Incorporation of external GEBV in the Dutch-Flemish dairy genetic evaluation

H. Eding

[1] *Animal Evaluation Unit, CRV, Rijksweg west 2 6842 BD Arnhem, The Netherlands*

Abstract

Incorporation of external breeding values in an evaluation is a convenient way to increase the information underlying the breeding values from a national evaluation. This provides improved estimates of breeding values of animals with mostly or wholly foreign pedigrees. In genomic analyses external breeding values can be used to increase the reference population. In this paper we present the approach to incorporating foreign or external breeding values taken for the Dutch-Flemish genetic evaluation. It consists of 1) deregression of external breeding values, removing national information to arrive at the deregressed proof containing foreign information only and 2) transformation of the deregressed proof to pseudo-observation records that can be used as 'own observations' in the routine evaluation. Solutions are presented for linear single animal effects models, correlated animal effects models and random regression models. Results are shown of a validating procedure in a random regression test day model for milk production.

Key words: External breeding values, deregression, genomic evaluation

Introduction

Incorporation of external breeding values in an evaluation is a convenient way to increase the information underlying the breeding values from a national evaluation. This provides improved estimates of breeding values of animals with mostly or wholly foreign pedigrees. In genomic analyses external breeding values can be used to increase the reference population. In this paper we present the approach to incorporate foreign or external breeding values taken for the Dutch-Flemish genetic evaluation.

Materials and Methods

The method to incorporate foreign or external BV in routine evaluations consists of 1) deregression of external breeding values, removing national information to arrive at the deregressed proof containing foreign information only and 2) transformation of the deregressed proof to pseudo-observation records that can be used as 'own observations' in the routine evaluation.

Deregression

The method of deregression is based on the work by Pitkänen et al. (2019). It requires the following components:

1. A VanRaden (2009) deregressed proof of the external breeding value (DRP_X) with corresponding expected record contribution (ERC_X)
2. A VanRaden deregressed proof of the national breeding value DRP_N with corresponding ERC_N

Note that for a trait the *expected daughter contributions* EDC and ERC are proportional, such that ERC = k*EDC and $k = (1 - h^2)/(4 - h^2)$. The relative weight of components does not change by using either ERC or EDC to deregress.

The target DRP to include in genetic evaluations is then obtained through:

$$ERC = (ERC_X - ERC_N)$$
$$DRP = [\ DRP_X * ERC_X - DRP_N * ERC_N] / ERC$$

The reliability of the DRP can be obtained by back transforming ERC. Foreign information of a bull is included in the evaluation if the reliability of the DRP is at least 0.10.

Transformation

To derive the correct transformation function it is necessary to distinguish between 1) *m* input traits, 2) *k* analyzed traits and 3) *n* target traits, where 1) input traits are the traits for which a BV or DRP are available, 2) analyzed traits are the traits actually in the evaluation and 3) target traits are the traits for which a pseudo-observation is desired. Usually two or all three trait categories are identical, but particularly in the case of random regression models this may not be the case. For example, in a milk production test day model, the input traits are cumulative 305 day BV, the analyzed traits are the polynomial variables shaping the production curve and the target traits are observations of milk production on a particular day in lactation.

General form

The general form of the transformation function is:

$$\mathbf{o} = \mathbf{Tb}$$

Where **o** is a vector with *n* desired pseudo-observations or target traits, **b** is a *m* size vector with input DRP/BV and **T** is the $n \times m$ transformation matrix derived from the genetic covariance matrix used in the evaluation.

The transformation matrix **T** is obtained through:

$$\mathbf{T} = \mathbf{DCV}^{-1}$$

Where **V** is a $m \times m$ genetic (co)variance matrix of *m* input traits and **C** is a $k \times m$ matrix with covariance between *m* input traits and *k* analyzed traits. The $n \times k$ matrix **D** links analyzed traits with target traits.

To obtain **C** and **V** a $m \times k$ matrix **F** is constructed linking input traits to analyzed traits. Both **C** and **V** are obtained using **F** through:

$$\mathbf{C} = \mathbf{GF'}$$
$$\mathbf{V} = \mathbf{FGF'}$$

Where **G** is the genetic (co)variance matrix for analyzed traits. This is usually the matrix used as parameter in genetic evaluations.

The explicit form of the complete transformation matrix is:

$$\mathbf{T} = (\mathbf{DGF'}) \cdot (\mathbf{FGF'})^{-1}$$

Linear trait breeding value

The most trivial of cases is when input, analyzed and target traits are identical. In that case **F** = **I** and **D** = **I**. If **F** = **I**, then **T** necessarily also is equal to **I**, reducing the transformation function to:

$$\mathbf{o} = \mathbf{Tb} = \mathbf{Ib} = \mathbf{b}$$

If the input trait is an index of underlying traits, for which pseudo-observations are required (with analyzed and target traits identical), we construct a *1 x n* matrix **F** = **w**, where **w** is a vector with index weights of the analyzed traits. For instance, if the input trait is an index of three underlying traits with index weights, such that:

$$\mathbf{C} = \mathbf{GF'} = \mathbf{Gw} = \mathbf{c}$$
$$\mathbf{V} = \mathbf{FGF'} = \mathbf{wGw'} = v$$

Since analyzed and target traits are identical **D** = **I** and can be omitted, reducing the transformation matrix **T** to a vector:

$$\mathbf{o} = \mathbf{Tb} = (\mathbf{CV}^{-1})b = b\mathbf{c}/v$$

The above can be readily extended to include multiple traits by extending **F** with lines for every index trait to be included.

Random regression breeding values

The proper construction of transformation matrix **T** in random regression model is illustrated using a milk production test day model as an example.

Assume a milk production RR model with 3 lactations and 5 Legendre polynomials for 15 analyzed traits in total. The input DRP are based on cumulative 305 day BV. Target traits are (expected) milk productions on day 60 of lactation for each lactation. Assume furthermore that **G** is ordered traits within polynomial (e.g. $poly_1$(lac1…lac5), $poly_2$(lac1…lac5), etc.). Additionally, we assume the presence of a c x l matrix **L** with l Legendre coefficients for each day in a lactation curve of c days in length. In this example $\mathbf{L} = \mathbf{L}_{5:420}$.

The transformation matrices **F** and **D** are of the following form:

$$\mathbf{F} = \mathbf{s} \otimes \mathbf{I}_3$$
$$\mathbf{D} = \mathbf{t} \otimes \mathbf{I}_3$$

Where **s** is a vector with cumulative Legendre coefficients $\sum \mathbf{L}_{5:305}$ and **t** is a vector with Legendre coefficients for DIM = 60, $\mathbf{L}_{60}$ and $\otimes$ denotes the Kronecker product of **s**/**t** and identity matrix **I** of size equal to the number of lactations.

These **F** and **D** are then used to construct **T** to provide pseudo-observations at DIM=60 corresponding to the DRP based on 305 day BV.

Weight or ERC of pseudo-observations

To accurately account for the reliability of the input DRP, weights must be calculated for the observations on the target traits. These can be obtained by transformation of the single trait ERC into corresponding multi trait ERC.

This cannot be done analytically, but a simple iterative procedure to obtain MT-ERC from ST-ERC is the following:

Reliability function

A vector with reliabilities **b** is a function of $\Re(\mathbf{Y}, \mathbf{G}, \mathbf{F})$, which calculates reliabilities according to the MT-ERC matrix **Y** (Liu et al. 2001), using genetic covariance matrix **G** and matrix **F**, where **F** is as before. The function is as follows:

1) $\mathbf{G_t} = \mathbf{G}[\mathbf{I} - (¼\mathbf{YG} + \mathbf{I})^{-1}]$
2) $\mathbf{b} = diag(\mathbf{F'G_tF})/(diag(\mathbf{F'GF})$, where **b** is now a vector with reliabilities for each trait, corresponding to observations enumerated in **Y**, on the diagonal. In this instance the operator / denotes element-by-element division.

Deriving multi-trait ERC from single trait ERC

Let **O** be a diagonal matrix with ERC. With function $\Re$ in place we can iterate on **O** until some convergence conditions are met. Let **b** be the matrix with DRP reliabilities and $\mathbf{b}_i$ the reliabilities from the ERC iterated on ($\mathbf{O}_i$). Since **O** corresponds to number of repeated observation records we assume values on the diagonal of $\mathbf{O}_i$ are integer values.

1) Calculate $\mathbf{Y}_i = 4(\mathbf{O}_i\mathbf{D})'\mathbf{R}^{-1}\mathbf{D}$
2) Calculate $\mathbf{b}_i = \Re(\mathbf{Y}_i, \mathbf{G}, \mathbf{F})$
3) Calculate $\mathbf{t} = (\mathbf{b}_i - \mathbf{b})$
4) Calculate convergence $c = (\mathbf{t't})/(\mathbf{b'b})$
5) Compare $\mathbf{b}_i$ with **b**
6) If $b_i(x) < b(x) - r$: $\mathbf{O}_{i+1}(x,x) = \mathbf{O}_i(x,x) + 1$
7) If $b_i(x) > b(x) + r$: $\mathbf{O}_{i+1}(x,x) = \mathbf{O}_i(x,x) - 1$ (with a certain minimum value, e.g. 1)
8) Repeat until c reaches a threshold value e or **O** stops changing: $\mathbf{O}_{i+1} - \mathbf{O}_i = \mathbf{0}$.

A theshold value can be $e = n*r^2/(\mathbf{b'b})$, where n is the number of elements in **b** and r is the maximum allowed error on $\mathbf{b}_i$ described above.

Validation of the transformation procedure

We applied the procedure described above to the MACE breeding values of a set of 54,194 Eurogenomic bulls for milk, fat and protein and incorporated the resulting pseudo observations in a full conventional milk production evaluation. Input MACE BV were from the Aug. 2023 Interbull evaluation. The full conventional evaluation was based on data for the Dutch/Flemish Dec. 2023 genetic evaluation. Pseudo-observations were fitted as own observations with a separate herd-testday-class. Apart from a random animal effect no other fixed or random effects were fitted.

From the evaluation we selected bulls without national information present in the data, only data from DRP, and compared the input Interbull BV to the BV estimated in the full conventional evaluation.

Results & Discussion

The results are presented in Table 1. Correlations between input BV and output BV were high (~0,99) with comparable standard deviations.

Table 1. Number of selected bulls, standard deviations and correlation *r* of input (MACE) and output EBV for milk production traits.

	N	*std MACE*	*std EBV*	*r*
Milk	37,429	851.6	865.8	0.988
Fat	37,398	31.6	33.2	0.985
Protein	37,413	27.6	26.7	0.990

Bulls were selected which did not have daughter data present in the evaluation, only DRP based on MACE BV.

The procedure outlined above provides a generalized and relatively convenient method of transforming deregressed proofs into pseudo-observations that can be used in a genetic evaluation directly. It precludes the need for the definition of correlated pseudo-traits and the redefinition of the statistical modelling of the evaluation.

The method relies on two matrices, **F** and possibly **D**, which must be constructed explicitly to carry out the transformation. If these are correctly defined in relation to input traits, analyzed traits and target traits, the construction of **T** is straightforward and the transformation of large numbers of DRP or EBV is a relatively simple and fast process.

Conclusions

A generalized approach to derive pseudo-observations from deregressed proofs was presented. The pseudo-observations thus obtained can be used in a genetic evaluation directly, as observations on existing traits, without the need for the definition of additional correlated pseudo-traits. The method is applicable to a variety of models and types of DRP.

References

Liu Z, Reinhardt F., Reents R. 2001. The effective daughter contribution concept applied to multiple trait models for approximating reliability of estimayed breeding values, *Interbull Bulletin* 27.

Pitkänen T, Koivula M., Strandén I. , Aamand G.P., and Mäntysaari E. 2019. Integration of external information from multiple traits into the national multitrait evaluations, *Book of Abstracts of the 70th Annual Meeting*, Ghent, 468

VanRaden, P., Van Tassell, C., Wiggans, G., Sonstegard, T., Schnabel, R., Taylor, J., & Schenkel, F. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92(1), 16-24.