Estimation of Variance Components and Genetic Evaluation with Large-scale Dominance Models

Ignacy Misztal
Department of Animal and Dairy Science, University of Georgia, Athens, 30605, USA
ignacy@uga.cc.uga.edu

Tom J. Lawlor
Holstein Association, Brattleboro, VT 05301, USA
lawlor@holstein.com

INTRODUCTION

The genotypic model of inheritance includes additive, dominance and epistatic effects (Cockerham, 1954). Genetic evaluations of farm animals ignore effects other than additive. Subsequently, evaluations are less accurate than they could be, with the loss of accuracy depending upon the variance of the nonadditive effects and on the number of animals with dominance relationships. Considering only the dominance effect, the largest loss of accuracy is for full-sibs. Despite the low number of full-sibs in cattle population, these animals, often results of ET, are the elite of the population. Similarly, the loss of accuracy can occur for young bulls with full brothers under testing. Papers and topics in nonadditive evaluation have recently been reviewed and discussed by Misztal et al. (1996).

In the past, the evaluation with dominance was restricted to pedigrees where the dominance relationship matrix could be created and inverted, i.e., to a few thousands of animals (Henderson, 1989). Discovery of rules to create the inverse of the dominance relationship matrix by Hoeschele and VanRaden (1991) has relaxed some computing limits. However, the rules can lead to a large number of equations, and the mixed model coefficient matrix can be denser than in the additive model. In a study by VanRaden et al. (1992), the evaluation considering dominance took approximately 40 times more CPU time than an additive procedure, and would be too time-consuming for commercial use.

Another reason why dominance was ignored in the evaluation was little knowledge about the magnitude of dominance variation, i.e., the dominance variance. Accurate estimates of dominance variances with the animal model need to be derived from datasets with at least 30,000-100,000 animals for populations with many full-sibs to much larger datasets for cattle populations. The sire model considers only about 1/4 of the dominance information and therefore even larger datasets are required for the same accuracy if the sire model is used. In the past, estimates of the dominance variance were derived either from data sets too small for meaningful results and/or with estimation methods not resistant to bias and/or ignoring most of genetic relationships (e.g., Rodriguez-Almeida et al., 1995; Tempelman and Burnside, 1990, VanRaden et al.,1992). Thus, accurate estimates of the dominance variance were not available.

The first purpose of this study was to develop methodology for 1) genetic evaluation with dominance for very large data sets, 2) estimation of the dominance variance for large data sets. The second purpose was to illustrate this methodology on conformation traits in U.S. Holsteins.

MATERIALS AND METHODS

Complete data included 5.2 million records on 14 conformation traits for all 2,997,312 Holsteins evaluated by the Holstein Association in July 1995. A small data set was created by

selecting 301,818 records of cows with the lowest herd codes. Additionally pedigrees of 218,693 bulls were available. The model was:

$$\mathbf{y} = \mathbf{X\hat{a}} + \mathbf{ZB}\ddot{A} + \mathbf{Za} + \mathbf{ZWf} + \mathbf{Zp} + \mathbf{e}$$

where **y** is vector of records, **â** is vector of management, age at calving and stage of lactation effects, Ä is coefficient of inbreeding depression, **a** is additive animal effect, **f** is parental dominance effect, **p** is permanent environment effect, **B** is vector of inbreeding coefficients, and **X**, **Z**, and **W** are matrices or vectors that relate records to respective effects. The inverse of the parental dominance relationship matrix was created as in Hoeschele and VanRaden (1991), and the system of equations was solved by a modified second-order Jacobi (Misztal, 1987). Variance components were estimated using the Method R (Reverter et al., 1994). This method is applicable to very large data sets, considers the relationship matrices, and returns all estimates in the parameter space. Also, in simulation studies, it has been shown to be resistant to some types of bias (Kaiser and Golden, 1994). Subsampling in Method R results in estimates drawn from the sampling distribution of Method R estimates, and can be used to obtain estimates of the sampling variance (Mallinckrodt et al., 1996). An overview of the complete methodology can be found in Misztal (1996).

Initial tests involved stature and the small data set. Estimates of variance components were obtained with 20 different subsets of the data, where each subset contained approximately 50% of the full data set. Selection of the subset was by a random number generator with a different seed for each subset. Then, the same variance components were estimated for three subsets of the complete data set. Finally, three estimates were obtained for each of the 14 conformation traits with the small data set. The analyses were run on a Sun Ultra 1 workstation.

RESULTS AND DISCUSSION

The graph below shows 20 estimates of the variances obtained with the small data set and three estimates obtained with the complete data set.
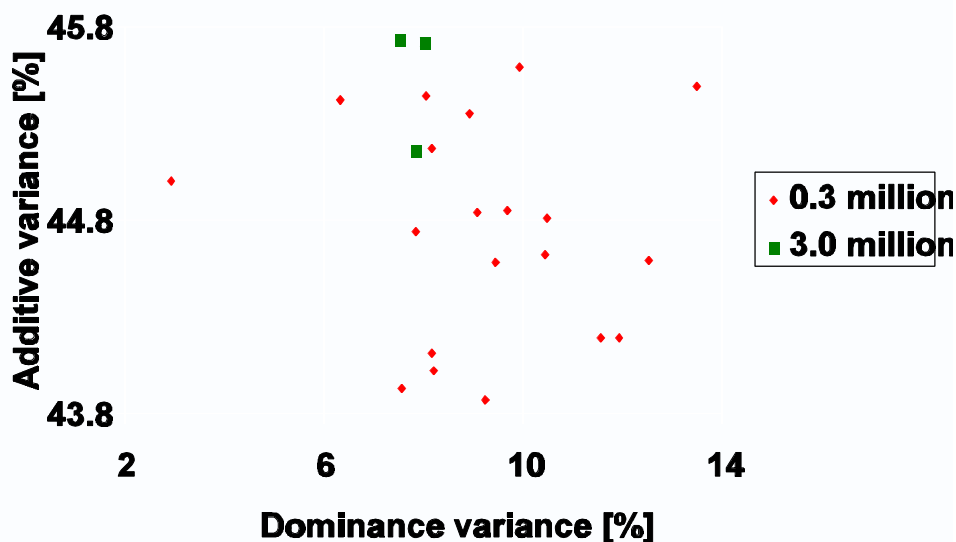
For the small data set, the estimates of the dominance variance had a range of 2.9% to 13.5% of the total variance, with an average of 9.2 ± 2.3. This illustrates a low number of dominance relationships in the data and a large size of data necessary to obtain accurate estimates of the dominance variance in dairy. For the complete data set, the range of the estimates of the dominance was much narrower: 7.5 to 8%, with an average of $7.8 \pm .2$ %. The lower range reflects not only increased data size but also a larger fraction of dominance relationships in the complete data because many full-sib groups were no longer separated. The estimates of the additive variance had a relatively small range in the small data set. Relatively large range for the complete data set is a result of a relaxed convergence criterion in the computing procedures to lower the computing costs.

Estimates of variances for the 14 conformation traits obtained from the small data set are shown in the graph below. The estimates of the dominance variance were in the range of 10-11% for three traits. For two traits, the estimates were in the range of 6-9 %. For the remaining traits, the estimates were below 5%. No clear relationship between the additive and dominance variances is evident; there is a large variation in ratios of estimates of dominance to additive variances. That ratio is 0.46 for dairy form, in the range of 0.32-0.36 for strength, body depth and udder width, and is smaller than 0.23 for all the other traits.
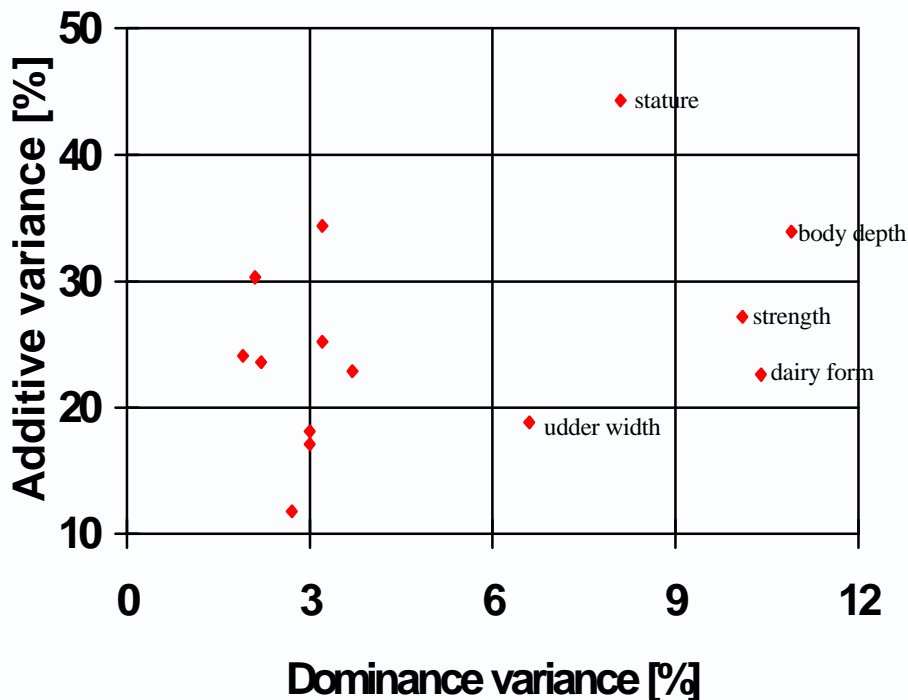
The accurate estimation of the dominance variance may be impossible for small populations because of a large sampling variance. As the inbreeding depression is closely associated with dominance and can be accurately estimated even from small populations, there is a

question whether a dominance variance can be predicted from inbreeding depression.
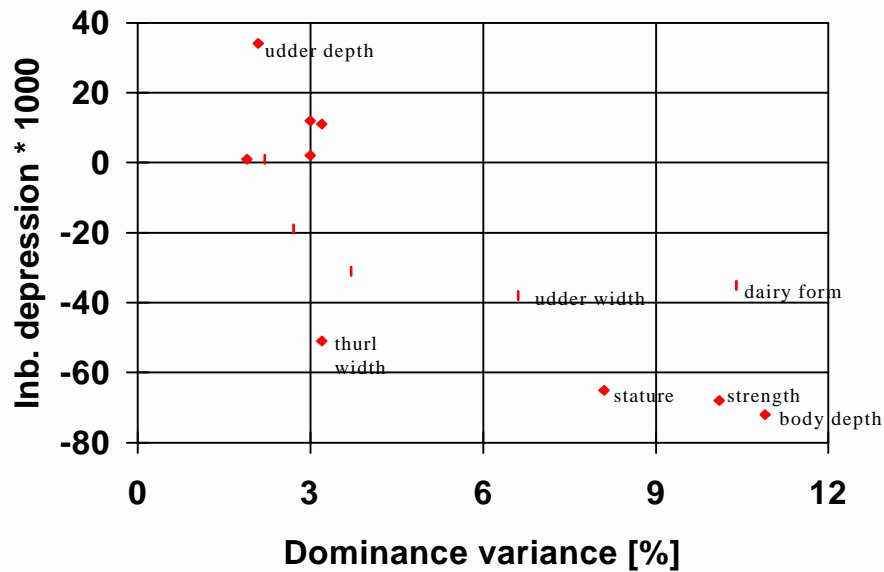
In the graph above, the three traits with the largest estimates of the dominance variance also have the largest negative estimate of the inbreeding depression. The dependence is only statistical, as evidenced by dairy form, which had the estimate of the dominance variance as large as body depth but only half of its estimate of inbreeding depression.

Utilization of the nonadditive evaluations

The genetic merit of a future progeny in the dominance model can be written as:

$$g = (ADD_{sire} + ADD_{dam}) / 2 + DOM_{sire,dam} + INB_{sire,dam}$$

where ADD, DOM and INB are estimates of the additive, parental dominance and inbreeding effects, respectively. Because the number of potential sire x dam combinations can be very high, the dominance and inbreeding information cannot be listed in sires summaries but can be considered in a mating system (DeStefano and Hoeschele, 1993). On input, such a system accepts a list of cows to be mated and selection criteria specified by the breeder, such as weights on individual traits and maximum costs of semen. Then, the system considers mating each cow to a number of sires, computing any nonadditive effects necessary in the process, and selects a combination with the highest overall index. To compute nonadditive adjustments for each potential mating, the mating system would need access to complete results of the recent evaluation and to a complete pedigree. Access to a mating system can now be easily implemented through the World Wide Web.

Computations

Computations with dominance were computationally demanding but feasible. Preparing the dominance pedigree, which included forming and pruning the dominance pedigree classes,

took approximately 40 Mbytes of memory and .5 hr of computing time for the partial data set and 400 Mbytes of memory and 7 hrs of computing time for the complete data set. A typical estimation of variance components took 30 Mbytes of memory and 5 hrs computing for the partial data set and 255 Mbytes of memory and 70 hrs computing for the full data set. A genetic evaluation with a complete data set augmented by later records and a permanent environment effect in the model took about 7 hrs of computing. The currently used additive model would take approximately 55% of the memory and 70% of the computing time. Algorithm optimizations can decrease the CPU and memory requirements; memory limitations are becoming less important due to memory prices falling rapidly.

CONCLUSIONS

Evaluation that ignores dominance may result in inaccurate evaluation. Evaluation with dominance is now feasible computationally for very large models. Also, estimates of the dominance variance can be obtained from complete populations. The decision whether to include the dominance in the model depends on the magnitude of the dominance variance. For conformation traits, the estimates were from 3 to 11% of the total variance, or up to 46% of the estimate of the additive variance.

REFERENCES

Cockerham, C.C.. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859.

DeStefano A.L., Höschele I., 1992. Utilization of dominance variance through mate allocation strategies. J. Dairy Sci. 75:1680.

Henderson C.R., 1989. Prediction of merits of potential matings from sire-maternal grandsire models with nonadditive genetic effects. J Dairy Sci. 72:2592.

Höschele I, VanRaden P.M., 1991. Rapid inversion of dominance relationship matrices for noninbred populations by including sire by dam subclass effects. J. Dairy Sci. 74:557.

Kaiser, C.J. and B.L. Golden. 1994. Heritability estimation and breeding value prediction using selected data. J. Animal Sci. (Abstr).

Mallinckrodt, C.H. B.L. Golden, A. Reverter. 1996. Confidence intervals for Method R estimates of heritability. J. Animal Sci. (Abstr.) 36.

Misztal, I. 1996. Estimation of variance components with large-scale dominance models. J. Dairy Sci. (submitted)

Misztal I., R.L. Fernando, M. Grossman,T. J. Lawlor and M. Łukaszewicz. 1996. Dominance and epistatic effects in genetic evaluation. Animal Science Papers Reports 13:251.

Misztal, I., and D. Gianola. 1987. Indirect solution of mixed model equations. J. Dairy Sci. 70:716.

Reverter A., Golden B.L., Bourdon R.M., 1994. Method R variance components procedure: application on the simple breeding value model. J. Anim. Sci. 72:2247.

Rodriguez-Almeida F.A., Van Vleck L.D., Wilham R.L., Northcutt S.L., 1995. Estimation of non-additive genetic variances in three synthetic lines of beef cattle using an animal model. J. Anim. Sci. 73:1002.

Tempelman R.J., Burnside E.J., 1990. Additive and nonadditive genetic variation for production traits in Canadian Holsteins. J. Dairy Sci. 73:2206.

VanRaden P.M., Lawlor T.J., Short T.H., Höschele I., 1992. Use of reproductive

technology to estimate variances and predict effects of gene interactions. J. Dairy Sci. 75:2892.

Trait	Effect		
	Inbreeding Depression ± 0.001	Variance [% phenotypic]	
		Additive	Dominance
strength	**-0.068**	**27.8 ± 0.2**	**10.1 ± 1.9**
body depth	**-0.072**	**33.9 ± 0.4**	**10.9 ± 1.8**
dairy form	**-0.035**	**22.6 ± 0.5**	**10.4 ± 1.8**
rump angle	**0.011**	**34.4 ± 0.4**	**3.2 ± 1.1**
thurl width	**-0.051**	**25.2 ± 0.0**	**3.2 ± 0.4**
rear leg set	**0.012**	**18.1 ± 0.7**	**3.0 ± 1.4**
foot angle	**-0.019**	**11.8 ± 0.6**	**2.7 ± 2.2**
fore udder att	**0.001**	**23.6 ± 0.5**	**2.2 ± 0.7**
udder height	**-0.031**	**22.9 ± 0.6**	**3.7 ± 1.8**
udder width	**-0.037**	**18.8 ± 0.4**	**6.6 ± 2.2**
udder cleft	**0.002**	**17.1 ± 0.7**	**3.0 ± 0.4**
udder depth	**0.034**	**30.3 ± 0.3**	**2.1 ± 0.5**
front teat placement	**0.001**	**24.1 ± 0.3**	**1.9 ± 0.5**

Data set	Statistics	Effect		
		Variance [% phenotypic]		Inbreeding Depression
		Additive	Dominance	
Complete	Mean ± SD	44.74 ± .56	9.19 ± 2.33	-.0646 ± .0004
	Range	43.87- 45.59	2.92 - 13.48	(-.0640) - (-.0657)
Subset	Mean ± SD	45.53 ± .27	7.79 ± .19	-0.0709 ± .0002

	Ranges	45.16 - 45.73	7.52-8.00	(-0.0712) - (-.0707)