The Journal of Animal & Plant Sciences, 23(6): 2013, Page: 1583-1587
ISSN: 1018-7081

COMPARING COVARIANCE STRUCTURES using DIFFERENT OPTIMIZATION TECHNIQUES in GLMM on SOME SEXUAL BEHAVIORS of MALE LAMBS

G. Ser, A. Yeşilova and A. Yılmaz*

Department of Animal Science, Faculty of Agriculture, Yuzuncu Yıl University, Van, Turkey
*Hizan Vocational College, Bitlis Eren University, Bitlis, Turkey
Corresponding author e-mail: gazelser@yyu.edu.tr

ABSTRACT

This study is concerned with use of generalized linear mixed models (GLMM) to analyse the repeated measurements based on count data obtained from the sexual behaviors of male lambs. A combination of different optimization techniques and covariance structures were applied to four constructed models. These models were defined in terms of random effect specifications. Therefore, residuals was assumed to be random (Model A), intercept assumed to be random effect (Model B), time (slope) assumed to be random effect (Model C) and both intercept and time assumed to be random effects (Model D). Five different techniques quasi-newton (QUANEW), newton-raphson (NEWRAP), trust region (TRUREG), newton-raphson ridge (NRRIDG) and double-dogleg (DBLDOG) optimization techniques were used for analyzing these models. Three different covariance structures compound symmetry (CS), unstructured (UN) and first-order autoregressive (AR(1)) were used. In conclusion, based on likelihood criteria, the Model A with CS structure outperformed other models for the repeated measurement data of sexual behavior characteristics.

Key words: Norduz male lambs, optimization techniques, subject specific model

INTRODUCTION

Mixed model methodology is essential to properly identify the variance-covariance structure among the data in the analysis of the repeated measures (Akbaş *et al.,* 2001; Eyduran and Akbaş 2010; Orhan *et al.,* 2010).

If a dependent variable in a data set is normally distributed, population-averaged model, general linear model, subject-specific model and general linear mixed model are used in combination with the standard models. However, in cases when the dependent variable is not normal distributed, population-averaged models and as an extension of these models, Generalized Linear Model (GLM) and Generalized Estimation Equation (GEE) are used. At same time, in subject-specific models, Generalized Linear Mixed Models (GLMM) are used (Singer and Willet 2003).

Generalized Linear Model makes an estimation combining likelihood based on approaches with regression analysis for random variables. While doing this, GLM defines the distribution of random variables in the form of exponential distribution uses the function of expected values of the variables instead of the variables themself (Dobson, 1990; McCulloch and Searle 2001).

Generalized Linear Mixed Model is an extension of GLM, consists of the random effects in addition to fixed effect. Thus, GLMM makes parameter estimates for both fixed and random effects in the model. At the same time, GLMM is practical because it takes into account the over-dispersion which is often observed among the random variables and it models the dependency among response variables specific to repeated measurements (McCullah and Nelder 1989).

MATERIALS AND METHODS

Data used in this study were obtained Norduz male lambs which were measured at different age periods. Study material consisted of total of 32 Norduz male lambs. The studies of determination sexual behaviors of the male lambs were started when the male lambs were 6- months old. Sexual behaviors tests were done once a fortnight until they were 12 months old and between 12 and 13 just one test was carried out. Using a data set of sexual behaviors of male lambs, four different models were constructed in this study. The vocalization were included to the models as dependent variable, while the frequency of mount, the response of flehmen, the weight of male lambs, the anogenital sniffing, the raising of tail and duration of mount were integrated as independent variables to the model.

Generalized Linear Mixed Models: Generalized Linear Mixed Model is an extension of GLM which consists of the random effects in addition to fixed effect. A standard linear mixed model,

$$Y = X \quad + Zu + e \tag{1}$$

In equation (1), X and Z are the design matrices of fixed and random effect, and u are the parameter vectors for fixed and random effects and e is an error vector, respectively.

In a standard mixed linear model; $y \mid u \sim MVN(X \quad + Zu, R)$, $u \sim MVN(0, G)$ and $e \sim MVN(0, R)$. However, in GLMM, there is no need to define $e \sim MVN(0, R)$ assumption for error term. As in standard linear mixed model, $u \sim MVN(0, G)$ the assumption is valid in GLMM, too. General form of GLMM with the link function determined according to the distribution form of the dependent variable as given below:

$g(E(Y_{ij} \mid u)) = X_i \quad + Z_i u$ or

$E(Y_{ij} \mid u) = g^{-1}(X_i \quad + Z_i u) = g^{-1}(«_i) = «_i$ (2)

(Yeşilova, 2003; Dang *et al*., 2008; Rabe-Hesketh and Skrondal 2009).

In our study we used four different models designed for GLMM method.

Random residual model (Model A): Random residual model is also considered as a marginal model. In the model, R-side covariance parameters are obtained and dependency among the results is modeled. This model is defined as,

$E(Y \mid u) = g^{-1}(X \quad + Zu) = g^{-1}(¦) = «$ (3)

Random intercept model (Model B): This is a random intercept model $(\ _0)$. The model assumes that each individual has a different intercept. At the minimum one explanatory variable is added to the model. In the model, time is the explanatory variable. This model is defined as,

Level1: $Y_{ij} = \ _{0j} + \ _1 t_{ij} + e_{ij}$ (4)

Level 2: $\ _{0j} = \ _0 + u_{0j}$

As it is understood from the above given explanation, intercept $(\ _{0j})$ is random while $(\ _{ij})$ are fixed. In addition, (u_{0j}) indicates random change in amount of the second level around $(\ _0)$. On the other hand, means of random terms in the model are zero, while variance and covariances are $Var(e_{ij}) = ±^2_e, Var(u_{0j}) = ±^2_u, Cov(e_{ij}, u_{0j}) = 0$ respectively.

Random time model (Model C): In this model, only time is random. The constructed model assumes that each individual has a different slope. This model is defined as followed,

Level 1: $Y_{ij} = \ _0 + \ _{1j} t_{ij} + e_{ij}$ (5)

Level 2: $\ _{1j} = \ _1 + u_{1j}$

In Equation (5), time or slope $(\ _{1j})$ is random, intercept $(\ _0)$ is fixed. Furthermore (u_{1j}) indicates random change in amount of second level around $(\ _1)$. The variance and covariances of the random components in the model can be written as followed respectively,

$Var(e_{ij}) = ±^2_e, Var(u_{1j}) = ±^2_{u_{1j}}, Cov(e_{ij}, u_{1j}) = 0$

Random intercept time model (Model D): In the model, both intercept $(\ _0)$ and time or slope $(\ _1)$ are random. The model expresses that each individual has a specific intercept and slope. This model is,

Level 1: $Y_{ij} = \ _{0j} + \ _{1j} t_{ij} + e_{ij}$ (6)

Level 2: $\ _{0j} = \ _0 + u_{0j}$

$\ _{1j} = \ _1 + u_{1j}$

where (u_{0j}) and (u_{1j}) are random variables of the second level for $(\ _{0j})$ and $(\ _{1j})$, respectively (Candy, 2000; Hox, 2002; Akkol *et al*., 2007; Jiang, 2007).

Five different optimization techniques were used in the analysis of these models. These models are Quasi-Newton (QUANEW), Newton-Raphson (NEWRAP), Trust Region (TRUREG), Newton-Raphson Ridge (NRRIDG), Double-Dogleg (DBLDOG). Furthermore, together with these optimization techniques, three different covariance structures were used which are Compound Symmetry (CS), Unstructured (UN) and First-order autoregressive (AR(1)). Data were analyzed using Proc Glimmix procedures in SAS 9.2 software (SAS, 2010).

Parameter Estimation: PQL (Penalized Quasi Likelihood) and MQL (Marginal Quasi Likelihood) methods were generally used for parameter estimations in GLMM. In this study, PQL method was used for parameter estimation. In PQL method, using Taylor expansion, linearity is achieved around fixed and random $(\hat{\ }$ and $\hat{u})$ effects, respectively. As a result, if Taylor expansion of $Y_{ij} = h(x'_{ij}\hat{\ } + z'_{ij}\hat{u}) + £_{ij}$ and $h(.)$ are written in matrix form, the PQL equation is obtained as following,

$Y_i^* \equiv \hat{V}_i^{-1}(Y_i - \hat{«}_i) + X_i \hat{\ } + Z_i \hat{u}_i \approx X_i \quad + Z_i u_i + £_i^*$ (7)

In Equation (7), $\hat{V}_i$ is equal to the diagonal matrix with diagonal element $V(\hat{«}_{ij})$, $(£_i^*)$ mean is zero, variance is $(\hat{V}_i^{-1} £_i)$, X_i and Z_i are the design matrices, respectively. Given starting values for the parameters , G and ¤ in the marginal likelihood function, empirical

Bayes estimates are calculated for u_i and then pseudo data Y_i^* are computed. Then, the approximate linear mixed model is fitted, yielding updated estimates for , G and ¤ . These are then used to update the pseudo data and this whole scheme is iterated until a convergence criterion is reached (Molenberghs and Verbeke 2005).

RESULTS

AIC and BIC results obtained from different optimization techniques and covariance structure from four models are given Table 1.

Table 1. Goodness of fit results obtained from different optimization techniques and covariance structures for Model A, Model B, Model C and Model D

Optimizatin Techniques	Covariace Structures	Model A		Model B		Model C		Model D	
		Pseudo-AIC	Pseudo-BIC	Pseudo-AIC	Pseudo-BIC	Pseudo-AIC	Pseudo-BIC	Pseudo-AIC	Pseudo-BIC
QUANEW	CS	744.18	747.11	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
	UN[1]	-	-	996.87	998.34	1107.57	1109.03	951.78	956.17
	AR(1)	762.07	765.00	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
NEWRAP	CS	744.18	747.11	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
	UN[1]	-	-	996.87	998.34	1107.57	1109.03	951.78	956.17
	AR(1)	762.07	765.00	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
TRUREG	CS	744.18	747.11	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
	UN[1]	-	-	996.87	998.34	1107.57	1109.03	951.78	956.17
	AR(1)	762.07	765.00	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
NRRIDG	CS	744.18	747.11	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
	UN[1]	-	-	996.87	998.34	1107.57	1109.03	951.78	956.17
	AR(1)	762.07	765.00	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
DBLDOG	CS	744.18	747.11	998.87	1001.80	1109.57	1112.50	1048.85	1051.78
	UN[1]	-	-	996.87	998.34	1107.57	1109.03	-	-
	AR(1)	762.07	765.00	998.87	1001.80	1109.57	1112.50	1048.85	1051.78

[1]AIC and BIC could not obtained

Table 2. Parameter estimates and standard error values of different models

Parameters	Model A	Model B	Model C	Model D
	Estimation (SEM)	Estimation (SEM)	Estimation (SEM)	Estimation (SEM)
Intercept	2.761^{***}	2.964^{***}	2.054^{***}	2.745^{***}
	(0.364)	(0.245)	(0.256)	(0.387)
Weight of male lambs	-0.024^{**}	-0.033^{***}	0.006	-0.029^{**}
	(0.007)	(0.004)	(0.006)	(0.007)
Response of flehmen	0.024	0.024	0.017	0.033^{*}
	(0.027)	(0.016)	(0.016)	(0.017)
Anogenital sniffing	0.071^{***}	0.074^{***}	0.068^{***}	0.074^{***}
	(0.012)	(0.007)	(0.007)	(0.007)
Raising of tail	0.010	-0.013	0.012	0.005
	(0.025)	(0.015)	(0.013)	(0.016)
Duration of mount	-0.001^{*}	-0.001^{***}	-0.001^{***}	-0.001^{**}
	(0.001)	(0.000)	(0.000)	(0.000)
Frequency of mount	0.024^{***}	0.025^{***}	0.021^{***}	0.025^{***}
	(0.004)	(0.003)	(0.002)	(0.003)

$^{*}P<0.05$; $^{**}P<0.01$; $^{***}P<0.001$

The smaller values of the model selection criteria (AIC and BIC) given in Table 1 indicate the best combination of optimization technique and covariance structure for models (Model A, B, C and D) used. The CS structure resulted in a smaller AIC and BIC than other covariance structures for all optimization tecniques. Typically the AIC and BIC values for all covariance structure and optimization tecnique combination increased drastically with model complexity. For example the AIC values for QUANEW-CS combination were 744.18, 998.87, 1109.57 and 1048.85 for Model A, Model B, Model C and Model D, respectively. This pattern was evident for AIC and BIC for all other optimization technique-covariance structure combinations.

Parameter estimates and their standard error values relating to CS in Model A and UN in Model B, C and D with all optimization techniques are shown in Table 2. The anogenital sniffing, duration of mount, frequency of mount and intercept were found to be statistically significant on vocalization for all models ($p<0.05$; $p<0.01$; $p<0.001$). Furthermore, the response of flehmen was significant on vocalization in Model D ($p<0.05$) only. The anogenital sniffing and frequency of mount had the greatest effect on vocalization regardless the model used. Likewise the duration of mount had a negative and significant effect on vocalization for all models. Similarly, except in Model C, the weight of male lambs had a negative effect and was significant on vocalization in Model A, B and D.

DISCUSSION

In the results of optimization techniques, all models reached a convergence (for pre-specified an acceptable difference between two consecutive log likelihood functions) at different number of iterations. For Model A, Model B, Model C and Model D iteration number varied between 11-42, 4-16, 4-7 and 5-97, respectively. It was found that, for Model A, the highest number of iterations was obtained from QUANEW and DBLDOG optimization techniques with CS structures (17-42 number of iterations); for Model B, the highest number of iterations was obtained from DBLDOG optimization technique with CS structure (16 iterations); for Model C, the highest number of iterations was obtained from QUANEW and DBLDOG techniques with CS, UN and AR(1) structures (7 iterations); the highest number of iterations was obtained from DBLDOG optimization technique and AR (1) covariance structures (97 iterations) for Model D. When compared to TRUREG, NRRIDG and NEWRAP optimization techniques, QUANEW and DBLDOG require a higher number of iterations. However, because of theoretical structure of these optimization techniques, each iteration can be obtained much faster to avoid time consuming for big data set (SAS, 2008; Ser, 2011).

In GLMM, goodness of fit criteria are calculated as pseudo-AIC and pseudo-BIC. As it is known, goodness of fit criteria like AIC and BIC are calculated from log-likelihood values. In GLMM, since pseudo-likelihood value is calculated, obtained goodness of fit criteria are pseudo-AIC and pseudo-BIC (SAS, 2008). When AIC and BIC goodness of fit criteria in Model A were evaluated on Table 1, in all optimization techniques, the smallest goodness of fit criterion was obtained from CS, which had a homogenous structure. In CS structure variance values in diagonal elements, in other words, the measurements at all times have the same variance while covariance among the observation values is fixed. In model where intercept (Model B) and time (Model C) are randomly constructed, when goodness of fit criteria is evaluated, it was found that UN covariance structure had the best adaptation to all optimization techniques. There are no assumptions for the variance and covariance in case of UN; however there is a heterogeneous structure (Akbaş *et al*., 2001; Weiss, 2005; Ser, 2011). Furthermore, no convergence problem was observed in these models. In Model D, in random time and intercept model, it was found that UN was the smallest goodness of fit criterion in all optimization methods. However, AR (1) covariance structure for all models to the data set had the worst goodness of fit criteria for the model specified. According to this finding, for the male lambs with low vocalization frequency, the time between the moment of entering test arena and realization of mount behavior is shortened. Similarly, male lambs which have a low live weight tend to show a higher ratio of vocalization behavior since they are unable to show mount behavior. In addition, high courting behavior frequency observed before breeding activity in male lambs results from lack of sexual experience. The effect of male lambs age on the quality of sexual activity was found to be significant in many studies (Katz *et al*., 1988; Price *et al.*, 1988; Yılmaz *et al*., 2009). This also indicates that it is important to enable the male lambs to gain sexual experience within the framework of flock management programs.

Based on the optimization techniques and covariance structures used in the models, since no convergence problem was encountered, it can be stated that Model B and Model C had the best adaptation to the used data set. In addition, when compared to other models, these models required lower number of iterations to reach convergence. While UN covariance structure provided the best adaptation to the data set in Model B, C and D, as a result of CS structure, a good adaptation was obtained only for Model A.

In conclusion covariance patterns and optimization techniques are two critical issue in repeated data analysis. There are many choices of covariance patterns for modeling repeated measures data. Choosing the most appropriate pattern is important in order to draw accurate conclusions. On the other hand the optimization technique plays and essential role for faster converge which is critically important for big data set with many parameters.

Acknowledgements: The data used in this study were collected from Ph.D thesis of Mr. Ayhan Yilmaz titled

"Determination of some reproduction characteristics in Norduz male lambs".

REFERENCES

Akbaş Y., M. Z. Fırat and Ç. Yakupoğlu (2001). Comparison of different models used in the analysis of repeated measurements in animal science and their sas applications. *Agricultural Information Technology Symposium*. 20-22 September 2001, Sütçü İmam University, Kahramanmaraş.

Akkol Gökdere S., H. Okut and S. Şehribanoğlu, (2007). Comparisons of Multilevel ZIP and ZIBN regression models for count data with excess zeros. Veterinary Medicine and Animal Husbandry Research Grant Committeee, Project no: Tubitak Tovag 1050248 (Project final report, unpublished).

Candy S.G. (2000). The application of generalized linear mixed models to multilevel sampling for insect population monitoring. Environmental and Ecological Statistics 7: 217-238.

Dang Q., S. Mazumdar and R.P. Houck (2008). Sample size and power calculations based on generalized linear mixed models with correlated binary outcomes. Computer Methods and Programs in Biomedicine 91: 122-127.

Dobson J.A. (1990). An introduction to generalized linear models. 2nd edn. Chapman and Hall, New York.

Eyduran E. and Y. Akbaş (2010). Comparison of different covariance structure used for experimental design with repeated measurement. The J. Anim. and Plant Sciences 20(1): 44-51.

Hox J. (2002). Multilevel analysis: Techniques and applications. 2nd edn. New Jersey.

Jiang J. (2007). Linear and genaralized linear mixed models and their applications. 2nd edn. Springer, New York.

Katz L.S., E.O. Price, S.J.R. Wallach and J.J. Zenchak (1988). Sexual performance of rams reared with or without females after weaning. J. Anim. Sci. 66:1166-1173.

McCullagh P. and J.A. Nelder (1989). Generalized linear models. 2nd edn. Chapman & Hall, London.

McCulloch C.E. and S.R. Searle (2001). Generalized linear and mixed model. 2nd edn. John Wiley&Sons, New York.

Molenbergs G. and G. Verbeke (2005). Models for discrete longitudinal data.1nd Springer, New York.

Orhan H., E. Eyduran and Y. Akbaş (2010). Defining the best covariance structure for sequential variation on live weights of Anatolian Merinos male lambs. The J. Anim. and Plant Sciences 20(3): 158-163.

Price E.O., L.S. Katz, S.J.R. Wallach and J.J. Zenchak (1988). The relationship of male-male mounting to the sexual preferences of young rams. Applied Animal Behaviour Science 21: 347-355.

Rabe- Hesketh S. and A. Skrondal (2009). Generalized linear mixed-effects models. In: Geert F (eds), Longitudinal Data Analysis. New York, CRC Press pp. 79–106.

SAS (2008). SAS/STAT 9.2 User's Guide: The GLIMMIX Procedure (Book Excerpt). 2nd edn. SAS Institute Inc., Cary, North Carolina.

SAS (2010). SAS/STAT software: Hangen and enhanced, Version 9.2, SAS, Inst. Inc., Cary, North Carolina.

Ser, G. (2011). Model selection and comparing optimization techniques in marginal and non-marginal multilevel generalized linear mixed model using missing observed longitudinal data. PhD Diss., Yüzüncü Yil University 111p. Available from: Yüzüncü Yıl University, Institute of Natural Science.

Singer J.D. and J.B. Willet (2003). Applied longitudinal data analysis: modeling chance and event occurence. 2nd edn. Oxford University Press,New York.

Weiss E.R. (2005). Modelling longitudinal data. 1nd edn. Springer, New York.

Yeşilova A. (2003). The use of mixed poisson regression models for categorical data in biology. PhD Diss., Yüzüncü Yil University 68p. Available from: Yüzüncü Yıl University, Institute of Natural Science.

Yılmaz, A., F. Karakuş and A. Yeşilova (2009). Sexual behaviour characteristics in Norduz and Karakas rams and changes with age. J. Agri. Sci., 15: 270-276.