Biometrical Letters
Vol. 43 (2006), No. 1, 9-17

Statistical modeling of egg production curves

Anna Wolc[1], Mirosław Lisowski[2], Tomasz Szwaczkowski[1]

[1] Department of Genetics and Animal Breeding, Agricultural University of Poznan, ul. Wołyńska 33, 60-631 Poznań, Poland
[2] National Research Institute of Animal Production, Poultry Research Branch at Zakrzewo, Poland

SUMMARY

Nine functions (Wood, McMillan, McNally, Ali and Schaeffer, Wilmink, Yang, Guo and Swalve, Grossman, fourth order polynomial) were used to model egg production curves. The goodness of fit was analysed according to several criteria. A numerical example was supplemented. The five parameter Ali and Schaeffer model, Grossman model and four parameter model of McMillan most adequately described the data.

Key words: egg production curves, model comparison criteria, laying hens

1. Introduction

Egg production is known to show some regularity especially when summarized on weekly or monthly basis in a group of hens (Yang *et al.*, 1989). There are several applications of mathematical models like prediction of total production based on part-record (McMillan, 1981), prediction of economical effectiveness (Adams and Bell, 1980), description of persistency (Grossman *et al.*, 2000) and genetic evaluation (Anang *et al.*, 2001). A typical egg production curve consists of successive phases of slow and rapid increase and of a constant production and decreasing phase (Grossman *et al.*, 2000). Several types of functions have been suggested to describe growth curves, many of which can be fitted to poultry data. It should be noted that some of them were developed to describe other longitudinal traits, mainly milk yield (Druet *et al.*, 2003, Perz and Sobek, 2004). The models differ in mathematical approach: compartmental model (McMillan *et al.*, 1970), gamma type function (Wood, 1967), logistic model (Cason and Britton, 1988), set of intersecting lines with continuous transition (Grossman *et al.*, 2000). These models vary in complexity (most commonly between 3 and 5 parameters), computing demands and adequacy. Recently, genetic evaluation of traits expressed over time is based on fixed or random regression models (Schaeffer *et al.*, 2000, Anang et al., 2002). These models facilitate more accurate modelling of variance-covariance structure and are able to predict covariance structures at any time point along a continuous scale by including production curve in the model (Huisman, 2002). From practical point of view they enable more accurate selection and the use of information on the course of traits. However, they are computationally more demanding so preferably would include minimal number of parameters but maximising the goodness of fit. Selection has substantially changed an average egg production and consequently the shape of typical egg production curve. Therefore, there appears a need to reevaluate known models of this process. Nine functions were fitted to the monthly records of modern commercial flock in order to verify which of them give the best description of the production records.

2. Growth functions

The following functions were fitted to the data:

- Wood model (Wood, 1967) – incomplete gamma function developed to describe milk yield
  $y = at^b \exp^{(-ct)}$,
  where: y is a production in time t, a is a parameter associated with peak production, b represents the inclining slope, c is related to the declining slope.
- McMillan model (McMillan *et al.*, 1970) – compartmental model representing egg production of *Drosophila melanogaster*
  $y = a\left(1 - \exp^{-b(t-t_0)}\right)\exp^{-ct}$,
  where: t_0 is initial day of egg laying, a is potential maximum daily output of eggs, b is rate of increase in egg laying, c is the rate of decay of egg production
- McNally model (McNally, 1971) – modification of Wood model to better reflect egg production

  $$y = at^b \exp\left(-ct + dt^{\frac{1}{2}}\right),$$

  where: a, b, c and d are the parameters to be estimated in the model
- Ali and Schaeffer model (Ali and Schaeffer, 1987)- lactation curve

  $$y = p_0 + p_1\left(\frac{t}{n}\right) + p_2\left(\frac{t}{n}\right)^2 + p_3 \ln\left(\frac{n}{t}\right) + p_4\left(\ln\left(\frac{n}{t}\right)\right)^2,$$

where: p_i are the regression coefficients and p_0 is associated with peak production, p_1 and p_2 are associated with the decreasing slope, p_3 and p_4 are associated with the increasing slope, n is the number of periods

- Wilmink model (Wilmink, 1987) – exponential model of milk production

  $$y = a + bt + c\exp\left(-\frac{1}{2}t\right)$$

  where: *a*, *b*, *c* and *d* are the parameters to be estimated in the model
- Yang model (Yang *et al.*, 1989) – modification of McMillan model to take into account age at maturing

  $$y = \frac{a\exp^{-bt}}{1+\exp^{-c(t-d)}},$$

  where: *a* is a scale parameter, *b* is the rate of decrease in laying ability, *c* is the reciprocal indicator of the variation in sexual maturity, *d* is the mean age of sexual maturity.

- Guo and Swalve model (Guo and Swalve, 1995) – mixed log model for analysis of test day records (milk production)

  $$f(x) = a + bt^{0.5} + c\ln t$$

  where: *a*, *b*, *c* and *d* are the parameters to be estimated in the model
- Grossman model (Grossman *et al.* 2000) – egg production curve proposing new definition of persistency

$$y = r\left(\frac{y_p}{t_1 - t_2}\right) * \left[\ln\left(\frac{\exp^{t/r} + \exp^{t_1/r}}{1+\exp^{t_1/r}}\right) - \ln\left(\frac{\exp^{t/r} + \exp^{t_2/r}}{1+\exp^{t_2/r}}\right)\right] + rb_4\ln\left(\frac{\exp^{t/r} + \exp^{(t_2+P)/r}}{1+\exp^{(t_2+P)/r}}\right),$$

  where: t_1 and t_2 are the times at transition, *r* is the duration of transition, y_p is the level of constant production, b_4 is the rate of decline in production, *P* is the persistency of constant production.
- Fourth order polynomial- proposed as growth function of infants (Ricco el al. 2001)

  $$y = a + bt + ct^2 + dt^3 + ft^4$$

  where: *a*, *b*, *c*, *d* and *f* are the parameters to be estimated in the model.

3. Model comparison criteria

A number of criteria to describe the goodness of fit of nonlinear models can be found in the literature (Ott and Longnecker, 2001).
The following criteria were used to describe the adequacy of the models:

- Sum of squared deviations (SSD) which gives the insight into magnitude of overall prediction error;
- Proportion of variance explained (R^2) indicates which part of sum of squares can be eliminated by using multiple regression equation;
- Adjusted coefficient of multiple determination (R_a^2) which is the R^2 statistics adjusted for the number of observations and number of parameters in the model;
- Durbin-Watson test for autocorrelation (D-W), which is used to test for the presence of the first-order autocorrelation in the residuals of a regression equation (Jensen, 2005);
- Difference between total production observed and predicted (O-P), very important criterion from practical point of view especially in the analysis of economic effectiveness;

4. Numerical example

Average egg production during the first 9 months of lay was recorded. The records of 13950 birds of A88 line (modern purebred Rhode Island White hens) were averaged over six generations. The line has been selected on egg production and shell colour. No rapid break down of production, as it may appear due to disease or environmental factors was observed. Nine listed above functions were used to describe egg production. The Marquardt method (Marquardt, 1963) was used to estimate the parameters. The PROC NLIN in the SAS package (SAS, 2002) was applied. Predicted egg production curves were grouped according to the number of parameters and shown on Figures 1 to 3.

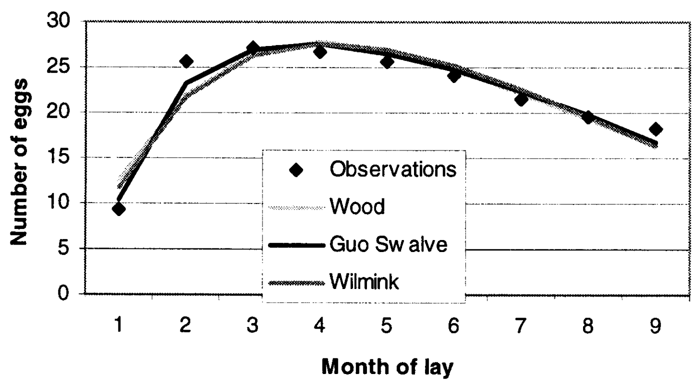

Figure 1. The fit of three parameter functions to egg production data

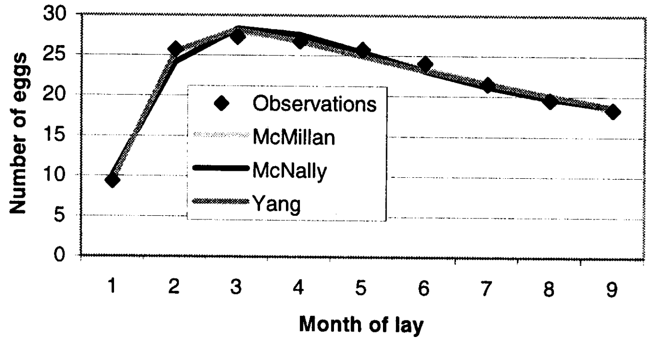

Figure 2. The fit of four parameter functions to egg production data

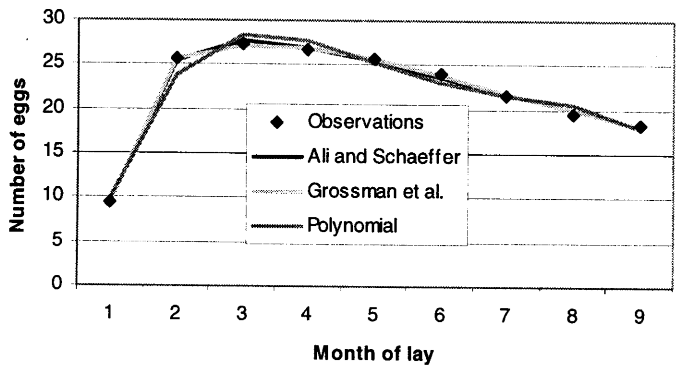

Figure 3. The fit of five parameter functions to egg production data

In the group of three parameter curves the mixed log model of Guo and Swalve gave the best description of the data with R_a^2 = 0.94 and substantially lower deviations from the actual records (Table 1). The regularity in the residuals was reflected in lower Durbin Watson values of Guo and Swalve and Wilmink models. However, it should be stressed that both of the models very well predicted total number of eggs.

From a group of four parameter functions the curve of McMillan outperformed other functions. Although Yang *et al.* (1989) developed a new model to reduce the limitations of McMillan model it's advantageousness was not confirmed in this study. McNally (1971) modified the Wood model to improve its fit to poultry data. According to all criteria it outperformed Wood model however it performed worse in the group of four parameter functions. Prediction of total production slightly deviated from the actual records for all models in this group.

Table 1. Comparison criteria of models describing egg production curve

Criterion\Model	Ali and Schaeffer	McMillan	McNally	Wood	Yang	Grossman	Guo and Swalve	Wilmink	Polynomial
SSD	0.910	1.495	6.184	33.715	2.930	0.546	11.751	31.220	8.030
R^2	0.997	0.994	0.976	0.871	0.989	0.998	0.955	0.881	0.969
R_a^2	0.993	0.991	0.962	0.828	0.982	0.996	0.940	0.841	0.939
D-W	2.086	1.835	2.381	2.004	1.497	2.323	1.828	1.864	2.494
O-P	0.000	-0.013	-0.203	-0.338	-0.059	0.000	0.000	0.000	0.000

Notes on symbols: SSD- sum of squared deviation, R^2 - proportion of variance explained, R_a^2 - adjusted coefficient of multiple determination, D-W- Durbin-Watson statistics, O-P - difference between total production observed and predicted

Ali and Schaeffer and Grossman functions gave excellent fit to the data although the last function had slightly higher Durbin Watson value. Their sums of square deviations were the lowest and proportion of variance explained exceeded 99%.

5. Final comments

The group of three parameter models did not follow the rapid increasing phase at the beginning of production and maintaining constant level for some time after a peak therefore tended to give underprediction at the beginning of peak production and overprediction at the end of this phase. They therefore retained systematic errors in this faze of laying cycle. Some limitations of Wood model (lack of inflection point in the initial period, systematic errors in model fitting) were already reported by Yang *et al.* (1989). This group of models is commonly used for description of lactation curves (Ferguson *et al.*, 2000) and genetic evaluation of dairy cattle (Rekaya *et al.*, 2000; Cobuci *et al.*, 2005) however, due to worse fit (especially around the peak production) except from the Guo and Swalve model can not be recommended for description of poultry data.

The fit of four parameter functions was in the faze of peak production substantially better than in the previous models. The McMillan model was developed to describe egg production of *Drosophila melanogaster*. However, already in 1970s, Gavora *et al.* (1971) reported that it was fitted closely to egg production of group of hens. Yang *et al.* (1989) modified compartment model of McMillan to take into account variation in sexual maturity. He reported improved fit of a new model which however does not apply to the studied population. It should be stressed that parameters of both models have biological meaning and therefore are easier to interpret.

The Ali and Schaeffer model is a function used to describe lactation curves. It is linear in parameters therefore can be relatively easy fitted to regression models in genetic evaluation. It was also suggested for genetic evaluation of laying hens (Anang *et al.*, 2002). The forth order polynomial could play the same role however its fit was worse. Grossman *et al.* (2000) proposed their model together with new definition of persistency as the number of weeks in which constant production is maintained. Its parameters may be of interest to breeders however the complexity of the function may lead to convergence problems (local maximum depending on starting values, parameters out of biological sense range).

REFERENCES

Adams C.J., Bell D.D. (1980). Predicting poultry egg production. *Poultry Science* 59, 937-938.

Ali T.E., Schaffer L.R. (1987). Accounting for covariances among test day milk yield in dairy cows. *Canadian Journal of Animal Science* 67, 637-644.

Anang A., Mielenz N., Schüler L. (2001). Monthly model for genetic evaluation of laying hens. I. Fixed regression. *British Poultry Science* 42, 191-196.

Anang A., Mielenz N., Schüler L. (2002). Monthly model for genetic evaluation of laying hens. II. Random regression. *British Poultry Science* 43, 384-390.

Cason J.A., Britton W.M. (1988). Comparison of compartmental and Adams-Bell models of poultry egg production. *Poultry Science* 67, 213-218.

Cobuci J.A., Euclydes R. F., Lopes, P. S. Costa C.M. Torres R., Pereira C.S. (2005). Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. *Genetics and Molecular Biology* 28, 75-83.

Druet T., Jaffrezic F., Boichard D., Ducrocq V. (2003). Modeling Lactation Curves and Estimation of Genetic Parameters for First Lactation Test-Day Records of French Holstein Cows. *Journal of Dairy Science* 86, 2480–2490.

Ferguson J. D., Beede D. K., Shaver R. D., Polan C. E., Huber J. T., Chandler P. T. (2000). A Method to Analyze Production Responses in Dairy Herds. *Journal of Dairy Science* 83, 1530–1542.

Gavora, J. S., Parker R. J., McMillan I. (1971). Mathematical model of egg production. *Poultry Science* 50, 1306-1315.

Grossman M., Grossman T.N., Koops W.J. (2000). A model for persistency of egg production. *Poultry Science* 79, 1715-1724.

Guo Z., Swalve H.H. (1995). Modelling of the lactation curve as a sub-model in the evaluation of test day records. INTERBULL Mtg., Prague, Sept. 7–8. *INTERBULL Bulletin* 11.

Huisman A. (2002). Genetic analysis of growth and feed intake patterns in pigs. *PhD Thesis*. Wageningen University, The Netherlands.

Jensen A. (2005). Durbin Watson Statistic. http://www.csus.edu/indiv/j/jensena/mgmt105/durbin.htm.

Marquardt D.W. (1963). An algorithm for the Least-Squares Estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics* 11, 431–441.

McMillan I., Fitz-Earle M., Butler L.,Robson D.S. (1970). Quantitative genetics of fertility I. Lifetime egg production of *Drosophila melanogaster*. *Genetics* 65, 349-353.

McMillan I. (1981). Compartmental model analysis of poultry egg production curves. *Poultry Science* 60, 1549-1551.

McNally D.H. (1971). Mathematical model for poultry egg production. *Biometrics* 27, 737-738.

Ott L.R., Longnecker M. 2001. *An introduction to statistical methods and data analysis*. Duxbury. Pacific Grove, USA.

Perz W.P., Sobek Z. 2004. The accuracy of 200-day milk yield estimation in cows on the basis of test yield from the 100-day period. *Biometrical Letters* 40, 73-80.

Rekaya Iii., Carabano M.J., Toro M.A. (2000). Bayesian analysis of lactation curves of Holstein-Friesian cattle using a nonlinear model. *Journal of Dairy Science* 83, 2691-701.

Ricco R.G., Nogueira de Almeida C.A., Del Ciampo L.A., Daneluzzi J.C., Ferlin M.L.S., Gerson MuccilloG. (2001). Growth of exclusively breast-fed infants from a poor urban population. *Archivos Latinoamericanos de Nutrición* 51, 122-126.

SAS Institute Inc. 2002-2003. *The SAS System for Linux* version 9.1, Cary, NC 27513-2414 USA.

Schaeffer L.R., Jamrozik J., Kistemaker G.J., Van Doormaal B.J. (2000). Experience with a Test-Day Model. *Journal of Dairy Science* 83, 1135-1144.

Wilmink J.B.M. (1987). Adjustment of test-day milk, fat and protein yields for age, season and stage of lactation. *Livestock Production Science* 16, 335.

Wood P.D.P. (1967). Algebratic model of the lactation curve in cattle. *Nature* 216, 164-165.

Yang N., Wu C., McMillan I. (1989). A new mathematical model of poultry egg production. *Poultry Science* 68: 476-481.