prediction error variance using beta . They are equal when $w_{Bj} = 1$.

3.2.3 Method 3

Following procedure 2 of Wilmink et al. (1985), for $P_{A.} = 0$ (average proof in country A of the group to which the bull belongs) and again assuming $V(w_B) = 0$ the bias in b_2 can be removed by the model

$$P_{Bj} = a + b_3 w_{Bj} P_{Aj} + e_j$$

Then

$$\begin{aligned} b_3 &= \text{Cov}(w_{Bj} P_{Aj}, P_{Bj}) / V(w_{Bj} P_{Aj}) \\ &= b_2 / w_{Bj} \\ &= \text{beta} \end{aligned}$$

Thus, b_3 should give similar prediction error variances as when beta is used except that b_3 is only an estimate of beta .

3.2.4 Method 4

Goddard (1984) proposed the following estimator of $\text{Cov}(P_A, S_B) / V(P_A)$. The model was

$$P_{Bj}^* = a + b_4 P_{Aj} + e_j$$

where

$$P_{Bj}^* = (P_{Bj} - P_B) w_{Bj}^{-1} + P_B.$$

P_B is the average of all bulls proven in country B, not just bulls that also have proofs in country A. If groups are included in the model, P_B is the mean of the group to which a bull belongs. Then,

$$b_4 = \text{Cov}(P_A, S_B) / V(P_A) = \text{Cov}(w_{Aj} S_A, S_B) / V(w_{Aj} S_A)$$

If $V(w_{Aj}) = 0$, then

$$b_4 = \text{Cov}(S_A, S_B) / V(S_A) = \text{beta}$$

Thus, b_4 should be similar to b_3 and beta in prediction error variances.

a is defined to be $P_B - b_4 P_A$, then

$$P_{Bj} = P_B - w_{Bj} b_4 P_A + w_{Bj} b_4 P_{Aj} + e_j$$

$P_B - w_{Bj} b_4 P_A$ is the same as a in Method 3, and, therefore, Methods 3 and 4 are the same.

Goddard (1984) also proposed a weighted least squares analysis with weights equal to

$$(w_{Bj}^{-1} - r_g^2 w_{Aj})^{-1}$$

However, if $r_g = 1$, then as w_{Bj} and w_{Aj} approach one, then this weight approaches infinity. This is not desirable in a weighting scheme. In practice, however, it would be sensible to set an upper limit to the weights. In the simulation study to compare methods, the weights were set equal to one, and P_B was calculated only on bulls with proofs in country A.

3.2.5 Method 5

Schulte-Coerne and Gravert (1984) proposed a maximum likelihood estimator for b where

$$b_5 = \frac{V(P_B) - K V(P_A) + ((V(P_B) - K V(P_A))^2 + 4 K \text{Cov}(P_B, P_A)^2)^{1/2}}{2 \text{Cov}(P_B, P_A)}$$

and

$$K = (1 - w_B) w_B V(S_B) / ((1 - w_A) w_A V(S_A))$$

Problems arise when $\text{Cov}(P_B, P_A) = 0$, but when $w_B = w_A$, then

$$K = V(S_B) / V(S_A) = V(P_B) / V(P_A), \text{ and}$$

$$\begin{aligned} b_5 &= \frac{4 K \text{Cov}(P_B, P_A)^2}{2 \text{Cov}(P_B, P_A)} = K \\ &= (V(S_B) / V(S_A))^{1/2} \end{aligned}$$

Otherwise, b_5 overestimates beta in general, and will lead to greater prediction error variances than beta . The estimate for the base difference is

$$a = P_B - b_5 P_A.$$

3.3 Simulation comparison

The five methods were compared under the following conditions. The true parameters were

$$V \begin{bmatrix} S_A \\ S_B \end{bmatrix} = \begin{bmatrix} 1 & y \\ y & x \end{bmatrix} \quad V(S_A) = G$$

where

$x = V(S_B) / V(S_A) \equiv$ (values of 1, 2, and .5 were used)

and

$$y = r_g (1/x)^{1/2}$$

Values of r_g were 1 and .8, $V(S_A)$ was equal to 1.

G was triangularized into TT' , where

$$T = \begin{bmatrix} 1 & 0 \\ y & z \end{bmatrix}$$

and

$$z = (x - y^2)^{1/2}, \text{ if } r_g = 1, \text{ then } z \text{ was } 0.$$

True S_A and S_B values were generated for 2000 bulls. If u_1 and u_2 are random normal deviates with zero means and variance 1, then

$$\begin{bmatrix} S_{Aj} \\ S_{Bj} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ y & z \end{bmatrix} \begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix} \quad \text{for } j = 1, \dots, 2000$$

Note that $V(Tu) = T V(u) T' = TT' = G$, which is the desired result.

The heritability of the trait in country A was .25, and in country B was .30.

Proofs in country A were generated for all 2000 bulls. Each bull was randomly assigned an effective number of daughters, n_{Aj} according to the following scheme.

1. Generate a random number between 0 and 1, say q .
2. If $q < .7$, then another random number was generated between 1 and 50 and this became n_{Aj} .
3. If $q \geq .7$, then n_{Aj} was generated between 1 and 500.

From n_{Aj} , then

$$w_{Aj} = n_{Aj} / (n_{Aj} + k_A)$$

$$k_A = V(e_A) / V(S_A) = (4 - h_A^2) / h_A^2$$

and

$$V(e_A) = k_A V(S_A)$$

The proof was formed by generating a random normal deviate, u_3 , and computing

$$P_{Aj} = w_{Aj} (S_{Aj} + u_3 V(e_A) / n_{Aj})$$

The mean of all P_{Aj} was therefore zero.

Two methods were used to decide which bulls had proofs in country B. In both cases, the minimum w_{Aj} was set at .75. In the first method, 200 bulls were randomly chosen from the 2000 provided that $w_{Aj} > .75$. In the second method, the best 200 bulls with $w_{Aj} > .75$ were chosen to have proofs in country B.

In both cases, n_{Bj} was generated for the 200 bulls chosen, by generating a number between 1 and 20 and adding this to the minimum n_{Bj} needed to have $w_{Bj} > .75$. Thus, P_{Bj} and P_{Aj} both had minimum repeatabilities of .75. This has been a common requirement of a few practical studies of actual data. The mean proof in country B was set to be 50.0.

Fifty replicates of each set of parameters were analyzed and the following statistics were calculated:

1. the average a - value
2. the average b - value
3. the average $\hat{S}_{Bj} - S_{Bj}$ over 2000 bulls
4. the average $V(\hat{S}_{Bj} - S_{Bj})$ over 2000 bulls

Although the a and b values were estimated from data on 200 bulls, they were applied to all 2000 bulls to study their validity on the entire population of bulls in country A.

Simple t-tests at the .05 level were used to determine if a was significantly different from 50, and if b was significantly different from β_a , and if the average $\hat{S}_{Bj} - S_{Bj}$ was different from zero, and if $V(\hat{S}_{Bj} - S_{Bj})$ was significantly different from that of Method 1 when β_a was used.

The correlation between S_{Bj} and S_{Aj} was calculated and was the same for all methods regardless of a or b values.

The following sets of parameters were studied.

Set	r_g	$V(S_B)/V(S_A)$	$w_A = w_B$	Sires randomly chosen
A	1.0	1.0	No	Yes
B	1.0	2.0	No	Yes
C	1.0	0.5	No	Yes
D	0.8	1.0	No	Yes
E	0.8	2.0	No	Yes
F	0.8	0.5	No	Yes
G	1.0	1.0	Yes	Yes
H	1.0	1.0	No	No
I	1.0	2.0	No	No
J	0.8	1.0	No	No
K	0.8	2.0	No	No

3.4 Results

When sires with proofs in country B were a random sample from country A, none of the methods gave biased estimates of either the intercept (i.e., difference in bases between proofs) or the predicted proof. However, ordinary Least Squares (method 2) and Maximum Likelihood (method 5) gave biased estimates of β_a . As shown earlier, Least Squares can only give an unbiased estimate of β_a

when all $w_{Bj} = 1$, and Maximum Likelihood requires all w_{Bj} and $w_{Aj} = 1$ and $r_g = 1$. Since bulls are never proven with a repeatability of 1, Maximum Likelihood and Least Squares will always give biased estimates of **beta**. Least Squares is always biased downwards by the average w_B and Maximum Likelihood is always biased upwards (see Tables 3.1 to 3.11).

When sires with proofs in country B were a selected group of bulls from country A, then

- 1) Maximum Likelihood (method 5) gave biased estimates of both **alpha** and **beta** and was the method most affected by selection.
- 2) Method 4 gave biased estimates of **alpha**, but not **beta**. This was due to the calculation of P_B , which in this simulation study was based only on the sample of bulls in country B, and obviously P_B is affected by the selection differential. Thus, using a correct value for P_B , the biases obtained in this study would most likely be removed.
- 3) Ordinary Least Squares (method 2) continued to give unbiased estimates of **a** and $\hat{S}_B - S_B$, but estimates of **beta** were biased downwards farther due to selection than when sires were random.
- 4) Method 3 was unbiased and similar to the use of the true **beta**, even when sires were selected.

Standard errors on the comparison statistics more than doubled (for all methods) when sires were selected. The results are restricted to situations of equal repeatabilities of the proofs within countries. A summary of the significant results is given in Table 3.12.

3.5 Conclusions

Methods 3 and 4 seem to yield very similar results. Both methods aim at removing the effects of different reliability of proofs in the exporting and the importing countries when estimating **beta**. This is done by utilizing the repeatability of the proof in the importing country.

The parameter **a** which is $G_B - b G_A$, pertain to bulls which are in a certain group in country A and country B. If breeding values of bulls pertaining to different groups in country A are converted, different **a** values are required. The required parameters **a** can be computed by $G_B - b G_A$, if G_A is provided by country A. However, if repeatabilities are $\geq .75$, **a** is well approximated by $P_B - b P_A$.

Ordinary Least Squares is generally underestimating **beta**. Still it is not a bad method, since S_B is estimated unbiasedly. However, with only a slight modification, methods 3 and 4 can be employed just as easily, and yield lower $V(\hat{S}_B - S_B)$.

All methods would suffer if P_{Bj} and P_{Aj} were biased estimates of S_{Bj} and S_{Aj} , respectively. Thus, nonrandom mating of sires to cows and preferential treatment of daughters of bulls could cause unknown problems in estimating **alpha** and **beta**.

For further studies it is suggested that the importance of variable repeatabilities of proofs within countries is illuminated. Furthermore, it would be desirable if the consequences of different definitions of **beta**, as applied in method 1 (used in the simulation study) and as used by **Goddard (1984)**, could be clarified. For practical purposes it seems logic that the prediction equation should aim at the true breeding value in the importing country rather than the proof. Thus, a method utilizing the repeatabilities of proofs in the importing country seems to be adequate. For most practical purposes, however, the differences between the methods proposed by **Wilmink et al. (1985)** and **Goddard (1984)** seem to be small, especially if only bulls with rather reliable proofs are used for international comparisons.

Table 3.1. Results of the simulation study.

Set A: $r_g = 1$, $V(S_B)/V(S_A) = 1$, $w_A \neq w_B$, sires randomly chosen.
Average correlation between \hat{S}_B and $S_B = .8183$, average $w_A = .6666$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	1.000	0.0017	0.3305
2	50.004	0.806**	0.0054	0.3570**
3	50.003	1.002	0.0050	0.3317
4	50.002	1.002	0.0036	0.3316
5	50.003	1.150**	0.0043	0.3472**

** Significantly different from Method 1 at .05 level

Table 3.2. Results of the simulation study.

Set B: $r_g = 1$, $V(S_B)/V(S_A) = 2$, $w_A \neq w_B$, sires randomly chosen.
Average correlation between \hat{S}_B and $S_B = .8135$, average $w_A = .6638$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	1.414	0.0010	0.3383
3	49.998	1.423	-0.0014	0.3395
4	49.995	1.422	-0.0041	0.3396
5	49.996	1.634**	-0.0025	0.3565**

** Significantly different from Method 1 at .05 level

Table 3.3. Results of the simulation study.

Set C: $r_g = 1$, $V(S_B)/V(S_A) = 0.5$, $w_A \neq w_B$, sires randomly chosen.
Average correlation between \hat{S}_B and $S_B = .8167$, average $w_A = .6661$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	0.707	0.0018	0.3331
2	50.004	0.564**	0.0056	0.3616**
3	50.004	0.702	0.0055	0.3344
4	50.004	0.703	0.0056	0.3344
5	50.004	0.805**	0.0055	0.3473**

** Significantly different from Method 1 at .05 level

Table 3.4. Results of the simulation study.

Set D: $r_g = .8$, $V(S_B)/V(S_A) = 1$, $w_A \neq w_B$, sires randomly chosen.
Average correlation between \hat{S}_B and $S_B = .6527$, average $w_A = .6656$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	0.800	-0.0006	0.5741
2	49.997	0.642**	-0.0045	0.5915**
3	49.996	0.799	-0.0044	0.5753
4	49.996	0.800	-0.0047	0.5754
5	49.996	1.102**	-0.0042	0.6375**

** Significantly different from Method 1 at .05 level

Table 3.5. Results of the simulation study.

Set E: $r_g = .8$, $V(S_B)/V(S_A) = 2$, $w_A \neq w_B$, sires randomly chosen.
Average correlation between \hat{S}_B and $S_B = .6550$, average $w_A = .6665$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	1.131	0.0003	0.5711
2	50.016	0.918**	0.0176	0.5872**
3	50.016	1.143	0.0166	0.5725
4	50.015	1.142	0.0153	0.5724
5	50.014	1.549**	0.0116	0.6332**

** Significantly different from Method 1 at .05 level

Table 3.6. Results of the simulation study.

Set F: $r_g = .8$, $V(S_B)/V(S_A) = 0.5$, $w_A \neq w_B$, sires randomly chosen.
Average correlation between \hat{S}_B and $S_B = .6535$, average $w_A = .6657$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	0.566	0.0006	0.5730
2	50.002	0.463**	0.0022	0.5890**
3	50.002	0.577	0.0026	0.5755
4	50.002	0.576	0.0030	0.5755
5	50.003	0.786**	0.0046	0.6419**

** Significantly different from Method 1 at .05 level

Table 3.7. Results of the simulation study.

Set G: $r_g = 1$, $V(S_B)/V(S_A) = 1$, $w_A = w_B$, sires randomly chosen.
 Average correlation between \hat{S}_B and $S_B = .8160$, average $w_A = .6664$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	1.000	-0.0026	0.3342
2	49.998	0.890**	-0.0044	0.3423**
3	49.998	0.992	-0.0047	0.3346
4	49.997	0.990	-0.0054	0.3347
5	49.998	1.119**	-0.0050	0.3449**

** Significantly different from Method 1 at .05 level

Table 3.8. Results of the simulation study.

Set II: $r_g = 1$, $V(S_B)/V(S_A) = 1$, $w_A \neq w_B$, sires selected.
 Average correlation between \hat{S}_B and $S_B = .8173$, average $w_A = .6660$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	1.000	0.0022	0.3321
2	49.988	0.811**	-0.0102	0.3590**
3	49.987	1.010	-0.0101	0.3363
4	49.769**	1.009	-0.2290**	0.3363
5	49.444**	1.623**	-0.5528**	0.6101**

** Significantly different from Method 1 at .05 level

Table 3.9. Results of the simulation study.

Set I: $r_G = 1$, $V(S_B)/V(S_A) = 2$, $w_A \neq w_B$, sires selected.
 Average correlation between \hat{S}_B and $S_B = .8149$, average $w_A = .6653$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	1.414	0.0013	0.3361
2	50.001	1.130**	0.0028	0.3659**
3	50.003	1.405	0.0043	0.3402
4	49.700**	1.404	-0.2986**	0.3407
5	49.208**	2.316**	-0.7927**	0.6268**

** Significantly different from Method 1 at .05 level

Table 3.10. Results of the simulation study.

Set J: $r_G = .8$, $V(S_B)/V(S_A) = 1$, $w_A \neq w_B$, sires selected.
Average correlation between \hat{S}_B and $S_B = .6576$, average $w_A = .6679$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	0.800	0.0029	0.5676
2	50.002	0.642**	0.0051	.5902**
3	50.001	0.801	0.0042	.5754
4	49.826**	0.800	-.1207**	.5755
5	48.480**	2.524**	-1.5203**	2.8556**

** Significantly different from Method 1 at .05 level

Table 3.11. Results of the simulation study.

Set K: $r_g = .8$, $V(S_B)/V(S_A) = 2$, $w_A \neq w_B$, sires selected.
Average correlation between \hat{S}_B and $S_B = .6541$, average $w_A = .6666$

Method	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)/V(S_B)$
1	50.000	1.131	0.0037	0.5722
2	50.040	0.859**	0.0433	0.6014**
3	50.041	1.069	0.0442	0.5804
4	49.804**	1.072	-0.1919**	0.5801
5	47.828**	3.565**	-2.1669**	2.7038**

** Significantly different from Method 1 at .05 level

Table 3.12. Results of the simulation study.

Summary of results in Tables 3.1 to 3.11. Comparisons to Method 1.

Method	Randomly chosen sires				Sires selected			
	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)$	a	b	$\hat{S}_B - S_B$	$V(\hat{S}_B - S_B)$
2		*		*		*		*
3								
4					(*)		(*)	
5		*		*	*	*	*	*

* Significantly poorer than Method 1 at .05 level