Random regression models for body weight from birth to 210 days of age in partridges (*Rhynchotus rufescens*), using different structures of residual variances

P. THOLON[1], S.A. QUEIROZ[1]*

[1]Departamento de Zootecnia – Faculdade de Ciências Agrárias e Veterinárias - UNESP, Via de Acesso Prof. Paulo Castellane, Jaboticabal, 14884-900, SP, Brazil.
*** Corresponding author: saquei@fcav.unesp.br**

Abstract: *Rhynchotus rufescens* is a native bird spread out in South America, showing traits that qualify it to be considered as an option by the poultry industry. Random regression models (RRM) allows considering heterogeneous residual variances for each age. However, this feature increases the number of parameters to be estimated in the maximization likelihood function process. Searching for more parsimonious RRM, several approaches have been suggested. One of them is the use of different structures of residual variances modelled through step function in different classes with similar variance or through variance functions. A total of 7,369 records of body weight measured from birth to 210 days of partridges born from 2000 to 2004 were used in this research. Data considered birds with at least 4 records of weight, sires with at least 4 progenies and groups of birth with at least 4 animals. The birds were raised in a commercial barn, grounded by grass hay litter. They received feed and water *ad libitum*. The eggs were hatched daily at 35.5 ºC and 70% of humidity. The chicks were sexed when reached around 250 g. The random regression models applied to the data set considered different structures of residual variances and were performed by the restricted maximum likelihood method. The residual variances were modelled using classes of 210 (R210) and 30 (R30) ages and variance functions with orders ranging from quadratic (VF2) to nine (VF9). The R30 considered birds weighted in the same week. The random effects included were the genetic additive direct and the permanent environment effects of the animal. It was not possible to include the maternal effects in the models. All random effects were modelled by cubic regression on Legendre polynomials. The models were compared by the likelihood ratio test, the Akaike's information criterion and the Schwarz's Bayesian information criterion. The models R210 and VF5 showed the best results. In conclusion, the most parsimonious model was VF5 and should be applied to fit growth records of partridges.

Key words: Legendre polynomials, partridges (*Rhynchotus rufescens*); residual variance function; step function; wild bird

Introduction

Recently, the random regression models have been applied in longitudinal data analysis in animal breeding. This method allows fitting random growth curves for each animal, expressed as deviation of an average curve of the population or individuals groups (Schaeffer, 1996). Since the random effects are described by continuous functions in the random regression models, Meyer and Hill (1997) and Meyer (1998a) showed that random regression models (RRM) are a special case of covariance function (CF), and covariance function coefficients can be estimated directly from random regression models by restricted maximum likelihood in infinitive dimensional data (Meyer, 1998a).

The RRM and CF allow to model random effects using polynomials functions and make possible to consider heterogeneous residual variance for each age.

The initial works using RRM (Jamrozik and Schaeffer, 1997; Jamrozik et al., 1997) have not considered heterogeneous residual variance and have produced overestimated additive genetics variances estimates. However, when considering heterogeneous residual variances for each age, the partition of the total variance is improved, but the number of parameters to be estimated increases in the maximization likelihood function process.

Searching for more parsimonious RRM, other approaches have been suggested as the utilization of different structures of residual variances modeled through step function in different classes with similar variances and also using variance functions on ordinary and orthogonal polynomials (Olori et al., 1999; Meyer, 1999; Albuquerque and Meyer, 2001).

This methodology was used for estimating genetic parameters for pig weight (Huisman et al., 2002) and number of piglet in the litter (Luković et al., 2004). However, in birds, studies on longitudinal weight data using RRM are not available yet and the researches of Mielenz et al. (2002), Quinton et al. (2002) and Anang et al. (2002) on egg yield of laying chickens should be mentioned. The aim of this study was to compare different structures of residual variances in RRM, decreasing the number of parameters to be estimated, and to facilitate the convergence process in order to obtain variances components of partridges (*Rhynchotus rufescens*) weights from birth to 210 days old.

Materials and methods

The information used in this paper was taken from birds of the Wild Animal Section of the Faculdade de Ciências Agrárias e Veterinárias (UNESP), at Jaboticabal, São Paulo State, Brazil. The birds were raised in reproduction wired boxes, with concreted floor, covered by a bed lid of Coast-cross (*Cynodon dactylon*) hay, whose dimensions were 1.0 x 2.0 x 2.0 m, placed inside an avian barn. The proportion of females per males ranged from 1:1 to 6:1. The water supply and feeding were *ad libtum* with peletized ration in tubular hods and hanging fountains.

Eggs were colected daily and identified individualy for pedigree control. After 15 days of incubation at 35.5 ºC and 70% of humidity, they were transfered to the birth place. The newborn chicks received temporary markers and were transfered to the breeding boxes, being the maximum allotment of 20 animals per box. The sex verification was made by cloaca reversion, when the birds weighted around 250 g and they received a definitive numered metal marker placed on the right wing when they reached around 90g of weight. The healthy management included the checking for endo and ectoparasites and the bed lid change every two months. The daily hatch procedure caused little synchronized births, originating lots of heterogeneous animals.

Records came from 408 animals born from 2000 to 2004 totaling 7369 observation. The animals were weighed weekly on digital balance (Filizola) from birth to 210 days of age. Data set included records on 67 sires, 17 dams and 27 grandsires, animals with at least 4 records on weight and sires with at least 5 progenies.

Random regression models using an animal model were performed to analyze the data set. Fixed effects included contemporary groups (formed by year and fortnight of birth), sex of bird and age of the animal at weighting effects. The last one was considered as covariates. This effect was modeled using quintic regression on Legendre polynomials.

The random effects included in the models considered the additive genetic direct and animal permanent environmental effects, modeled by quintic regression on Legendre polynomials. The additive genetic maternal effect was not included in the models because very few females could be identified on the maternal side of pedigree. Different structures of residual variances were tested using:

1) homogeneous residual variance (1 class);
2) heterogeneous residual variances using 210 classes (all ages – R210)) and 30 classes of ages (birds born in the same week – R30), and
3) variances functions (VF) modeled by ordinary polynomials with orders ranging from two to nine.

The number of parameters estimated for each model was k(k+1)/2=6(6+1)/2 coefficients for additive genetic direct and animal permanent environmental (with k=6 being the order of polynomial fitted) besides residual variances.

The variance components were estimated by REML using the DXMRR option of the DFREML program (Meyer, 1998b).

The residual variances (σ^2_e) using variance functions were estimated by regression coefficients of the variances functions, represented as follows:

$$\sigma^2_{e_i} = \sigma^2_{e_o}\left(1 + \sum_{r=1}^{q} \beta_r t_{ij}^r\right)$$

Where: σ^2_{eo} is the intercept variance, β_r are the r regression coefficients from variance functions of order q and t_{ij} are the ages i for each animal j.

The models could be represented by:

$$Y_{ij} = F + \sum_{m=0}^{k_b-1} \beta_m \phi_m(t_{ij}) + \sum_{m=0}^{k_A-1} \alpha_{im} \phi_m(t_{ij}) + \sum_{m=0}^{k_R-1} \gamma_{im} \phi_m(t_{ij}) + \varepsilon_{ij}$$

where:

Y_{ij} is the ith age from the jth animal;

F is a set of fixed effects ;

β_m are the fixed regression coefficients to model the population mean;

$\phi_m(t_{ij})$ is the mth Legendre polynomial of age;

t_{ij} is the standardized (-1 to +1) age at recording;

α_{jm} and γ_{jm} are the random regression coefficients for additive genetic direct and animal permanent environmental effects for each animal j; and

ε_{ij} is the temporary environmental effect.

Using matrix notation, the models can be described as:

$$\mathbf{Y = Xb + Z\alpha + W\gamma + \varepsilon}$$

where:

Y is the vector of observations; **b** is the vector of fixed effects; $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are the solution vectors for the random effects of animal additive genetic and animal permanent environmental coefficients, respectively; $\boldsymbol{\varepsilon}$ is the vector of residual effects and **X, Z e W** are the incidence matrices for fixed, animal additive genetic and animal permanent environment effects, respectively.

The models consider the following assumptions:

$$E\begin{bmatrix} y \\ \alpha \\ \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad V\begin{bmatrix} \alpha \\ \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} K_A \otimes A & 0 & 0 \\ 0 & K_R \otimes I & 0 \\ 0 & 0 & \sigma^2_\varepsilon D \end{bmatrix}$$

Where:

$\mathbf{K_A}$ and $\mathbf{K_R}$ are the (co)variances matrices between the random effects of animal genetic and permanent environmental coefficients; respectively. A zero correlation between animal genetic and permanent environmental coefficients is an assumption of this model. **A** is the relationship matrix between birds ; **I** is an identity matrix; $\otimes$ is the Kroenecker product between matrices; **D** is a diagonal block matrix containing the residual variances and $\boldsymbol{\sigma^2_\varepsilon}$ is the residual variance.

The models were compared by likelihood ratio test (LRT). LRT allows comparisons between nested models and tends to favor models with higher number of parameters. Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC) (Wolfinger, 1993) were also applied. These favored more parsimonious models imposing penalties according to the number of parameters to be estimated.

Results and discussion

Table 1 presents the values of LRT (nested models), AIC and BIC for the models studied. Comparing the models with different residual variance classes, according to LRT, the model with homogeneous residual variance was inadequate, differing significantly (P<0.01) of the other models.

The models with heterogeneous residual variances (30 and 210) also show significant (P<0.01) between among them, with the decrease in the value of the likelihood function for the model with more number of parameters. The three criteria pointed out the model with 210 classes as the best one for residual variances considered in classes. On the other hand, the CL210 model had too much parameters to be considered which makes the computing procedures slow. Among the models using residual variances functions, the best one according to the three criteria used (LRT, AIC and BIC) was the model fitting ordinary polynomial of 5th order to the residual variances.

Table 1. Number of parameters (p), value of the maximum likelihood function (Log L) , Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (BIC) e likelihood ratio test (LRT) applied to models using homogeneous residual variances (G_6_6_6_0), classes of residual variances (CL) and residual variance functions (VF).

	Models	p	LOG L	AIC	BIC	LRT
(1)	G_6_6_6_0	43	-28354,376666	56794,753331	57091,669935	-
(2)	G_6_6_6_CL30	72	-27242,859127	54629,718255	55126,880939	(2-1) 1111,5175**
(3)	G_6_6_6_CL210	252	**-26363,789007**	**53231,578014**	**54971,647411**	(3-2) 879,0701**
(1)	G_6_6_6_VF2	45	-24645,929039	49381,858078	49692,584756	-
(2)	G_6_6_6_VF3	46	-24545,743878	49183,487755	49501,119471	(2-1) 100,1852**
(3)	G_6_6_6_VF4	47	-24485,153321	49064,306641	49388,843393	(3-2) 60,5905**
(4)	G_6_6_6_VF5	48	**-24196,543818**	**48489,087637**	**48820,529427**	(4-3) 288,6095**
(5)	G_6_6_6_VF6	49	-24219,590431	48537,180863	48875,527690	(5-4) 23,0466
(6)	G_6_6_6_VF7	50	-24501,463471	49102,926941	49448,178806	(6-5) 281,87304**
(7)	G_6_6_6_VF8	51	-24500,314507	49102,629013	49454,785915	(7-6) 1,1489
(8)	G_6_6_6_VF9	52	-24482,353575	49068,707149	49427,769089	(8-9) 17,9609

G_6_6_6_0 = order of fitting for fixed, animal genetic and animal permanent environment effects = 6, CL210 = residual classes considering 210 classes of ages, CL30 = residual classes considering 30 classes of age, VF2, VF3, VF4, VF5, VF6, VF7, VF8 and VF9 = variance functions, ** P<0,01.

Estimates of phenotypic, genetic direct, animal permanent environmental, maternal permanent environmental and residual standard deviations (SD) for the G666VF5 model studied are presented in Figure 1.

The standard deviation estimates increased from birth to 168 days of age. It can be seen on Figure 1 a disproportional increase in phenotypic, animal direct and animal permanent environment standard deviations in the last portion of their curves. Probably these increases were due to the polynomial order used to model the animal permanent environmental effect. Brotherstone et al. (2000) suggested the use of polynomials with greater orders to model the changes in permanent environmental variances in such cases, but the use of this tool did not solve the problem in our study.

Olori et al. (1999) studying different structures of residual variances applied to lactation curve, noticed that the changes in the residual variance models did not affect the estimates of variance for the genetic direct effect. However, the estimates of SD for genetic direct and animal permanent environmental effects were smoothly influenced by the shape of the residual variance models, when comparing the better two models (210 classes and variance function of 5th order).

The models using 30 and 210 classes presented estimates of phenotypic, additive genetic direct and animal permanent environmental SD estimates very similar, suggesting that the model using variance function of 5th order should be picked.

According to the results, there is no need in considering all the 210 residual variances classes. It is feasible modeling the residual variances using the 5th order polynomial variance functions. In addition, decreasing the number of parameters estimated makes the model more parsimonious, providing faster convergence and less processing computer time.

In conclusion the models considering heterogeneous residual variance were more appropriate than the model using homogeneous residual variance and in order to decrease the number of parameters estimated the variance function of 5th order could applied to this data set.

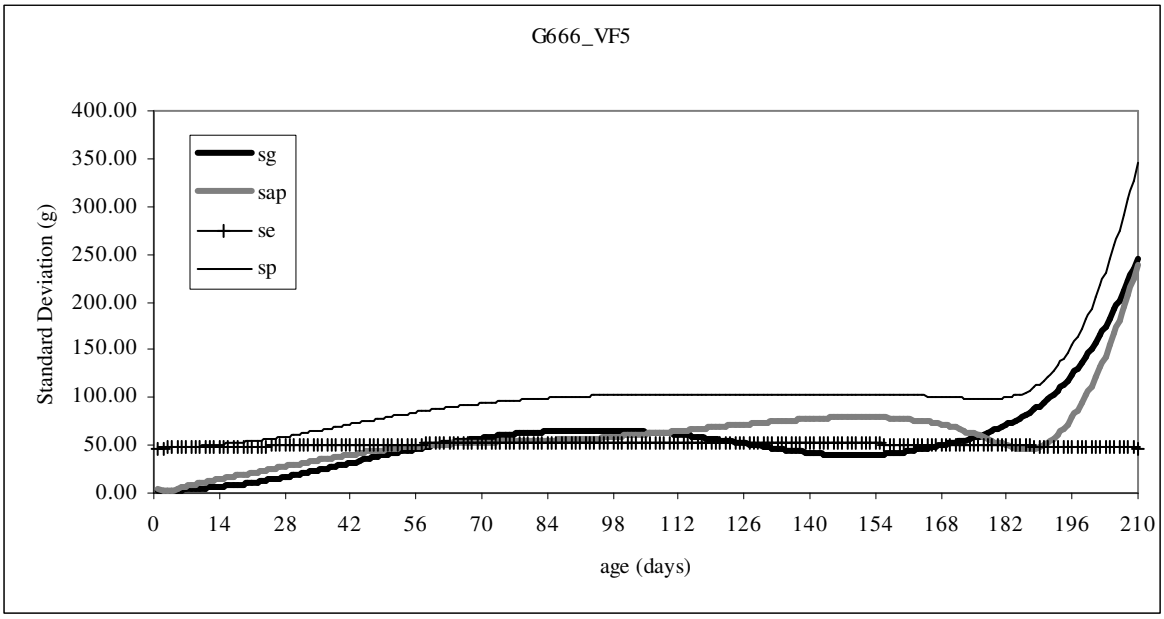

Figure 1. Estimates of phenotypic (sp), genetic additive direct (sg), animal permanent environment (sap) and residual (se) standard deviation using a variance function of order five for body weight of partridges (*Rhynchotus rufescens*).

References

ALBUQUERQUE, L. G., MEYER, K. (2001) Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle**.** *J. Anim. Sci.* **79**:2776-2789.

ANANG, A., MIELENZ, N., SCHULER, L. (202) Monthly model for genetic evaluation of laying hens. II Random regression. *British Poultry Science*, **43**: 384-390.

BROTHERSTONE, S., WHITE, I. M. S., MEYER, K.(2000) Genetic modeling of daily milk yield using orthogonal polynomials and parametric curves. *J.Anim. Sci.* **70**: 416-417.

HUISMAN, A. E., VEERKAMP, R. F., van ARENDONK, J. A. M. (2002) Genetic parameters for various random regression models to describe the weight data of pigs. *Journal of Animal Science* **80**: 575-582.

JAMROZIK, J., SCHAEFFER, L. R. (1997) Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holsteins. *J. Dairy Sci.***,** 80:762-770.

JAMROZIK, J., KISTEMAKER, G.J., DEKKERS, J.C.M., SCHAEFFER, R. (1997) Comparison of possible covariates for use in a random regression model for analyses of test day yields. *J. Dairy Sci.* **80**:2550-2556.

LUKOVIĆ, Z., MALOVRH, S., GORJANK, G., KOVAĆ, M. (2004) A random regression model in analysis of litter size in pigs. *South African Journal of Animal Science* **34**: 241-248.

MEYER, K., HILL, W. G.(1997) Estimation of genetic and phenotypic covariance functions for longitudinal or "repeated" records by restricted maximum likelihood. *Livest. Prod. Sci.* **47**:185-200.

MEYER, K.(1988a) Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.* **30**:221-240.

MEYER, K. (1988b) "DXMRR" - A program to estimate covariance functions for longitudinal data by REML. In: *6th World Congress on Genetics Applied to Livestock Production*, 1998b, Armidale, Australia, *Proceedings*.**..** CDRom.
MEYER, K. (1999) Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. *J. Anim. Breed. Genet.* **116**:181-203.
MIELENZ, N., ANANG, A., PREININGER, R., SCHUMUTZ, M., SCHUELER, L. (2002) Gennetic evaluation of laying performance data – comparison of models based on monthly records. In: *7th World Congress on Genetics Applied to Livestock Production*, 2002, Montpellier, France, *Proceedings*.**..** CDRom.
OLORI, V. E., HILL, W. G., MCGUIRK, B. J. et al.(1999) Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model.*Livest. Prod. Sci.* **61**:53-63.
QUINTON, V.M., McMILLAN,I., FAIRFULL, R.W. (2002) Changes in the variance structure of egg production over three lines cycles in white Leghorns. In: *7th World Congress on Genetics Applied to Livestock Production*, 2002, Montpellier, France, *Proceedings*.**..** CDRom.
SCHAEFFER, L. R**.**(1996) *Tópicos Avançados em Melhoramento Animal: Random Regression Models*. FCAV - UNESP, Câmpus de Jaboticabal, Jaboticabal, SP. p.25-33 (course notes).
WOLFINGER, R. D. (1993) Covariance structure in general mixed models. *Commun. Statist.* **22B**:1079-1106.