

Estimation of parameters and breeding values

E. J. Pollak and R. L. Quaas
Cornell University

The topic "Estimation of parameters and breeding values" is quite general, allowing for a wide range of problems and analytical procedures to be addressed. To the animal breeder the necessity of obtaining good estimates of population parameters and prediction of breeding values is obvious. However, the best approach in analysis often is not as obvious nor is the correct interpretation of the results obtained. Dr. Jay L. Lush, in his Notes, considers the concept of parameter estimation as two components-- first, a statistical component whereby one obtained solutions, and, second, the interpretation for use of the solutions obtained.

"Estimating the magnitude of a variance component is a purely statistical procedure in describing the variability of a population. Interpreting in genetic or other biological terms the causes of each component, on the other hand, depends on correctly appraising the cause and effect relationships. No set of rules will automatically guarantee that this has been done correctly. It is necessary always to keep a clear distinction between the actually observed variance component as it is derived from the table of analysis of variance and the genetic interpretation of each such component."

It would seem appropriate that a session on estimating parameters and breeding values should contain a mixture of both papers addressing methodology and papers associated with applying methodology to actual data. There are usually two sources of data from which variance components or breeding values

are estimated. The first being data collected at experimental stations, and the second being data collected from producer herds, which are often referred to as field data. There are advantages and disadvantages to using both types of data sets in estimation of parameters or breeding values. In experimental herds, data are usually more reliable, the sources of variation affecting each observation are perhaps better documented, and the environment in which the animal has performed is either better understood or under better control. However, the number of observations available at such units is usually quite small. Also, the question arises as to whether the particular group of animals at that unit is truly representative of the population at large. On the other hand, with field data the data sets themselves can be quite large, in fact, large enough at times to cause problems in using the appropriate analytical procedure to obtain results. Also, the observations themselves are obtained from a wide range of environments. However, the experimenter sometimes finds himself at a loss to be able to explain or appropriately identify the sources of variation influencing these types of records. Hence, in considering papers as examples of application of methodology to real data, both types of data sets should be represented. We feel the papers contained in this session amply reflect the variety of problems which may be encountered in parameter estimation or breeding value prediction as well as original and refreshing suggestions concerning the theory involved.

A fundamental requirement of any parameter estimation or breeding value prediction problem is obtaining accurate information. This does not only mean that the measurements or observations on the animals involved be as

accurate as possible but also that the pedigree information and the identification of the animals involved be accurate. The paper of Christensen et al. addresses the influence of incorrect sire identification of progeny on the estimates of genetic parameters and breeding values. Although the corrections of the problems of misidentification are logistic considerations for the data collection agencies, especially when field data is involved, an appreciation of the potential biases or problems associated with misidentification is necessary. Two papers in this session address the concept of genetic parameter estimation from populations subjected to selection. The paper of Sorensen and Kennedy considers genetic variance estimation when selection is across generations as well as estimation from control populations. On the other hand, the paper of Meyer addresses the problem of genetic parameter estimation from later lactation records in dairy cattle, most probably records obtained as the result of selection on previous lactation performance. Several of the papers in this session address the concept of application of current prediction methodology to the evaluation of animals from a variety of species. The paper of Graser et al. looks at the application of mixed model multiple trait evaluation for records on Australian Simmental. Their multiple trait model contains effects for direct influence of the sire on progeny growth as well as effects of the dam's maternal contribution to progeny growth through weaning as well as postweaning performance. Kennedy considers the application of mixed model methodology for best linear unbiased prediction to swine data. In considering traits such as growth rate and backfat thickness of swine, the goal of this

approach is to combine both progeny test records with individual performance records. The paper of Cady and Burnside is involved with the evaluation of dairy bulls for dystocia and the mortality of their progeny. Both of these traits are categorical in nature, which presents unique problems to the concepts of prediction. The approach of Cady and Burnside is to apply the methodology developed by Quaas and Van Vleck (1980) which considers each category of observations for a trait as a separate trait. The evaluation obtained for each category is the probability of that event for a particular sire's future progeny. The paper of Gianola and Foulley is also concerned with categorical data analysis. In this paper they briefly discuss some of the disadvantages of the current methodology used in the evaluation for categorical data. As an alternative they discuss a nonlinear prediction procedure using as an example a trait with binary expression. All of the papers in this session that are concerned with prediction make the basic assumption that the underlying variance-covariance structure of the data is known. In reality, this is never the case. However, an accepted practice by tradition has been to substitute estimates of the variances and/or covariances into the analyses in order to obtain the evaluation. The paper of Taylor and McClintock presents analytical methods for examining bias and sampling and prediction error variances when estimated covariance matrices are used in the model in place of true values. The remaining two papers in this session are concerned with the estimation of variance components and genetic correlations, using as examples data from swine. Ollivier examines combined estimation method for genetic correlations for data in which one trait is measured on a particular group of

animals and the second trait is measured on an independent set of animals. Haussmann and Reiser, on the other hand, examine the estimation of variance components from swine test data paying particular attention to which models should be used and which methods of estimation can be used.

The purpose of this brief summary of the papers contained in the session "Estimation of parameters and breeding values" was to give the reader, one, an appreciation for the diversity of the material contained in this session, and, two, to show how the material relates to the general theme.