GENETIC ANALYSIS OF NUMBER OF SERVICES PER CONCEPTION AND DAYS OPEN AS CONTINUOUS AND BINARY TRAITS USING GIBBS SAMPLING AND REML IN HOLSTEIN COWS

E. Mousa[1], U.M. EL-Saied[2] and M.A.M. Ibrahim[3]

1- Department of Animal Production, Faculty of Agriculture, Assiut University, Assiut, Egypt, 2- Animal Production Research Institute, Ministry of Agriculture, Dokki, Giza, Egypt, 3- Department of Animal Production, Faculty of Agriculture, Cairo University, Giza, Egypt

SUMMARY

Variance components were estimated for a total of 3652 lactation records obtained from 1501 Holstein cows raised in a single commercial farm, using two procedures Multiple Trait Derivative Free-REML (MTDFREML) via restricted maximum likelihood and Gibbs Sampling via Baysian analysis with a multiple-trait repeatability animal model. The model included year-season of calving (18 groups) and parity (first three lactations) as fixed effects and additive genetic and permanent environmental as random effects.

Records were analyzed to compare estimates for variance components for number of services per conception (NS) and days open (DO) when measured once as continuous traits and another as binary traits through the two previous procedures.

When NS and DO were considered as continuous traits, heritability estimates were 0.010 and 0.030 through MTDREML and 0.034 ± 0.023 and 0.054 ± 0.014 using GS, respectively. Repeatability estimates were 0.048 and 0.069 via MTDREML and 0.054 ± 0.019 and 0.060 ± 0.022 through GS, respectively. Genetic and phenotypic correlations between the two traits were 0.680 and .770 using MTDFREML and .680 and .740 using Gibbs Sampling (GS), respectively.

When the two traits were considered as binary traits heritability estimates were 0.003 and 0.030 through MTDREML and 0.067 ± 0.030 and 0.062 ± 0.030 using GS, respectively. Repeatability estimates were 0.047 and 0.035 via MTDREML and 0.175 ± 0.012 and 0.127 ± 0.007 through GS, respectively. Genetic and phenotypic correlations between the two traits were 0.350 and 1.000 , respectively ,through MTDFREML and 0.990 and 0.310, respectively, via GS.

The obtained results revealed that differences in the estimates are due to both the form considered for the trait (continuous or binary) and the employed methodology. The size of the variance components plays an important role in determining these differences.

Keywords: Holstein-Friesian, genetic parameters, REML, Gibbs Sampling

INTRODUCTION

Estimation of variance components (VC) has long been an important aspect for genetic improvement of production traits. Several traits of interest in animal breeding are measured as binary variables. Analysis of such variables by linear methodology violates several assumptions of linear model and is not optimal (Gianola, 1982). A more satisfactory method is based on the threshold model concept (Wright, 1934). Recently, Gibbs Sampling (GS) has been extended to Bayesian inference in threshold model for binary data (Albert and Chib, 1993; Sorensen *et al.*, 1995). Statistical treatment of the threshold model has been developed for the animal model by Wang *et al.* (1994, 1997) and Van Tassell *et al.* (1995). The multiple-trait Gibbs Samples for animal model (MTGSAM) program has been developed to implement the GS algorithm for Bayesian analysis of a broad range of animal models (Van Tassell and Van Vleck, 1995,1996).

The purpose of this study was to compare posterior mean estimates using GS for genetic parameters with those obtained using MTDFREML when a pair of traits was analyzed as continuous and binary traits.

MATERIALS AND METHODS

Data on reproductive performance of 1501 lactating Holstein cows from a single commercial farm (International Company for Animal Wealth) in Giza governorate, Egypt were collected during the period from 1991 to 1998.

Measurements of reproductive performance included days open (DO) and number of services per conception (NS). Data included 3652 records for the1501 Holstein cows which were daughters of 220 sires. These cows were distributed over 18 years-season groups. Most of cows were imported as pregnant heifers from the U.S.A. Cows were artificially inseminated at the first observed estrus after uterine involution using frozen semen imported from the U.S.A. Each year was divided into two seasons: warm (March to August) and mild (September to February).

Model

The MTDFREML programs of Boldman *et al.* (1995) were used to obtain REML estimates of (co)variance components. Principles of derivative-free restricted maximum likelihood have been described by Smith and Graser (1986) and Meyer (1989) and reviewed by Boldman *et al.* (1995).

As an alternative method to estimate variance components (VC) through REML, GS is investigated here. GS is one method in a larger class of methods, referred to Monte Carlo Marcov Chain methods. Gibbs Sampling is based on Bayesian methods for estimation of VC but is evaluated based on its frequents.

The general representation of the multiple-trait repeatability animal model is the following:

Where:

$$Y = xb + Z_a a + Z_e c + e$$

Y   is a 'n by 1 vector of records,

b   Denotes the fixed effects in the model with association matrix X;

a   is the vector of direct animal genetic effects with association matrix Za;

c   is the vector of permanent environmental effects of the animal with association matrix Z_c; and

e   denotes the vector of residual effects.

The following multiple-trait repeatability animal model was used to obtain (co)variance components for the studied traits (NS and DO) through REML and GS procedures:

$$y_{ijkl} = A_i + PE_i + ys_j + p_k + e_{ijkl}$$

where:

y_{ijkl} = records of trait 1 (DO or NS) for the kth parity of the jth year-season of calving of the ith animal;

A_i = the random effect of the additive genetic effect of the animal;

PE_i = the random permanent environmental effect on the animal;

ys_j = the fixed effect of the jth year-season of calving (18 levels);

p_k = the fixed effect of the kth parity (3 levels representing the first three parities);

and

e_{ijkl} = the random residual effect associated with each observation.

The variance-covariance structure for the model is as follows:

$$V = \begin{bmatrix} a_1 \\ a_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\sigma^2 a_1 & \sigma a_1 a_2 & O & & O \\ \sigma a_2 a_1 & A\sigma^2 a_2 & & & \\ & & I\sigma^2 c_1 & \sigma c_1 c_2 & O \\ O & & \sigma c_2 c_1 & I\sigma^2 c_2 & \\ & & & & In_1\sigma^2 e_1 & \sigma e_1 e_2 \\ O & & O & & \sigma e_2 e_1 & In_2\sigma^2 e_2 \end{bmatrix}$$

Where:

A   is the numerator relationship matrix;

$\sigma^2 a_1$, $\sigma^2 a_2$ are the direct genetic variance for traits 1 and 2;

$\sigma^2 c_1$, $\sigma^2 c_2$ are the variance due to permanent environmental effects;

In_1 , In_2 is an identity matrix of order equal to the records of trait 1 and 2;

$\sigma a_1 a_2$ is the direct genetic covariance between the two traits;

$\sigma c_1 c_2$ is the permanent environmental covariance between the two traits; and

$\sigma e_1 e_2$ is the error covariance between the two traits.

The convergence criterion used in this analysis was the variance of likelihood in the current set of simplex value used in the derivative-free REML algorithm (Boldman *et al.*, 1995). The desired range of this variance at convergence is 10^{-6} to

10^{-10}. According to the software authors (Boldman *et al.*, 1995) the estimation programs should be restarted repeatedly with the converged values and no prior simplex information until convergence occurs to the same values that were supplied as starting values. This approach should ensure that a global, rather than a local, maximum likelihood estimate is obtained. All estimates were based on restarted system where a Gibbs chain length of 250,000 was run for each trait of each case with burn in 10,000 rounds and were considered to be an effective number of rounds (Mousa and Van Vleck, 1998). In order to create the binary form of NS and DO from its continuous form (NS and DO), each of them was classified into two categories depending on the normal range of each of them for this breed. Cows conceived after one, two or three services (79%) were considered to be normal and therefore they were formed as the first category. While, cows conceiving after 4 services or more (21%) were formed as the second category. Similarly, cows with DO less than 120 d (33%) were considered to be normal forming the first category and those with more than 120 d (67%) formed the second .

RESULTS AND DISCUSSION

Preliminary least squares analysis using PROC GLM option of SAS (SAS, 1990) showed highly significant effect of year-season of calving on the two traits (DO and NS). The mean effect for DO (220 ± 3.4 d) and NS (2.8 ± 0.05 services) were clearly higher than the values frequently reported in the literature for the same breed. Detailed results were presented by Abou-Bakr *et al.* (2000).

Table 1 shows heritability and repeatability estimates for the studied traits. For the continuous form of the two traits, estimates of heritability by GS were 3.4 and 1.8 times as those by MTDFREML for NS and DO, respectively. It is clear that estimates from both procedures were not comparable except for the repeatability estimate for DO.

Table 1. MTDFREML and GS heritability ($h^2 \pm SD$) and repeatability ($t \pm SD$) estimates for days open (DO) and number of services per conception (NS) of Holstein cows as continuous and binary traits

	MTDFREML		0	
	h^2	t	h^2	t
NS as a continuous trait	0.010	0.048	0.034 ± 0.023	0.054 ± 0.019
NS as a binary trait	0.003	0.047	0.067 ± 0.030	0.175 ± 0.012
DO as a continuous trait	0.030	0.069	0.054 ± 0.014	0.060 ± 0.022
DO as a binary trait	0.030	0.035	0.062 ± 0.030	0.127 ± 0.007

The same trend was observed , but with more differences, for the binary form of the traits as for its continuous form except for the heritability estimate for DO and the repeatability estimate for NS. Gibbs Sampling estimates for heritability and repeatability were 22.33 and 3.7 times as those by MTDFREML for NS and 2.07 and 3.6 times for DO, respectively.

In summary, the GS and MTDFREML estimates were quite different from each other, but the differences were higher in the binary form than for the continuous form of the traits.

From these results, it appears that GS had consistently higher estimates of VC than MTDFREML probably due to the influence of the prior distribution of the variance component on the postrior distribution. These results are in good agreement with those observed in a pervious study (Van Tassel, 1994). Mousa and El-Sayed (2001) added that the effect of prior distribution decreases as the heritability or the amount of information increase. This means that, the difference between estimates from the two procedures is large when the heritability estimate is low and the distribution of this trait is not continuous. The advantage of GS is the possibility to analyse the trait under the threshold model. Therefore, it appeared to be relatively unbiased, that is none of estimators seemed to be superior based on biasedness (Van Tassel, 1994 and Mousa and El-Sayed, 2001).

For the continuous form of the studied traits, estimates of heritability for DO from MTDFREML (0.030) and GS (0.054) are comparable to other estimates reported in the literature (Berger *et al.*,1981; Hayes *et al.*, 1992, Marti and Funk, 1994, Rekaya, *et al.*, 1996, Dematawewa and Berger, 1998 and El-Saied *et al.,* 2000).

MTDFREML estimate for heritability of NS was 0.010 and that form GS was 0.034 as compared to previous study estimates of 0.008 (Moore *et al.*, 1990) and 0.060 (Berger *et al.*, 1981). Therefore, the direct genetic selection to improve this trait would be meaningless.

The small additive genetic variance of NS (0.0007) could render the genetic correlation between this trait and DO in this study low.

Phenotypic and genetic correlations between NS and DO were 0.680 and 0.770 by MTDFREML and 0.680 and 0.470 by GS considering the two traits as continuous traits. For the binary form of the two traits, the corresponding estimates by MTDFREML were 0.350 and 1.000, while by GS they were 0.990 and 0.310 for phenotypic and genetic correlation, respectively. Although values for phenotypic and genetic correlations between the two traits differ in their respective values from one procedure to another and from continuous to binary form of the trait, the same positive relationship between the two traits existed. This trend is in agreement with the results of Dematawewa and Berger (1998) for the same breed analyzing the continuous form. These results indicate that the longer the DO the higher the NS.

CONCLUSION

A clear difference was observed between the estimates of genetic parameters from the two procedures. The differences between the estimates by the two procedures (large or small) depend on the distribution of the considered trait and on the size of its variance components. Therefore, if the variance component is large and the trait is continuous the estimates by the two procedures would be very similar. In contrast, the binary traits with small variance components create more difference between the two procedures. This is due to that the GS is adding a prior distribution to the final estimates of variance components i.e. the posterior distribution.

ACKNOWLEDGMENT

REFERENCES

Abou-Bakr, S., U.M. El-Saied and M.A.M. Ibrahim 2000. Genetic and phenotypic parameters for milk yield days open and number of services per conception of Holstein cows of commercial herd in Egypt. Egyptian Journal of Animal Production 37 (1): 9-17.

Albert, J.H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. 88: 669-679.

Berger, P.J., R.D. Shanks; A.E. Freeman and R.C. Laben. 1981. Genetic aspects of milk yield and reproductive performance. J. Dairy Sci., 64 (1): 114-122.

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 variance and covariances [DRAFT]. U.S. department of Agriculture, Agricultural Research Services. pp 114, 120.

Dematawewa, C.M.B. and P.J. Berger 1998. Genetic and phenotypic parameters for 305 milk yield, fertility and survival in Holsteins. J. Dairy Sci., 81 (10): 2700-2709.

El-Saied, U.M.; E. Mousa; S. Abou-Bakr. 2000. Genetic parameters for milk yield and some reproductive traits of Holstein cows using REML and Gibbs sampling. Proceeding of the 3rd All Africa Conference on Animal Agriculture and 11th Conference of the Egyptian Society of Animal Production, Alexandria, Egypt 6-9 November 2000, in press.

Gianola, D. 1982. Theory and analysis of threshold characters. J. Animal Sci. 54: 1079-1096.

Hayes, J.F.; R.I. Cui and H.G. Monardes 1992. Estimates of repeatability of reproductive measures in Canadian Holsteins. J. Dairy Sci., 75 (6): 1701-1706.

Marti, C.F. and D.A. Funk. 1994. Relationship between production and days open at different levels of herd production. J. Dairy Sci., 77(6): 1682-1690.

Meyer, R. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm Genet Sel. Evol. 21: 317-340.

Moore, R.K., B.W. Kennedy; L.R. Schaeffer and J.E. Moxley 1990. Relationships between reproduction traits age and body weight at calving and days dry in first lactation of Ayrshires and Holsteins. J. Dairy Sci., 73(3): 835-842.

Mousa, E. 1999. Genetic analysis for growth and reproductive traits in sheep data using restricted maximum likelihood and Gibbs sampling. Ph.D. Thesis, Assiut. Univ., Assiut, Egypt.

Mousa, E. And El-Sayed. 2001. The use of gibbs sampling for variance component estimation with simulated three level of heretabilities under random mating and selection. Prepared manuscript.

Mousa, E.; L.D. Van Vleck. 1998. Effective number of rounds of Gibbs sampler needed for estimation of parameters for threshold and continuous traits. Egyptian Journal Animal Production, 35, Suppl. Issue, Dec. 493-505.

Rekaya, R., R. Alenda and M. Carabano. 1996. Days open and milk production in Friesian Cattle: A genetic study. Investigacion-Agraria, Production-Y-Sanidad-Animales. 11 (3): 253-266.

SAS, 1990. SAS statistics. Guide release 6.03. edition SAS Inst., Inc., Cary NC. USA.

Smith, S.P., and H.U. Graser, 1986. Estimating of Variance components in a class of mixed models by restricted maximum likelihood J. Dairy Sci. 69: 1156-1165.

Sorensen, D.A., S. Anderson, D. Gianola, and I. Korsgaad 1995. Bayesian inference in Threshold models using Gibbs sampling. Genet Sel. Evol. 27: 229-249.

Van Tassel 1994. The use of Gibbs Sampling for variance component estimation with simulated and weaning weight data using animal and maternal effects models. Ph.D. Diss., Cornell Univ. Ithaca, NY. pp 93-95.

Van Tassel , C.P. and L.D. Van Vleck 1996. Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co) variance component inference. J. Anim. Sci 74:2586-2597.

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 Agriculture, Agriculture Research Service. pp. 86.

Van Tassell, C.P., G. Casella and E.J. Pollak 1995. Effects of selection on estimation on estimates of variance components using Gibbs sampling and restricted maximum likelihood J. Dairy Sci. 78: 678-629.

Wang, C., S., J.J. Rutledge and D. Gianola 1994. Baysian analysis of mixed linear model via Gibbs sampling with an application to litter size in Iberian pigs. Genet Sel. Evol. 26: 91-115.

Wang, C.S., R.L. Quaas, and E.J. Pollak 1997. Baysian analysis of calving ease scores and birth weight. Genet Sel. Evol. 29: 117-143.

Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of Guinea pigs Genetics 19: 506-536.

تحليلات وراثية لصفتي عدد التلقيحات اللازمة للإخصاب و الفترة المفتوحــة كصفـــات سّلمية ومستمرة بإستخدام طريقتي جبس و التشابه العظمى المحددة لأبقار الهولستين

عمـــاد موسى[1]،   أســـامة محمد الســعيد[2]، محمد عبد العزيز محمد إبراهيم[3]

١- قسم الإنتاج الحيواني، كلية الزراعة ، جامعة أسيوط، أسيوط – مصــر – مصــر، ٢- معــهد بحــوث الإنتـاج الحيواني، الدقى ، الجيزة   مصر، ٣- قسم الإنتاج الحيواني، كلية الزراعة، جامعة القـاهرة، الجــيزة – مصر

أجريت الدراسة على ٣٦٥٢ سجل تم جمعها من سنة ١٩٩١ إلى سنة ١٩٩٨ من ١٥٠١ بقـرة هولستين فريزيان تابعة لمزرعة تجارية بهدف تقدير المعالم الورثية و المظهرية لكل مــن صفـة عدد التلقيحات اللازمة للإخصاب والفترة المفتوحة باستخدام طريقتى جبــس و التشابه العظمــى المحددة، وذلك عندما قيست هاتين الصفتين مرة كصفات مستمرة ومرة أخرى كصفات ســلمية. اشتمل النموذج الإحصائي على موسم وسنة الولادة (١٨ مجموعة) و ترتيب موسم الوضـــع (أول ثلاثة مواسم) وذلك كتأثير ثابت بينما استخدم التأثير الوراثي التجميعي و البيئــي الدائــم كتــأثير عشوائى .

عندما اعتبرت صفتا عدد التلقيحات اللازمة للحمل والفترة المفتوحة صفـات مســتمرة قـدر المكافئ الوراثي باستخدام طريقة MTDFREML بنحو ٠,٠١ و ٠,٠٣٠ أما باستخدام طريقــة جبس فكانت التقديرات ٠,٠٣٤ ± ٠,٢٣ و ٠,٠٥٤ ± ٠,٠١٤ على الترتيب، بينمــا كــانت قيـم المعامل التكــراري ٠,٠٤٨ و ٠,٠٦٩ بالطريقــة الأولــى و ٠,٠٥٤ ± ٠,٠١٩ و ٠,٠٦٠ ± ٠,٠٢٢ بالطريقة الثانية على الترتيب. و كانت تقديرات الارتباط المظهرى و الوراثى بين الصفتين ٠,٦٨٠ و ٠,٧٧٠ باستخدام طريقة MTDFREML و ٠,٦٨٠ و ٠,٤٧٠ باستخدام طريقة جبس على الترتيب.

و عند تحليل نفس الصفتين كصفات سلمية كانت تقديرات المكافئ الوراثى باســتخدام طريقــة MTDFREML ٠,٠٠٣ و ٠,٠٣٠ أما باستخدام طريقة جبــس فكانـت التقديــرات ٠,٠٦٧ ± ٠,٠٣٠ و ٠,٠٦٢ ± ٠,٠٣٠ على الترتيب، بينما كانت تقديرات المعــامل التكــراري ٠,٠٤٧ و ٠,٠٣٥ بالطريقة الأولــى و ٠,١٧٥ ± ٠,٠١٢ و ٠,١٢٧ ± ٠,٠٠٧ بالطريقــة الثانيــة علــى الترتيب. و كانت تقديرات الارتباط المظهرى و الوراثى بين الصفتين ٠,٣٥٠ و ١,٠٠٠ باستخدام طريقة MTDFREML و ٠,٩٩٠ و ٠,٣١٠ باستخدام طريقة جبس على الترتيب.

يمكن أن نستنتج من ذلك أن الاختلافات المشاهدة بين تقديرات المعالم الوراثية تخضـع لكل مـن أسلوب التحليل و طبيعة الصفة و أن لقيم مكونات التباين دور هام فى التأثير علــــى حجـم هـذه الاختلافات.