r_elsaid@hotmail.com Session G4.69 Abstract No. 411

Quality of heritability estimates as affected by heritability level, progeny number per sire, type of algorithm, type of model and type of trait

R[*]. Elsaid[1]; M. Elsayed[2]; S. Galal[2], H. Mansour[2], [1]Animal Production, Desert Environment Research Institute, Menofiya University, Sadat City, Egypt, [2]Animal Production, Faculty of Agriculture, Ain Shams University, Cairo, Egypt

The study objectives were to investigate the effect of heritability level, progeny number per sire, type of algorithm, type of model and type of trait on the quality of heritability estimates. Two simulation programs were used, one to simulate a continuous trait and another to modify it into a binary trait. Twelve populations were created (three heritability levels*four levels of progeny number per sire), each with three parities and twenty replicates. Each replicate was analyzed by sire and animal models, using two algorithms (MTDFREML or Gibbs Sampling, GS). Bias and mean squared errors (MSE) of heritability estimates were used to assess the quality of heritability estimates. The effect of heritability level, progeny number per sire, type of algorithm, type of model, type of trait and the interactions on bias and MSE were examined. For estimating variance components, for continuous trait, the animal model was the best in the case of using MTDFREML and GS. For binary trait, within GS, the sire model was the best at heritability equals to 0.1 with progeny number more than 5 whereas, at heritability equals to 0.25 or 0.5 with 20 progeny, the use of animal and sire models were equivalent.

INTRODUCTION

Variance components estimation is always an important tool in developing animal breeding improvement programs. (**Schaeffer, 1984**).

For categorical traits, genetic parameters are usually computed from sire or animal variances and (co)variances. For the estimation of genetic parameters, the heritability estimates resulting from threshold model were higher than those resulting from linear model**;** the primary reason for the difference being that the heritability from the linear model is expressed on the observed scale while habitability from the threshold model is on an underlying liability scale **(Luo *et al,* 2001)**. The same author reported that the threshold animal model using Gibbs Sampling may yield biased estimates, so the threshold sire model is an alternative model for genetic analysis of categorical trait. To test the quality of heritability estimates, the bias and MSE are to be used; an estimator with lower MSE being more precise.

The objective of this study is to investigate the effect of heritability level, number of progeny per sire, type of algorithm, type of model and type of trait on the quality of the heritability estimates as judged by bias and MSE.

Material and Methods

Simulation procedure

A Mont Carlo simulation technique using **SAS (1996)** with assumed mean (0) and variance (1) was used to simulate the continuous trait according to **Analla *et al* (1995)** the following model was used for simulation: $G_k = 0.5\,(S_i + D_j) + X\,\sqrt{0.5h^2\sigma_p^2}$, (Model 1)

G_k is equal to the genetic value of an individual k, a progeny of sire (S_i) and dam (D_j), X is random number taken from normal distribution with mean 0 and variance 1, h^2 is the heritability and σ^2_p is the phenotypic variance.

Parametric phenotypic, genetic, permanent environmental and residual variances used to generate the studied samples were generated according to **Al-shorepy and Notter, 1996** and modified to meet the need of simulating three levels of heritability. With three levels of heritability (0.1, 0.25 and 0.5) and four classes for the number of half-sib progeny per sire (5, 10, 15 and 20), twelve populations were simulated with three levels of parity, as the only fixed effect, and twenty samples (replicates) for each population were generated. Number of records in the simulated population in each of the three levels of parity was 250, 500, 750 and 1000 belonging to 5, 10, 15 and 20 half-sib daughters for each sire. Number of sires used in each case was 50. Each sample was categorized using a random variety from a binomial distribution (RANBIN Function) with **SAS (1996)** to obtain the binary response variable studied. So, two copies of each generated sample were obtained, the first contained the underlying continuous variable and the second contained the binary response.

Statistical analysis

Heritability estimates of the studied variable were estimated for each copy of each sample in the 12 simulated populations (i.e. three levels of heritability * 4 family size), obtained from the animal and sire models using two algorithms (multiple trait animal model program (MTDFREML) proposed by **Boldman *et al* (1995)** and Gibbs Sampling program proposed by **Van Tassell and Van Vleck (1995)**), though only a single trait was analyzed at a time. The linear animal model used for continuous and binary traits was:

$$y = X\beta + Z_a a + Z_c c + e, \qquad \text{(Model 2)}$$

where,

y is the vector of observation; X is the incidence matrix for fixed effects; β is the vector of an overall mean and parity (3 classes); Z is the incidence matrix for random effects; a is the vector of direct genetic effects of cow; c is the vector of permanent environment effects; and e is a vector of random errors normally and independently distributed with zero mean and variance $\sigma^2_e I$.

The linear sire model used for continuous and binary traits was:

$$y = X\beta + Z_s s + Z_c c + e, \qquad \text{(Model 3)}$$

where,

s is the vector of direct genetic effects of sire; and other terms in the model are defined as in model 2.

Two measures were calculated to estimate the correspondence between assumed and estimated values. These are bias and MSE. The estimates of bias in heritability estimates were calculated as the difference between the heritability values obtained from each analysis and the parametric value (**Elsayed, 1997**). The bias was calculated as follows:

$$\text{Bias} = [E(b^R)-B] \qquad \textbf{(Neter } \textbf{\textit{et al,}} \textbf{ 1985)}$$

where,

b^R the expected value of the deviation of the biased estimator from the true parameter B.

The MSE (equals the variance of the estimator plus the squared bias) was calculated as follows : $MSE = E(b^R - B)^2 = \sigma^2(b^R) + [E(b^R) - B]^2$ (**Neter *et al.*, 1985**),

Analysis of variance

Analysis of variance was performed to study the effect of heritability level (0.1, 0.25 and 0.5), number of progeny per sire (5, 10, 15 and 20), type of algorithm (MTDFREML or Gibbs Sampling), type of model (animal or sire model) and type of trait (continuous or binary trait) on the bias and MSE for heritability estimates obtained from the studied samples. To analyze MSE, data were transformed to their square root form to meet the characteristics of hypothesis testing. Preliminary analysis with full model was performed to identify significant terms, then the analysis was repeated with only significant terms retained.

The following model was applied using **SAS (1996)** to analyze the bias**:**

$$Y_{ijklmn} = \mu + h_i + n_j + a_k + m_l + t_m + e_{ijklmn}, \qquad \text{(Model 4)}$$

All possible significant interactions were included in the analysis.

where,

Y_{ijklmn} is the dependent variable of the nth record in the ith heritability, jth number of progeny, kth type of algorithm, lth type of model and mth type of trait; μ the overall mean of bias; h_i the effect of the ith heritability, i=1 to 3; n_j the effect of the jth number of progeny, j=1 to 4; a_k the effect of the k^{th} type of algorithm, k=1 and 2; m_l the effect of the l^{th} type of model, l=1 and 2; t_m the effect of the m^{th} type of trait, m=1 and 2 and e_{ijklmn} the effect of random error, associated with each observation assumed to be normally and independently distributed with 0 mean and variance $I\ \sigma^2_e$.

The same model was used using **SAS (1996)** to analyze the mean squared errors.

RESULTS AND DISCUSSION

All studied effects (heritability level, number of progeny, type of algorithm, type of model and type of trait) had significant effects on bias and MSE ($p<0.0001$).

Based on the results of bias, Figures (1) and (2) show a decision chart for estimating variance components related to continuous and binary traits. For continuous trait, Figure (1) indicates that at any level of h^2 with any number of progeny per sire, the use of MTDFREML by animal model was the best. For binary trait, Figure (2) indicates that generally, the use of GS was appropriate for analysis of binary traits at any level of h^2. This result confirms the findings of **Luo *et al.* (2001)** indicating linear models as inappropriate for analysis of binary response traits. Also, for binary trait within GS at h^2 equals 0.1 or 0.25, the use of sire model was appropriate whereas at h^2 equals to 0.5, the use of animal model and sire model was equivalent. This indicates that threshold single trait sire model could be a good alternative model compared to animal model for genetic analysis of binary traits. This result is in agreement with those reported by **Hoeschele and Tier (1995)** who reported that for categorical traits, because of the extreme category problem in which all observations from some subclasses are in the same category, threshold animal model using GS may yield biased estimates, poor or slow mixing of Gibbs chain, or even "blowing up" of the Gibbs chain. Figures (3) and (4) show the decision chart for estimating variance components for a continuous and binary traits, based on MSE criterion. Figure (3) indicates that for continuous trait at any level of h^2 with any number of progeny per sire, the use of GS by animal model was the best. This result confirms those obtained by **Mousa and Elsayed (2001)** who indicating that GS had consistently smaller MSE than MTDFREML, due to the influence of the prior distribution of the variance components on the posterior distribution. Figure (4) indicates that, generally, the use of GS was appropriate for analysis of binary traits at any level of h^2. Within GS at h^2 equals 0.1, the use of sire model was appropriate whereas at h^2 equals to 0.25, the use of animal model and sire model was equivalent. At h^2 equals to 0.5, the use of animal model was appropriate.

From results of bias and MSE, for binary trait, the conclusion as what methods to use based on bias agreed with those based on MSE, i.e. for the cases of use of sire model by GS at h^2 of 0.1 with 10, 15 and 20 progeny, the use of sire model by MTDFREML at the same level of h^2 with 5 progeny per sire and use of sire model by GS at h^2 equals to 0.25 or 0.5 with 20 progeny per sire (Figure 5) judging based on MSE and bias leads to the same conclusion.

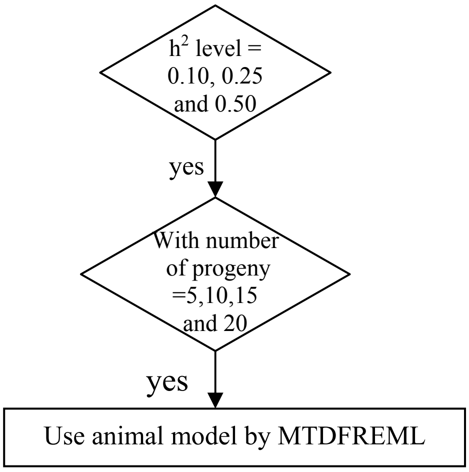

Figure 1. Decision chart for recommending the best model type – algorithm combination to estimate variance components in case of a continuous trait, based on bias.

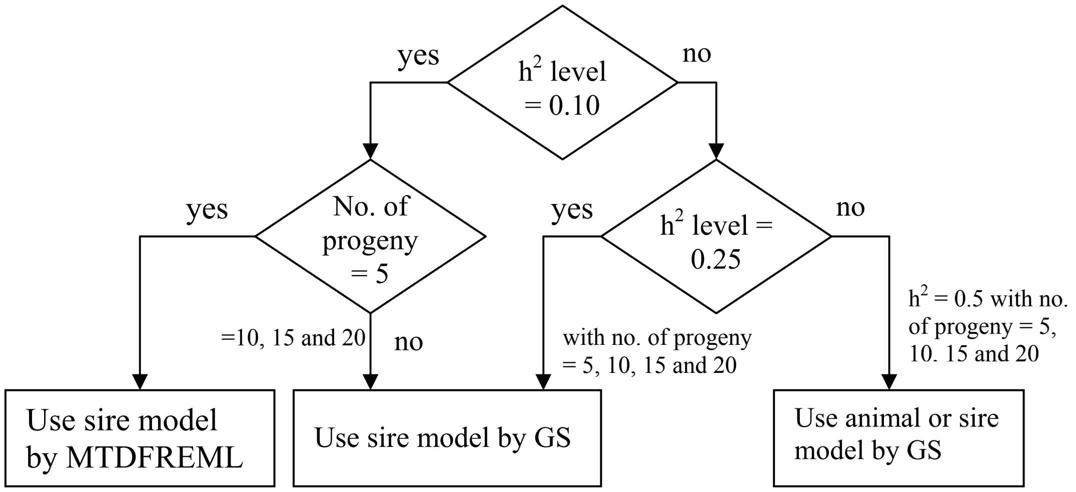

Figure 2. Decision chart for recommending the best model type – algorithm combination to estimate variance components in case of a binary trait, based on bias.

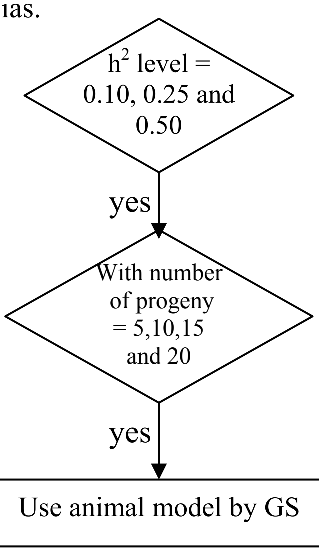

Figure 3. Decision chart for recommending the best model type – algorithm combination to estimate variance components in case of a continuous trait, based on MSE.

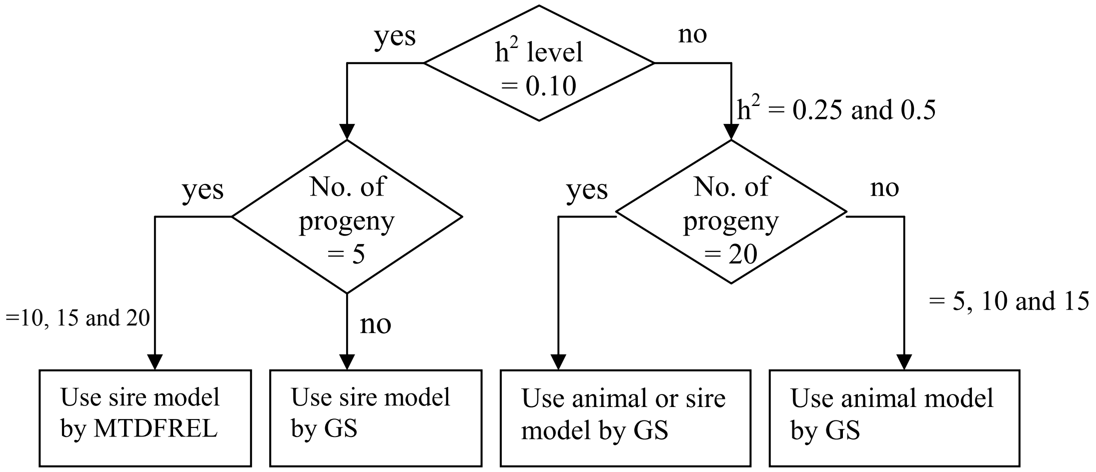

Figure 4. Decision chart for recommending the best model type – algorithm combination to estimate variance components in case of a binary trait, based on MSE.

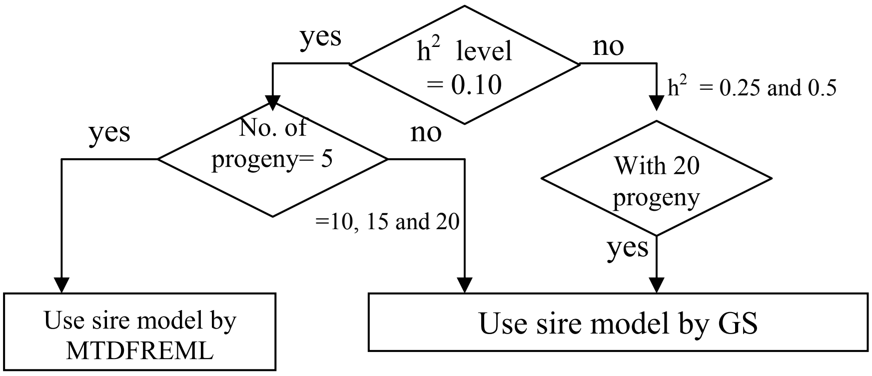

Figure 5. Decision chart for recommending the best model type – algorithm combination to estimate variance components in case of a binary trait, based on bias jointly with MSE.

CONCLUSION

If the aim was to estimate variance components, the conclusion is that for continuous trait, the animal model is the best with MTDFREML or GS at all levels of h^2 with any number of progeny per sire. For binary trait, GS is the best algorithm at all levels of h^2. Within GS, the sire model is the best at low h^2 with any number of progeny more than 5 whereas, at h^2 equals to 0.25 or 0.5 with 20 progeny per sire, the animal model is equivalent to sire model.

REFERENCES

Al-Shorepy, S.A. and D.R. Notter (1996). Genetic variation and covariation for ewe reproduction, lamb growth, and lamb scrotal circumference in a fall-lambing sheep flock. J. Anim. Sci. 74:1490-1498.

Analla, M.; A. Sanchez; A. Munaz and J.M. Serradilla (1995). Simulation analysis with BLUP methodology of different data structures in goat selection schemes in Spain. Small Ruminant Research, 17:51-55.

Boldman, K.G.; L.A. Kriese; L.D. Van Vleck; C.P. Van Tassell and S.D. Kachman (1995). A Manual for Use of MTDFREML. A Set of Programs To Obtain Estimates of Variances and Covariances [Draft]. U.S. Department of Agriculture, Agricultural Research Service. pp 114.

Elsayed, M.M. (1997). Modeling to determine possible problems in application of the animal model in dairy cattle. PhD. Thesis, Fac. Agric., Ain Shams Univ., Egypt, p. 24.

Hoeschele, I. and B. Tier (1995). Estimation of variance components of threshold characters by marginal posterior modes and means via Gibbs sampling. Genet. Sel. Evol. 27:519:540.

Luo, M.F.; P.J. Boettcher; L.R. Schaeffer and J.C.M. Dekkerst (2001). Bayesian inference for categorical traits with an application to variance component estimation. J. Dairy. Sci. 84:694-704.

Mousa, E., and M. Elsayed. 2001. Quality of variance components estimated by Gibbs sampling in populations with and without selection and varying heritability. Egyptian J. Anim. Prod. 38:97-110.

Neter, J.; W. Wasserman and H. Kutner (1985). Applied linear statistical model regression, analysis of variance and experimental designs. Second edition Library of Congress Catalog. Card No. 84:395.

SAS (1996). Statistical Analysis System, SAS User's Guide : Statistics. SAS Institute Inc. Editors, Cary, NC.

Schaeffer, L.R. (1984). Sire and cow evaluation under multiple trait models. J. Dairy Sci. 67:1567-1580.

Van Tassell, C.P. and L.D. Van Vleck (1995). A Manual for Use of MTGSAM. A Set of FORTRAN Programs To Apply Gibbs Sampling to Animal Models for Variance Component Estimation [Draft]. U.S. Department of Agriculture, Agricultural Research Service. pp 86.