

An Analytical Approach to Sire-by-Year Interactions in Direct and Maternal Genetic Evaluation

C. Lee¹

Laboratory of Statistical Genetics, Institute of Environment & Life Science, Hallym University, Chuncheon, Kangwon-do 200-702, Korea

ABSTRACT: The negative direct-maternal genetic correlation (r_{dm}) for weaning weight is inflated when data are analyzed with model ignoring sire-by-year interactions (SY). An analytical study investigating the consequences of ignoring SY was undertaken. The inflation of negative correlation could be due to a functional relationship of design matrices for additive direct and maternal genetic effects to that for sire effects within which SY effects were nested. It was proven that the maternal genetic

variance was inflated by the amount of reduction for sire variance; the direct genetic variance was inflated by four times the change for maternal genetic variance; and the direct-maternal genetic covariance was deflated by twice the change for maternal genetic variance. The findings were agreed to the results in previous studies.

(Key Words: Variance Components, Genetic Correlation, Weaning Weight)

INTRODUCTION

Additive direct and maternal genetic correlation has been a concern in genetic evaluation of weaning weight (WW) in beef cattle (Garrick et al., 1989; Mallinckrodt et al., 1995; Lee, 1995; Lee and Pollak, 1997a; Lee and Pollak, 1997b; Lee et al., 1997). The correlations estimated from field data were negative for many beef cattle breeds (Pollak et al., 1994). Inflation of negative direct-maternal genetic correlation estimate was caused by ignoring sire-by-year interaction (SY) effects in the United States Simmental cattle data (Lee and Pollak, 1997a). In analysis of data simulated with SY but without genetic correlation, a spurious negative genetic correlation was observed using a model with direct-maternal genetic correlation instead of SY (Robinson, 1994; Lee and Pollak, 1997a). Lee and Pollak (1997a) reported the proportional changes for variance and covariance estimates obtained from ignoring the interaction in real and simulated data sets. This implied intimate relationships among the components from the consistent changes in every component estimate. The objective of this study was to analytically investigate the consequences of ignoring SY in direct and maternal genetic analysis.

ANALYTICAL STUDY

An assumption

According to the national genetic evaluation of WW for Simmental cattle (Garrick et al., 1989), a WW observation (y_i) for animal i can be constituted as $y_i = cg_i + d_i + m_j + p_j + \varepsilon_i$ where cg_i is i^{th} contemporary group effect (usually herd-year-season effect), d_i is additive direct genetic effect for the animal i , m_j is additive maternal genetic effect for dam j of the animal i , p_j is permanent environmental effect for the dam j , and ε_i is residual for the animal i . The residual variance may be partially explained by non-additive genetic, systematic environmental, and their interaction effects. For example, they can be dominance and epistasis effects as established genetic phenomena. Yet these effects were not included in the model. However, let's assume that some notable SY effects contribute to the residual variance, i.e., $\varepsilon_i = sy_k + e_i$ where sy_k is k^{th} SY effect. Sire-by-year interaction represents a partial interaction between genetic effect and contemporary group effect. In the Simmental data, there were small interactions (3% of phenotypic variance) with a sizable effect on the direct-maternal genetic correlation (Lee and Pollak, 1997a). Biological explanation of the interaction has not been clearly understood presently. But the effect may capture some other effect such as preferential treatment and misidentification (Lee and Pollak, 1997b).

This research was supported by a grant from the Hallym Academy of Sciences, Hallym University.

¹ Address reprint requests to C. Lee. (Tel: +82 (0)361-240-1974; Fax: +82 (0)361-241-3422; E-mail: clee@sun.hallym.ac.kr

Received August 23, 1997; Accepted February 11, 1998

Matrix form of sire-by-year interaction model

Based on the scalar model including SY effects, a mixed linear model (MLM) is presented with matrices:

$$y = X\beta + Z_d u_d + Z_m u_m + Z_p u_p + Z_{sy} u_{sy} + e,$$

where y is a vector of WW observations, β is a vector of contemporary group fixed effects, u_d is an unknown random vector of additive direct genetic effects, u_m is an unknown random vector of additive maternal genetic effects, u_p is an unknown random vector of permanent environmental effects, u_{sy} is an unknown random vector of SY interaction effects, and e is an unknown random vector of residuals. The X , Z_d , Z_m , Z_p , and Z_{sy} are known incidence matrices relating observations to β , u_d , u_m , u_p , and u_{sy} , respectively. The first and second moments of random components in the model are $[u_d' u_m' u_p' u_{sy}' e' y']' \sim N([0' 0' 0' 0' 0' (X\beta)']', \Sigma)$, where

$$\Sigma = \begin{bmatrix} A\sigma_d^2 & A\sigma_{dm} & 0 & 0 & 0 & AZ_d'\sigma_d^2 \\ A\sigma_{dm} & A\sigma_m^2 & 0 & 0 & 0 & AZ_m'\sigma_m^2 \\ 0 & 0 & I\sigma_p^2 & 0 & 0 & Z_p'\sigma_p^2 \\ 0 & 0 & 0 & I\sigma_{sy}^2 & 0 & Z_{sy}'\sigma_{sy}^2 \\ 0 & 0 & 0 & 0 & I\sigma_e^2 & I\sigma_e^2 \\ Z_d A\sigma_d^2 & Z_m A\sigma_m^2 & Z_p\sigma_p^2 & Z_{sy}\sigma_{sy}^2 & I\sigma_e^2 & V \end{bmatrix}$$

A is the numerator relationship matrix (NRM), σ_d^2 is the additive direct genetic variance, σ_{dm} is the additive direct and maternal genetic covariance, σ_m^2 is the additive maternal genetic variance, σ_p^2 is the permanent environmental variance, σ_{sy}^2 is the SY interaction variance, σ_e^2 is the residual variance, and $V = Z_d A Z_d' \sigma_d^2 + Z_m A Z_m' \sigma_m^2 + (Z_d A Z_m' + Z_m A Z_d') \sigma_{dm} + Z_p Z_p' \sigma_p^2 + Z_{sy} Z_{sy}' \sigma_{sy}^2 + I \sigma_e^2$. Sire-by-year interaction in this model represents the interaction between paternal effects and year effects because the relationship among sires was not applied in Lee and Pollak (1997a).

Relationship among design matrices

No apparent trend for SY effects was detected in Simmental data. Considering the randomness, expected were some sire effects if the sire effects, instead of SY effects, were included in the model because SY effects are nested within the sire effects. The relationship of the genetic variances and covariance to the sire variance was examined.

Assume a reduced animal model (RAM) including the additional sire effects:

$$y = Z_d^* u_d^* + Z_m^* u_m^* + Z_s^* u_s^* + e^*,$$

where u_d^* is an unknown random vector of additive direct genetic effects for parents only, and Z_d^* is known matrix relating observation to u_d^* . Rows of Z_d^* include, at most, two nonzero elements, .5; as shown by Quaas and Pollak (1980). The u_m^* and u_s^* are unknown random vectors of additive maternal genetic and extra sire effects, respectively. In general, u_s^* contains the effects for only sires, but here include the effects corresponding to all parents. Then the sizes of both matrices, u_m^* and u_s^* , are equivalent to that of u_d^* . The Z_m^* and Z_s^* are the corresponding incidence matrices. All elements in columns of Z_s^* corresponding to the effects for dams are all zeros. The e^* is a random residual vector, and its covariance is not identity when there are unknown parents (Quaas and Pollak, 1980). Under the assumption that all animals with records have known parents, the phenotypic variances can be expressed as:

$$\text{var}(y) = Z_d^* A^* (Z_d^*)' \sigma_d^{*2} + [Z_d^* A^* (Z_m^*)' + Z_m^* A^* (Z_d^*)'] \sigma_{dm}^* + Z_m^* A^* (Z_m^*)' \sigma_m^{*2} + Z_s^* (Z_s^*)' \sigma_s^{*2} + I \sigma_e^{*2}$$

where A^* is reduced NRM (RNRM) excluding nonparent animals.

For the RAM, the design matrices have the following relationship:

$$Z_d^* = 1/2 (Z_m^* + Z_s^*) \text{ or } 2Z_d^* - Z_m^* = Z_s^*.$$

If both sides are squared, then

$$(2Z_d^* - Z_m^*)(2Z_d^* - Z_m^*)' = Z_s^* (Z_s^*)' \\ 4Z_d^* (Z_d^*)' - 2(Z_d^* (Z_m^*)' + Z_m^* (Z_d^*)') + Z_m^* (Z_m^*)' = Z_s^* (Z_s^*)'.$$

The components in the above equation are the coefficient matrices that represent the contribution of variances for the random effects to the total phenotypic variance if RNRM is identity. The phenotypic covariance matrix under the assumption that RNRM is identity is as follows:

$$\text{var}(y) = Z_d^* (Z_d^*)' \sigma_d^{*2} + [Z_d^* (Z_m^*)' + Z_m^* (Z_d^*)'] \sigma_{dm}^* + Z_m^* (Z_m^*)' \sigma_m^{*2} + Z_s^* (Z_s^*)' \sigma_s^{*2} + I \sigma_e^{*2} \\ = Z_d^* (Z_d^*)' \sigma_d^{*2} + [Z_d^* (Z_m^*)' + Z_m^* (Z_d^*)'] \sigma_{dm}^* + Z_m^* (Z_m^*)' \sigma_m^{*2} + \{4Z_d^* (Z_d^*)' - 2[Z_d^* (Z_m^*)' + Z_m^* (Z_d^*)'] + Z_m^* (Z_m^*)'\} \sigma_s^{*2} + I \sigma_e^{*2} \\ = Z_d^* (Z_d^*)' (\sigma_d^{*2} + 4\sigma_s^{*2}) + [Z_d^* (Z_m^*)' + Z_m^* (Z_d^*)'] (\sigma_{dm}^* - 2\sigma_s^{*2}) + Z_m^* (Z_m^*)' (\sigma_m^{*2} + \sigma_s^{*2}) + I \sigma_e^{*2}$$

This implies, when RNRM equals identity or is close to identity, that direct and maternal genetic variances and covariance are confounded with sire variance. A numerical example for the relationships of the matrices

was featured in Lee (1995).

Application to data

Even though sire variances were not estimated in actual and simulated data, the proportional changes in genetic variance and in covariance estimates ignoring SY interaction were given in Lee and Pollak (1997a). The proportional changes were calculated as the changes of component estimates per unit decrease of SY variance estimate. In this study, relative changes in (co) variance components to the change in additive maternal genetic variance were calculated in table 1 with the (co) variance components estimates obtained by Robinson (1994) as well as by Lee and Pollak (1997a). This calculation was to examine the applicability of the findings in this analytical study to results in the previous studies. The figures in table 1 confirmed that the direct genetic variance was inflated by four times the increase for maternal genetic variance. Also confirmed was that the direct-maternal genetic covariance was deflated by two times the increment for maternal genetic variance. This substantiates the relationship of the coefficient matrix for the sire variance to those for genetic (co) variances. Moreover, the sire variances could be approximated from

the formula presented as:

$$\tilde{\sigma}_s^2 \approx (\Delta \sigma_d^2/4 - \Delta \sigma_{dm}^2/2 + \Delta \sigma_m^2)/3$$

where $\Delta \sigma_d^2$, $\Delta \sigma_{dm}^2$, and $\Delta \sigma_m^2$ are the increases caused by ignoring SY effects for direct variance, direct-maternal covariance, and maternal variance, respectively. For example, the approximated sire variances were $9.7 \times \{4.06/4 - (-1.96)/2 + 1\}/3 = 9.7 \text{ kg}^2$ for REAL, $61.9 \times \{4.35/4 - (-2.12)/2 + 1\}/3 = 64.9$ for SIM1, $25.9 \times \{4.20/4 - (-2.07)/2 + 1\}/3 = 26.6$ for SIM2, and $27.3 \times \{3.99/4 - (-2.00)/2 + 1\}/3 = 27.3$ for SIM3 from table 1. The changes in (co) variance component estimates for SIM1 were not as functional as those shown in the other data sets. This might have happened because the RNRM differed more from identity than those in the other data sets. Although similarity of two matrices may not be clearly defined, the determinant or sparsity of an RNRM (or its inverse) gives some indication of how close it is to identity matrix. The determinant of an identity matrix is one, so the determinant of an RNRM is closer to one, indicating RNRM closer to identity matrix. As one of the measures for the sparsity, the ratio of diagonals to off-diagonal elements for inverse of RNRM would be another indication for the similarity to identity matrix.

Table 1. Relative changes^a of variance and covariance components to change of additive maternal genetic variance by ignoring sire-by-year interaction effects in simmental (REAL) and simulated (SIM 1, SIM 2, and SIM 3) data^b

Data	Change of additive maternal variance estimate	Relative change of			
		Additive direct variance	Additive direct and maternal covariance	Permanent environmental variance	Residual variance
REAL	9.7 ^c	4.06	-1.96	.12	-1.19
SIM 1	61.9	4.35	-2.12	-	-2.00
SIM 2	25.9	4.20	-2.07	.00	-1.32
SIM 3	27.3	3.99	-2.00	.00	-1.28

^a Relative change was calculated as the change of component estimate per unit increase of additive maternal genetic variance estimate.

^b REAL; Simmental data collected from 1966 to 1994, SIM 1; Data simulated with zero direct-maternal genetic correlation (These data were simulated without permanent environmental effects and analyzed without those.), SIM 2; Data simulated with zero correlation, and SIM 3; Data simulated with negative (-.32) correlation. (Co) variance components were estimated in REAL, SIM 2, and SIM 3 by Lee and Pollak (1997a) and in SIM 1 by Robinson (1994).

^c The additive maternal genetic variance estimated in Simmental data is in kg².

CONCLUSIONS

Generally, (co) variance component estimates are affected by ignoring any effects in an analytical model. Lee and Pollak (1997a) presented that when ignoring an effect, expected values for a variance of the remaining effect were changed through the ANOVA method. Furthermore, as shown in the current study, biases in genetic parameters can be caused by ignoring some

components in the equation that demonstrates a relationship among the design matrices for random effects. This leads to reduction in genetic progress because of reduction of correlation between estimated and true genetic merit. The analytical study showed that proportional changes in direct and maternal genetic variance and covariance components were expected by ignoring SY interaction effects. Sire effect described in the model was confounded with SY interaction effects

nested within the sire effect. For example, a portion (9.7 kg²) of SY interaction variance (18.7 kg²) in Simmental data was attributed to sire effects due to this nesting.

IMPLICATIONS

This analytical study implied that some overestimation or underestimation may be possible due to ignoring some significant effects in equation that demonstrates a relationship among the design matrices for random effects. Another example can be found in the WW genetic evaluation model where the design matrix for maternal genetic effects is equal to that for permanent environmental effects (Lee, 1995). Ignoring maternal permanent environmental variances led maternal genetic variances to be inflated by about the amount of the permanent environmental variances in weaner greasy fleece weight, weaner body weight, and yearling body weight in Marino sheep (Swan and Hickson, 1994). In a bivariate analysis for heterogeneous components by gender, covariances between genders for each component fit in the model (Lee et al., 1997). However, the restriction of permanent environmental covariance between genders, which is set zero, led to the underestimation of permanent environmental variances and the overestimation of maternal genetic variances for both genders (Lee, 1995). The design matrix in common for both effects is responsible for this compensation.

REFERENCES

- Casella, G. and R. L. Berger. 1990. *Statistical Inference*. Wadsworth & Brooks/Cole. Pacific Grove, CA.
- Garrick, D. J., E. J. Pollak, R. L. Quaas and L. D. Van Vleck. 1989. Variance heterogeneity in direct and maternal weight traits by sex and percent purebred for Simmental-sired calves. *J. Anim. Sci.* 67:2515.
- Lee, C. 1995. A simulation and analytical study of sources affecting genetic correlation estimates for additive direct and maternal genetic effects of weaning weight. Ph. D. Dissertation, Cornell Univ., Ithaca, NY.
- Lee, C. and E. J. Pollak. 1997a. Relationship between sire X year interactions and direct-maternal genetic correlation for weaning weight of Simmental cattle. *J. Anim. Sci.* 75:68.
- Lee, C. and E. J. Pollak. 1997b. Influence of sire misidentification on sire x year interaction variance and direct-maternal genetic covariance for weaning weight in beef cattle. *J. Anim. Sci.* 75:2858.
- Lee, C., C. P. Van Tassell and E. J. Pollak. 1997. Estimation of genetic variance and covariance components for weaning weight in Simmental cattle. *J. Anim. Sci.* 75:325.
- Mallinckrodt, C. H., B. L. Golden and R. M. Bourdon. 1995. The effect of selective reporting on estimates of weaning weight parameters in beef cattle. *J. Anim. Sci.* 73:1264.
- Pollak, E. J., C. S. Wang, B. E. Cunningham, L. Klei and C. P. Van Tassell. 1994. Considerations on the validity of parameters used in national cattle evaluations. In: *Proc. Fourth Genet. Prediction Workshop*, January 21-22, Kansas City, MO, USA.
- Quaas, R. L. and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* 51:1277.
- Robinson, D. L. 1994. Models which might explain negative correlations between direct and maternal genetic effects. *Proc 5th World Congr Genet Appl Livest Prod*, Guelph, 18:378.
- Searle, S. R., G. Casella and C. E. McCulloch. 1992. *Variance Components*. Wiley, New York.
- Swan, A. A. and J. D. Hickson. 1994. Maternal effects in Australian Marinos. *Proc 5th World Congr Genet Appl Livest Prod*, Guelph, 18:143.