Original article

Estimation of non-linear growth models by linearization: a simulation study using a Gompertz function

Kaarina VUORI*, Ismo STRANDÉN, Marja-Liisa SEVÓN-AIMONEN, Esa A. MÄNTYSAARI

MTT Agrifood Research Finland, Biotechnology and Food Research, Biometrical Genetics, FIN-31600 Jokioinen, Finland

Abstract – A method based on Taylor series expansion for estimation of location parameters and variance components of non-linear mixed effects models was considered. An attractive property of the method is the opportunity for an easily implemented algorithm. Estimation of non-linear mixed effects models can be done by common methods for linear mixed effects models, and thus existing programs can be used after small modifications. The applicability of this algorithm in animal breeding was studied with simulation using a Gompertz function growth model in pigs. Two growth data sets were analyzed: a full set containing observations from the entire growing period, and a truncated time trajectory set containing animals slaughtered prematurely, which is common in pig breeding. The results from the 50 simulation replicates with full data set indicate that the linearization approach was capable of estimating the original parameters satisfactorily. However, estimation of the parameters related to adult weight becomes unstable in the case of a truncated data set.

Gompertz function / non-linear mixed effects / variance components / breeding values / likelihood approximation

1. INTRODUCTION

Non-linear functions are particularly suited to model growth data, because predictions outside the data range can be made more reliably than by linear models, and the entire growth process can be described by few parameters. For example, growth data models commonly apply the Gompertz function, where the estimated parameters can have biological meaning. Non-linear models are, however, more complicated to solve than linear models, and several algorithms

---

* Corresponding author: kaarina.vuori@mtt.fi

have been proposed to estimate the parameters and variance components of non-linear mixed effects models [4].

In animal production research, the Bayesian framework has received much attention in growth curve analysis [2, 11]. This popularity is due to the use of Markov chain Monte Carlo methods that allow the solution of numerically complicated posterior density integration and calculation of confidence interval estimates. The cost of this procedure is, however, intensive calculations and the need to assure a sampling equilibrium [3]. Another possibility is to approximate the likelihood function using linearization [4, 10, 17, 18] or numerical integration [10]. Also, both of these alternatives are computationally difficult, yet, linearization may enable the inference of linear mixed effects models. All linearization methods in the literature are quite similar. Early methods use first-order Taylor series expansion of non-linear functions around expectation of the random effects, and are solved by either maximum likelihood (ML) or generalized least squares (GLS) estimation [4]. Lindstrom and Bates [9] suggested a more accurate method of making the expansion around current estimates of the random effects. Subsequent research has focused more on the second-order Taylor series expansion of integrals invoked by the Laplacian approximation [10, 17, 18].

Because of generality and familiar formulation, the most interesting choice of approximation is based on the second-order Taylor series expansion with respect to random effects that were presented by Wolfinger and Lin [18]. They gave two alternative approaches to select points of expansion: a zero-expansion method using expected values, and an EBLUP-expansion method using the empirical best linear unbiased predictors of the random effects. Both approximations lead to algorithms that iteratively fit mixed linear models to the suitably transformed data using either ML or restricted maximum likelihood (REML). Therefore, they allow the use of commonly applied methods for linear mixed effects models, and the use of existing programs after small modifications. A similar algorithm was proposed by Breslow and Clayton [3] in the context of generalized linear mixed models. Because conditions to functionality of the approximation methods are difficult to identify, Wolfinger and Lin [18] recommended simulation studies for assessing the performance of the methods in diverse kinds of non-linear models and data sets.

The aim of this work was to describe and examine the performance of the EBLUP-expansion method for the Gompertz function applied to the analysis of growth in the pig through simulation. The EBLUP-expansion is recommended especially for cases where the variance components are large, which may be the case for an adult weight parameter for pigs. Also, Lindstrom and Bates [9]

suggested that the expansion around the expected zero value may lead to poor estimates when substantial inter-individual variation exists. We chose to examine the method through the analysis of two data sets. The first analysis tested general performance of the EBLUP-expansion technique, and the second analysis tested performance of the method for incomplete data. Incomplete data are common in pig production, where the adult weight is unavailable due to an earlier slaughter age.

2. MATERIALS AND METHODS

2.1. Simulations

The Gompertz function has been shown to fit pig growth data, such as live weight and protein retention, well [14–16]. We assumed that weights of an individual i followed the Gompertz model:

$$y_{ij} = \alpha \exp(-\beta \exp(-\kappa\, t_{ij})) + e_{ij}, \quad j = 1, \ldots, n_i$$

where n_i is the number of observations for individual i, y_{ij} is the observed weight at age t_{ij} (in days), α, β and κ are the parameters of the Gompertz function, and e_{ij} is the random residual. The parameters have biological meaning: α is the adult weight, κ is the rate of exponential decay of the initial growth rate, and β is the logarithm of the ratio of birth weight to adult weight.

Each of the parameters α, β and κ can be described by a linear mixed effects model. In this study, we will consider a sire model, although notation could be for an animal model. The full model for observation j of animal i is

$$
\begin{aligned}
y_{ij} = \ & (\mathbf{x}_{\alpha_{ij}}\mathbf{b}_\alpha + \mathbf{z}_{s,\alpha_i}\mathbf{s}_\alpha + \mathbf{z}_{p,\alpha_i}\mathbf{p}_\alpha) \\
& \exp(-(\mathbf{x}_{\beta_{ij}}\mathbf{b}_\beta + \mathbf{z}_{s,\beta_i}\mathbf{s}_\beta + \mathbf{z}_{p,\beta_i}\mathbf{p}_\beta) \\
& \exp(-(\mathbf{x}_{\kappa_{ij}}\mathbf{b}_\kappa + \mathbf{z}_{s,\kappa_i}\mathbf{s}_\kappa + \mathbf{z}_{p,\kappa_i}\mathbf{p}_\kappa)t_{ij})) + e_{ij},
\end{aligned} \tag{1}
$$

where $(\mathbf{b}_\alpha, \mathbf{b}_\beta, \mathbf{b}_\kappa)^T = \mathbf{b}$ is a $d \times 1$-vector of fixed effects, $(\mathbf{s}_\alpha, \mathbf{s}_\beta, \mathbf{s}_\kappa)^T = \mathbf{s}$ is a $l \times 1$-vector of random additive genetic sire effects and $(\mathbf{p}_\alpha, \mathbf{p}_\beta, \mathbf{p}_\kappa)^T = \mathbf{p}$ is a $q \times 1$-vector of random animal effects other than sire. Vectors $\mathbf{x}$, $\mathbf{z}_s$ and $\mathbf{z}_p$ are from the design matrices of fixed, random sire and random animal effects $\mathbf{X}$, $\mathbf{Z}_s$ and $\mathbf{Z}_p$, respectively. It is assumed that

$$
\begin{pmatrix} \mathbf{s} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{pmatrix} \right).
$$

Here, $\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$, where $\mathbf{A}$ is a matrix of additive relationships between sires and $\mathbf{G}_0$ is a 3×3 genetic covariance matrix for the Gompertz parameters. Similarly, $\mathbf{P} = \mathbf{P}_0 \otimes \mathbf{I}_q$, where $\mathbf{P}_0$ is a 3×3 covariance matrix, *i.e.* the random animal effects $\mathbf{p}$ were identically and independently distributed for the animals. Furthermore, the residuals were assumed to be independently distributed and homoscedastic, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_n \sigma_e^2)$.

The Gompertz function coefficients were generated to simulate pig growth. The model had one fixed effect with two levels, and the random effects were the genetic sire effect and the animal effect other than the sire. The first fixed effect level had values 210, 5 and 0.017 for the parameters α, β and κ, respectively. The other fixed effect level had values 220, 4.7 and 0.016 for α, β and κ, respectively. The random effects were assumed to be normally distributed with mean zero and block diagonal covariance matrices. Variance and covariance components in the matrices for the genetic sire effect and for the animal effect are shown in Tables I and II. The residual variance σ_e^2 was one. These parameters approximated the variances calculated by the NLMIXED procedure of the SAS® program that fitted the Gompertz model to growth performance data of Finnish pigs [12].

Simulation of the random sire effect required taking into account the pedigree. The pedigree had three generations of animals with 10 unrelated founder grandsires. Each of the 10 grandsires was mated with 20 unrelated dams that produced one son each, *i.e.*, 20 half-sibs. The half-sibs were mated with unrelated dams to produce 24 progeny per sire. Only the last generation of animals had records. Thus, the data included 4800 tested animals. Two data sets were made: a complete set, and a truncated time trajectory set. The complete data contained 30 equally-distanced observations per animal between 50 and 253 days. The truncated time trajectory data contained slaughter weights up to 115 kg, which is similar to the common slaughter weight in pigs, and occurs at about 120 days of age. Consequently, the number of observations was reduced from 30 to about 11 per animal, *i.e.*, almost two thirds of the data were discarded.

2.2. Method to estimate the values of the growth parameters

The non-linear mixed model considered was

$$\mathbf{y} = f(\mathbf{X}, \mathbf{b}, \mathbf{Z}_s, \mathbf{s}, \mathbf{Z}_p, \mathbf{p}) + \mathbf{e},$$

where $\mathbf{y}$ is an $n \times 1$-vector of observations, f is the Gompertz function, and $\mathbf{e}$ is an $n \times 1$-vector of random residuals. Vectors $\mathbf{b}$, $\mathbf{s}$ and $\mathbf{p}$, with matrices $\mathbf{X}$, $\mathbf{Z}_s$

Table I. Relative bias, relative standard deviation (Rel. SD) and relative mean squared error (Rel. MSE) (as percent from the true value) for (co)variance components of genetic sire effects from the 50 replicates of full and truncated time trajectory data. Subscripts α, β and κ denote the three parameters in the Gompertz function.

Parameter	True	Full data			Truncated data		
		Rel. Bias (%)	Rel. SD (%)	Rel. MSE (%)	Rel. Bias (%)	Rel. SD (%)	Rel. MSE (%)
σ^2_α	10.0	−1.9	15.9	25.0	4.7	41.3	169.5
σ^2_β	0.01	0.3	15.2	2.27e−02	7.1	13.6	2.31e−02
σ^2_κ	3.0e−07	3.0	15.6	7.41e−07	−17.1	33.4	4.16e−06
$\sigma_{\alpha\beta}$	−0.06	6.0	55.1	1.8	−19.9	84.9	4.5
$\sigma_{\alpha\kappa}$	−3.0e−04	−9.7	64.5	1.25e−02	71.4	196.2	0.1
$\sigma_{\beta\kappa}$	1.0e−05	21.7	65.7	4.70e−04	−8.9	75.0	5.59e−04

Table II. Relative bias, relative standard deviation (Rel. SD) and relative mean squared error (Rel. MSE) (as percent from the true value) for (co)variance components of the animal effect from the 50 replicates of full and truncated time trajectory data. Subscripts α, β and κ denote the three parameters in the Gompertz function.

Parameter	True	Full data			Truncated data		
		Rel. Bias (%)	Rel. SD (%)	Rel. MSE (%)	Rel. Bias (%)	Rel. SD (%)	Rel. MSE (%)
σ^2_α	90.0	2.98e−02	2.3	4.5	18.8	10.6	416.5
σ^2_β	0.09	−0.2	1.8	2.88e−03	2.2	2.9	1.16e−02
σ^2_κ	3.7e−06	0.2	2.1	8.91e−07	−0.9	3.0	3.47e−07
$\sigma_{\alpha\beta}$	−0.54	−1.9	9.5	0.5	−36.7	24.5	10.4
$\sigma_{\alpha\kappa}$	−3.75e−03	−1.5	6.5	1.64e−03	−13.3	25.5	3.05e−02
$\sigma_{\beta\kappa}$	9.0e−05	−0.9	9.9	8.72e−05	4.5	16.9	2.71e−04

and $\mathbf{Z}_p$, were defined as before. For the random effects, denote $\mathbf{u}^T = (\mathbf{s}^T \mathbf{p}^T)$, $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_s \ \mathbf{Z}_p \end{bmatrix}$ and

$$\mathbf{D} = \begin{pmatrix} \mathbf{G}_0 \otimes \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_0 \otimes \mathbf{I}_q \end{pmatrix}.$$

Now the distribution assumptions were

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{pmatrix} \right).$$

Although $\mathbf{R}$ is diagonal here, any form is allowed, so the general form $\mathbf{R}$ will be used hereinafter. Unknown elements of covariance matrices $\mathbf{G}_0$, $\mathbf{P}_0$ and $\mathbf{R}$ are denoted by parameter vector $\boldsymbol{\theta}$.

The maximized likelihood function was

$$L(\mathbf{b}, \boldsymbol{\theta}|\mathbf{y}) = (2\pi)^{-\frac{n}{2}} |\mathbf{R}|^{-\frac{1}{2}} (2\pi)^{-\frac{l+q}{2}} |\mathbf{D}|^{-\frac{1}{2}}$$
$$\int \exp\left(-\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u}))^T \mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u})) - \frac{1}{2}\mathbf{u}^T \mathbf{D}^{-1}\mathbf{u} \right) d\mathbf{u}. \quad (2)$$

Only in some cases is the closed form of (2) found, so the integral is often solved numerically. However, numerical methods for the non-linear functions are usually slow to converge and numerically unstable. Instead, the integral may be approximated by quadratic Taylor-series expansion of the exponent. The second-order expansion was made about the EBLUP before integration of the likelihood function (see Appendix). This gave approximation to the logarithm of the likelihood function (2):

$$l^*(\mathbf{b}, \boldsymbol{\theta}|\mathbf{y}) = -\frac{1}{2}n\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{R}||\mathbf{I} + \mathbf{Z}^{*T}\mathbf{R}^{-1}\mathbf{Z}^*\mathbf{D}|)$$
$$-\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T \mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})) - \frac{1}{2}\tilde{\mathbf{u}}^T \mathbf{D}^{-1}\tilde{\mathbf{u}}, \quad (3)$$

where $\mathbf{Z}^* = \partial f / \partial \mathbf{u}^T |_{\mathbf{u}=\tilde{\mathbf{u}}}$ and $\tilde{\mathbf{u}}$ is the empirical BLUP-estimate of random effects. For the Gompertz function and two random effects in the model, $\mathbf{Z}^*$ had elements

$$\frac{\partial f}{\partial \alpha_i} = \exp(-\beta_i \exp(-\kappa_i t_j))\, c_{\alpha_i}$$
$$\frac{\partial f}{\partial \beta_i} = \alpha_i \exp(-\beta_i \exp(-\kappa_i t_j))\,(-\exp(-\kappa_i t_j))\, c_{\beta_i} \quad (4)$$
$$\frac{\partial f}{\partial \kappa_i} = \alpha_i \exp(-\beta_i \exp(-\kappa_i t_j))\,(-\beta_i \exp(-\kappa_i t_j))\,(-t_j)\, c_{\kappa_i}$$

where c is z_s or z_p depending on the random effect differentiated (see (1)).

Pinheiro and Bates [10] used the approximation (3) in estimation of parameters by the Laplacian approximation. However, no straightforward generalization to the REML-estimation was presented. Wolfinger and Lin [18] developed formula (3) further. Denote $\mathbf{V} = \mathbf{Z}^*\mathbf{D}\mathbf{Z}^{*T} + \mathbf{R} \mid_{\mathbf{u}=\tilde{\mathbf{u}}}$. Then,

$$l^*(\mathbf{b}, \boldsymbol{\theta}|\mathbf{y}) = -\frac{1}{2}n \ln(2\pi) - \frac{1}{2} \ln |\mathbf{V}|$$
$$- \frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}) + \mathbf{Z}^*\tilde{\mathbf{u}})^T \mathbf{V}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}) + \mathbf{Z}^*\tilde{\mathbf{u}}),$$

where $\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}^*\mathbf{D}(\mathbf{I} + \mathbf{Z}^{*T}\mathbf{R}^{-1}\mathbf{Z}^*\mathbf{D})^{-1}\mathbf{Z}^{*T}\mathbf{R}^{-1}$ and $|\mathbf{V}| = |\mathbf{R}||\mathbf{I} + \mathbf{Z}^{*T}\mathbf{R}^{-1}\mathbf{Z}^*\mathbf{D}|$ [5]. This led to a similar estimation function for variance component estimation presented by Lindstrom and Bates [9], although through different derivation.

2.2.1. Estimation of the fixed and random effects

Assume that the variance component vector $\boldsymbol{\theta}$ is known. Maximum likelihood estimation for the parameters $\mathbf{b}$ and $\mathbf{u}$ leads to solving equations:

$$\mathbf{X}^{*T}\mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \tilde{\mathbf{b}}, \mathbf{Z}, \tilde{\mathbf{u}})) = \mathbf{0}$$
$$\mathbf{Z}^{*T}\mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \tilde{\mathbf{b}}, \mathbf{Z}, \tilde{\mathbf{u}})) = \mathbf{D}^{-1}\tilde{\mathbf{u}},$$

$$(5)$$

where $\mathbf{X}^* = \partial f / \partial \mathbf{b}^T \big|_{\mathbf{b}=\tilde{\mathbf{b}}}$ and $\tilde{\mathbf{b}}$ is the estimate of fixed effects $\mathbf{b}$. Elements of $\mathbf{X}^*$ are similar to $\mathbf{Z}^*$, except that coefficient c in (4) is replaced by x due to the differentiated fixed effect. However, in order to arrive to these simple equations, dependency of $\mathbf{V}$ on $\mathbf{b}$ through $\mathbf{Z}^*$ has to be ignored. On the basis of arguments made by Bates and Watts [1], Wolfinger and Lin [18] justified this by appealing to intrinsic non-linearity instead of non-linearity of the parameters.

Denote $\mathbf{Y} = \mathbf{y} - f(\mathbf{X}, \tilde{\mathbf{b}}, \mathbf{Z}, \tilde{\mathbf{u}}) + \mathbf{X}^*\tilde{\mathbf{b}} + \mathbf{Z}^*\tilde{\mathbf{u}}$. Equations (5) can now be written as

$$\begin{bmatrix} \mathbf{X}^{*T}\mathbf{R}^{-1}\mathbf{X}^* \ \mathbf{X}^{*T}\mathbf{R}^{-1}\mathbf{Z}^* \\ \mathbf{Z}^{*T}\mathbf{R}^{-1}\mathbf{X}^* \ \mathbf{Z}^{*T}\mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{G}^{-1} \end{bmatrix} \begin{pmatrix} \tilde{\mathbf{b}} \\ \tilde{\mathbf{u}} \end{pmatrix} = \begin{bmatrix} \mathbf{X}^{*T}\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}^{*T}\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix}. \qquad (6)$$

This is similar to the mixed model equations (MME) for the linear models. Thus, already established methods for solving linear models can be used to analyse the pseudo-data $\mathbf{Y}$ created from the original data $\mathbf{y}$ with $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{u}}$ equal to their most recent estimates.

2.2.2. *Estimation of the variance components*

After finding estimates of the location parameters, profile likelihood can be used to estimate the variance components by setting $\mathbf{b} = \tilde{\mathbf{b}}(\boldsymbol{\theta})$. The logarithmic likelihood function of the parameter vector $\boldsymbol{\theta}$ can be written with the pseudo-data as

$$l^*_{ML}(\boldsymbol{\theta}) = -\frac{1}{2}n\ln(2\pi) - \frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}}). \tag{7}$$

Differentiating equation (7) with respect to $\boldsymbol{\theta}$ gives

$$-\frac{1}{2}\mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\boldsymbol{\theta}_j}\right) + \frac{1}{2}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}})^T\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\boldsymbol{\theta}_j}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}}). \tag{8}$$

Maximum likelihood estimates of variance components are found by equating (8) to zero and solving for $\boldsymbol{\theta}$.

Instead of the ML-estimates, REML-estimates are commonly used in practise. These estimates account for losses in degrees of freedom caused by the estimation of fixed effects $\mathbf{b}$ [5]. The logarithmic likelihood function is now

$$l^*_{REML}(\boldsymbol{\theta}) = -\frac{1}{2}n\ln(2\pi) - \frac{1}{2}\ln|\mathbf{V}| - \frac{1}{2}\ln|\mathbf{X}^{*T}\mathbf{V}^{-1}\mathbf{X}^*|$$
$$-\frac{1}{2}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}}). \tag{9}$$

Differentiation with respect to $\boldsymbol{\theta}$ and equating to zero gives

$$-\frac{1}{2}\mathrm{tr}\left(\mathbf{P}\frac{\partial\mathbf{V}}{\partial\boldsymbol{\theta}_j}\right) + \frac{1}{2}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}})^T\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\boldsymbol{\theta}_j}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}^*\tilde{\mathbf{b}}) = \mathbf{0}, \tag{10}$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}^*(\mathbf{X}^{*T}\mathbf{V}^{-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{V}^{-1}$. Solutions in $\boldsymbol{\theta}$ are REML-estimates of variance components.

2.2.3. *EBLUP-algorithm*

The approximate ML-solutions of location parameters and variance components can be obtained by iteratively solving the equations (6) and (8) until convergence. Correspondingly, the REML-solutions for the EBLUP-expansion are obtained by iteratively solving the equations (6) and (10). Hence, the algorithm fits the linear mixed effects model $\mathbf{Y} = \mathbf{X}^*\mathbf{b} + \mathbf{Z}^*\mathbf{u} + \mathbf{e}$ for the pseudo-data $\mathbf{Y}$ and the working vectors $\mathbf{X}^*$ and $\mathbf{Z}^*$, where $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}))$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}))$.

2.3. Implementation

We chose to implement the REML-based EBLUP-algorithm, because programs to solve the linear mixed effects models are available and commonly used by animal breeders. MiX99 [13] was used to solve the mixed model equations (6), and DMU [6], modified for random regression by Kettunen *et al.* [7], was used to solve the REML estimates of covariance components (10). The capability of fitting fixed and random regression models is crucial for implementation, because the coefficients in $\mathbf{X}^*$ and $\mathbf{Z}^*$ can have any values. Implementation of the linearization procedure required the Gompertz function formulas to be included in MiX99. However, there was no need to make changes to the variance component estimation program.

Starting values for both the location parameter effects and the variance components had to be assigned before first iteration. A natural choice was to initialize random effects with the expected value zero. However, initial values for fixed effects were derived with the model function of growth curve and available data. When the Gompertz model is used, only the asymptotic weight parameter has a natural initial value, which is the maximum value of the dependent variable. In the other cases, complex equations were derived in order to have a stable algorithm. Initial values for covariance matrices of the genetic sire effect and animal effect were diagonal matrices having values diag{100,10,1}. The initial value for the residual variance was 100.

Additionally, variance components were reparametrized for computational reasons, because the variance component κ was close to zero. Convergence was improved by scaling the time before every round of the EBLUP-algorithm. Each time of measurement t_{ij} was multiplied by a scaling factor c, which was set equal to the most recent estimate of κ. Consequently, the variance component estimate for the scaled parameter κ^* was $\frac{1}{c^2}Var(\kappa)$, and thus larger than the original parameter κ when $c < 1$.

Convergence of the EBLUP-algorithm was assumed when the relative round to round change was less than 10^{-3}. Furthermore, within every iteration of the EBLUP-algorithm, the location parameters were iterated until the relative difference between right-hand and left-hand sides of the MME was less than 1×10^{-7}. Covariance component estimates were calculated by the Expectation Maximization (EM) -algorithm, and convergence was assumed when the round to round change was less than 5×10^{-7}.

The results are from 50 simulation replicates. Relative bias, relative standard deviation (Rel. SD) and relative mean squared error (Rel. MSE), as percentage from the true value, were calculated for the difference of two levels of fixed effect and for the variance component parameter estimates. The relative bias

was calculated as (mean-true)/|true|, where mean is the arithmetic average of 50 estimates and true is the original value used to generate the data.

3. RESULTS

3.1. Complete data

The average number of iterations of the EBLUP-algorithm was 7 in the full data simulations. The residual error variance converged well and its estimate was equal to the original value used in the simulations (relative bias and relative standard deviation were 0% and 0.5%, respectively).

The estimated (co)variance components of the genetic sire effects were in fairly good agreement with the original parameter values used to simulate the data (Tab. I). Both the relative bias and the relative SD were higher for the covariance components (12.5% and 61.8% on average, respectively) than for the variance components (1.7% and 15.6% on average, respectively).

The estimated (co)variance components of animal effects were more accurately estimated than for the genetic sire effects (Tab. II). Estimated variance components had negligible relative bias and an average relative SD of 2.1%, but covariance components had an average relative bias of 1.4% and an average relative SD of 8.6%.

Simulation results for the difference of the two levels of fixed effects in the model are shown in Table III. The results of the parameters β and κ showed fairly good agreement with the initial values, with relative bias (Rel. SD) being 3.0% (5.2%). However, estimates for the parameter α were slightly biased. Relative bias was −10.0% and relative SD was 6.1%.

3.2. Truncated time trajectory data

The average number of iterations of the EBLUP-algorithm increased from 7 to 8 for the truncated time trajectory data. Residual error variance converged in this case as well and its estimate was equal to the original input value for the simulations (the relative bias and the relative standard deviation as percentages were −0.2% and 0.6%, respectively).

Compared to the full data, analysis for the truncated time trajectory data showed larger bias and SD for both the (co)variance components of genetic sire effect (Tab. I) and animal effect (Tab. II). The genetic sire effect had average relative bias (Rel. SD) of 9.6% (29.4%) for the estimated variance components and 33.4% (118.7%) for the estimated covariance components. The estimation

Table III. Relative bias, relative standard deviation (Rel. SD) and relative mean squared error (Rel. MSE) (as percent from the true value) for difference of two levels of fixed effect from the 50 replicates of full and truncated time trajectory data. Subscripts α, β and κ denote the three parameters in the Gompertz function and the following numbers denote the level of fixed effect.

Parameter	True	Full data			Truncated data		
		Rel. Bias (%)	Rel. SD (%)	Rel. MSE (%)	Rel. Bias (%)	Rel. SD (%)	Rel. MSE (%)
$b_{\alpha,1} - b_{\alpha,2}$	−10.0	−10.0	6.1	13.6	14.1	12.6	35.6
$b_{\beta,1} - b_{\beta,2}$	0.3	2.2	3.2	4.40e−02	−3.8	4.5	0.1
$b_{\kappa,1} - b_{\kappa,2}$	0.001	3.7	7.2	6.45e−04	−7.3	10.1	1.53e−03

of covariance between α and κ parameters was especially unstable. Animal effects had an average relative bias (Rel. SD) of 7.3% (5.5%) for the estimates of variance components. This increase was mostly due to increased uncertainty with parameter α. For the same reason, average relative bias (Rel. SD) increased to 18.2% (22.3%) for the estimates of covariance components.

Simulation results for the difference of the two fixed effect levels corresponded to results with the full data (see Tab. III). The results of parameters β and κ again showed fairly good agreement with the initial values, but the results for the parameter α were 14% biased. For the truncated time trajectory data, both bias and SD were on average 60% larger than for the full data.

4. DISCUSSION

The results from the simulation showed fairly good agreement with the original values for the data, when observations from the whole growing period were available. The largest discrepancies were seen in the estimates of covariance components for the genetic effect. When the animals were slaughtered prematurely, the adult weight was not reached and the latter part of the growth function curve contained no data. This especially influenced the estimation of (co)variance components related to adult weight. For the genetic effect, the uncertainty of estimation was also seen in the (co)variance components related to the exponential decay of the initial growth rate.

A direct comparison of our study to the literature cannot be made. Furthermore, the non-linear curves in the animal breeding literature generally rely on Bayesian analysis for case-specific problems [2, 11], and therefore comparisons are difficult to make. However, Wolfinger and Lin [18] considered a logistic model where variance component estimation results were similar to our case with full data. With respect to the fixed effects, the results were more biased for the Gompertz model than for the logistic model. For the truncated time trajectory data, no earlier results to compare were found in the scientific literature, where observations were available far over the inflection point [2, 8]. This improves the quality of results compared to our truncated data, where the slaughter time exceeded the inflection point only slightly.

The initial parameter values for simulation of pig growth in our study are approximations from field data. Correct values may differ, but the close proximity of slaughter time and the inflection point is a common situation in field data. However, there may be pigs that have observations until adult weight. This is due to selection of tested pigs as breeding animals. To test the effect of partial truncation, the principles of full and truncated time trajectory data

simulations were combined, *i.e.* approximately 5% of the tested progeny had observations until day 253, whereas observations from the rest of the progeny were truncated at 115 kg weight. Even a small proportion of fully observed animals improved the results when compared to the estimates produced from completely truncated data. For the genetic effect, improvements were especially shown as smaller relative bias and standard deviations for the covariance components and the variance component of parameter κ. For the variance and covariance components of genetic sire effect, the average relative biases (Rel. SD) were 3.4% (20.2%) and 6.6% (74.0%), respectively. For the animal effect, smaller biases were seen for the variance and covariance components of parameter α. The average relative bias (Rel. SD) was 4.6% (3.7%) for the variance components and 10.2% (13.5%) for the covariance components. We cannot make general recommendations about the proportion of animals with full data, because the results may be influenced by the population structure.

Starting values are important for the non-linear models in order for the algorithm to converge. We tried different strategies for defining the starting values, but general and simple equations were not discovered. Thus, the convergence of the algorithm with the presented parametrization depends on the proper starting values. However, the convergence may be improved by a different parametrization of the Gompertz model. Alternatively, a completely multiplicative model using log-transformed data can be analysed. This takes account of the common nature of the residuals in real growth data. In addition, log-transformation removes the dependence of the derivative of the adult weight parameter on the others.

The procedure presented is similar to that commonly used in animal breeding for linear models. The variance components estimated by REML are used in the mixed model equations to solve the location parameters. Consequently, even large models and data sets can be analysed when the variance components are assumed known. Easy implementation in already existing programs for linear mixed effects models is an advantage, although the two-step iterative procedure with each step itself being iterative can be regarded as computationally intensive. Another advantage of the method presented by Wolfinger and Lin is generality. It can be used for different types of models because it is developed for general non-linear mixed models. Also, generalization of this simple model to have multiple effects and traits is straightforward.

Linearization enables the linear mixed model procedures given the validity of the linear approximation. Simulation study is a convenient way to verify the appropriateness of the approximation method to specific situations. The full data set in our simulations shows that linearization works moderately well for

the Gompertz function. Some enhancement may be achieved by reparametrization, which is a subject for additional research. Another motivation for linearization is allowance for sparse data. This is common in field data, where varying amounts of information are available for the animals. The truncated data analysis showed, however, that if observations are missing from the tails of all animal growth curves, uncertainty increases and the estimation method can be distorted. This distortion diminished considerably when at least some of the animals had observations until or close to their mature weight. Therefore, the success of the Gompertz model greatly depends on the amount and nature of available information.

REFERENCES

[1] Bates D.M., Watts D.G., Relative curvature measures of nonlinearity, J. Roy. Stat. Soc. B 42 (1980) 1–25.

[2] Blasco A., Piles M., Varona L., A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits, Genet. Sel. Evol. 35 (2003) 21–41.

[3] Breslow N.E., Clayton D.G., Approximate inference in generalized linear mixed models, J. Am. Stat. Assoc. 88 (1993) 9–25.

[4] Davidian M., Giltinan D.M., Nonlinear Models for Repeated Measurement Data, Chapman & Hall, London, 1995.

[5] Harville D.A., Maximum likelihood approaches to variance component estimation and to related problems, J. Am. Stat. Assoc. 72 (1977) 320–338.

[6] Jensen J., Madsen P., DMU: A package for the analysis of multivariate mixed models, in: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, 7–12 August 1994, Vol. 22, University of Guelph, Guelph, pp. 45–46.

[7] Kettunen A., Mäntysaari E.A., Pösö J., Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models, Livest. Prod. Sci. 66 (2000) 251–261.

[8] Lewis R.M., Emmans G.C., Dingwall W.S., Simm G., A description of the growth of sheep and its genetic analysis, Anim. Sci. 74 (2002) 51–62.

[9] Lindstrom M.J., Bates D.M., Nonlinear mixed effects models for repeated measures data, Biometrics 46 (1990) 673–687.

[10] Pinheiro J.C., Bates D.M., Approximations to the log-likelihood function in the nonlinear mixed-effects model, J. Comput. Grap. Stat. 4 (1995) 12–35.

[11] Rekaya R., Weigel K.A., Gianola D., Hierarchical nonlinear model for persistency of milk yield in the first three lactations of Holsteins, Livest. Prod. Sci. 68 (2001) 181–187.

[12] Sevón-Aimonen M.-L., The Parameters of Growth Curve and Composition of Growth for Finnish Pigs, Book of Abstracts of the 52nd Annual Meeting of the European Association for Animal Production, 25–29 August 2001, Wageningen Press, The Netherlands, p. 290.

[13] Strandén I., Lidauer M., Solving large mixed linear models using preconditioned conjugate gradient iteration, J. Dairy Sci. 82 (1999) 2779–2787.

[14] Wellock I.J., Emmans G.C., Kyriazakis I., Describing and predicting potential growth in pig, Anim. Sci. 78 (2004) 379–388.

[15] Whittemore C.T., Green D.M., The description of the rate of protein and lipid growth in pigs in relation to live weight, J. Agric. Sci. 138 (2002) 415–423.

[16] Whittemore C.T., Tullis J.B., Emmans G.C., Protein growth in pigs, Anim. Prod. 46 (1988) 437–445.

[17] Wolfinger R.D., Laplace's approximation for nonlinear mixed models, Biometrika 80 (1993) 791–795.

[18] Wolfinger R.D., Lin X., Two Taylor-series approximation methods for nonlinear mixed models, Comput. Stat. Data Anal. 25 (1997) 465–490.

APPENDIX

Linearization of the likelihood function

Assume $\mathbf{y}|\mathbf{u} \sim \mathrm{N}(f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u}), \mathbf{R})$, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$. Then, the likelihood function to be maximized is the following:

$$L(\mathbf{b}, \boldsymbol{\theta}|\mathbf{y}) = (2\pi)^{-\frac{n}{2}}|\mathbf{R}|^{-\frac{1}{2}}(2\pi)^{-\frac{l+q}{2}}|\mathbf{D}|^{-\frac{1}{2}}$$
$$\int \exp\left(-\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u}))^T\mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u})) - \frac{1}{2}\mathbf{u}^T\mathbf{D}^{-1}\mathbf{u}\right)\mathrm{d}\mathbf{u}.$$

The term in the exponent to be integrated can be linearized using a second-order Taylor series expansion around the predicted value of random effects $\mathbf{u}$:

$$-\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u}))^T\mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \mathbf{u})) - \frac{1}{2}\mathbf{u}^T\mathbf{D}^{-1}\mathbf{u}$$

$$\approx -\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T\mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})) - \frac{1}{2}\tilde{\mathbf{u}}^T\mathbf{D}^{-1}\tilde{\mathbf{u}}$$

$$+ \left[(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T\mathbf{R}^{-1}f'(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}) - \mathbf{D}^{-1}\tilde{\mathbf{u}}\right](\mathbf{u} - \tilde{\mathbf{u}})$$

$$+ \frac{1}{2}(\mathbf{u} - \tilde{\mathbf{u}})\left[-f'(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})^T\mathbf{R}^{-1}f'(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})\right.$$

$$\left.+ (\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T\mathbf{R}^{-1}f''(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}) - \mathbf{D}^{-1}\right](\mathbf{u} - \tilde{\mathbf{u}})$$

$$\approx -\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T\mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})) - \frac{1}{2}\tilde{\mathbf{u}}^T\mathbf{D}^{-1}\tilde{\mathbf{u}}$$

$$- \frac{1}{2}(\mathbf{u} - \tilde{\mathbf{u}})^T\left[\mathbf{Z}^*\mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{D}^{-1}\right](\mathbf{u} - \tilde{\mathbf{u}}).$$

Here $\mathbf{Z}^* = \partial f / \partial \mathbf{u}^T |_{\mathbf{u}=\tilde{\mathbf{u}}}$ and $\tilde{\mathbf{u}}$ is the empirical BLUP-estimate of the random effects. In addition, the linear term in the expansion vanishes, because the first derivative of the function at ML-solutions is zero. Also $(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T \mathbf{R}^{-1} f''(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})$ is assumed to be negligible, because the residual vector $(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T \mathbf{R}^{-1}$ has mean zero.

Now, approximation for the likelihood function L is

$$
\begin{aligned}
L^*(\mathbf{b}, \boldsymbol{\theta}|\mathbf{y}) &= (2\pi)^{-\frac{n}{2}} |\mathbf{R}|^{-\frac{1}{2}} (2\pi)^{-\frac{l+q}{2}} |\mathbf{D}|^{-\frac{1}{2}} \\
&\quad \int \exp\left(-\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T \mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})) \right. \\
&\quad \left. -\frac{1}{2}\tilde{\mathbf{u}}^T \mathbf{D}^{-1}\tilde{\mathbf{u}} - \frac{1}{2}(\mathbf{u} - \tilde{\mathbf{u}})^T \left[\mathbf{Z}^* \mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{D}^{-1}\right](\mathbf{u} - \tilde{\mathbf{u}})\right) d\mathbf{u} \\
&= (2\pi)^{-\frac{n}{2}} |\mathbf{R}|^{-\frac{1}{2}} |\mathbf{D}|^{-\frac{1}{2}} \left|\mathbf{Z}^* \mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{D}^{-1}\right|^{-\frac{1}{2}} \\
&\quad \exp\left(-\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T \mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})) - \frac{1}{2}\tilde{\mathbf{u}}^T \mathbf{D}^{-1}\tilde{\mathbf{u}}\right) \\
&\quad \int (2\pi)^{-\frac{l+q}{2}} \left|\mathbf{Z}^* \mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{D}^{-1}\right|^{\frac{1}{2}} \\
&\quad \times \exp\left(-\frac{1}{2}(\mathbf{u} - \tilde{\mathbf{u}})^T \left[\mathbf{Z}^* \mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{D}^{-1}\right](\mathbf{u} - \tilde{\mathbf{u}})\right) d\mathbf{u} \\
&= \exp\left(-\frac{1}{2}n\ln(2\pi) - \frac{1}{2}\ln|\mathbf{R}| - \frac{1}{2}\ln|\mathbf{D}| - \frac{1}{2}\ln\left|\mathbf{Z}^* \mathbf{R}^{-1}\mathbf{Z}^* + \mathbf{D}^{-1}\right| \right. \\
&\quad \left. -\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T \mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})) - \frac{1}{2}\tilde{\mathbf{u}}^T \mathbf{D}^{-1}\tilde{\mathbf{u}}\right)
\end{aligned}
$$

and the logarithm of $L^*(\mathbf{b}, \boldsymbol{\theta}|\mathbf{y})$ is

$$
\begin{aligned}
l^*(\mathbf{b}, \boldsymbol{\theta}|\mathbf{y}) &= -\frac{1}{2}n\ln(2\pi) - \frac{1}{2}\ln(|\mathbf{R}||\mathbf{I} + \mathbf{Z}^{*T}\mathbf{R}^{-1}\mathbf{Z}^*\mathbf{D}|) \\
&\quad -\frac{1}{2}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}}))^T \mathbf{R}^{-1}(\mathbf{y} - f(\mathbf{X}, \mathbf{b}, \mathbf{Z}, \tilde{\mathbf{u}})) - \frac{1}{2}\tilde{\mathbf{u}}^T \mathbf{D}^{-1}\tilde{\mathbf{u}}.
\end{aligned}
$$