Longitudinal Analysis of Body Weight and Feed Intake in Selection Lines for Residual Feed Intake in Pigs*

W. Cai[1,2], H. Wu[2], and J. C. M. Dekkers[1], **

[1] Department of Animal Science, Iowa State University, Ames 50011, USA

ABSTRACT : A selection experiment for reduced residual feed intake (RFI) in Yorkshire pigs consisted of a line selected for lower RFI (LRFI) and a random control line (CTRL). Longitudinal measurements of daily feed intake (DFI) and body weight (BW) from generation 5 of this experiment were used. The objectives of this study were to evaluate the use of random regression (RR) and non-linear mixed models to predict DFI and BW for individual pigs, accounting for the substantial missing information that characterizes these data, and to evaluate the effect of selection for RFI on BW and DFI curves. Forty RR models with different-order polynomials of age as fixed and random effects, and with homogeneous or heterogeneous residual variance by month of age, were fitted for both DFI and BW. Based on predicted residual sum of squares (PRESS) and residual diagnostics, the quadratic polynomial RR model was identified to be best, but with heterogeneous residual variance for DFI and homogeneous residual variance for BW. Compared to the simple quadratic and linear regression models for individual pigs, these RR models decreased PRESS by 1% and 2% for DFI and by 42% and 36% for BW on boars and gilts, respectively. Given the same number of random effects as the polynomial RR models, i.e., two for BW and one for DFI, the non-linear Gompertz model predicted better than the polynomial RR models but not as good as higher order polynomial RR models. After five generations of selection for reduced RFI, the LRFI line had a lower population curve for DFI and BW than the CTRL line, especially towards the end of the growth period. (**Key Words :** Longitudinal Analysis, Pigs, Residual Feed Intake, Selection)

INTRODUCTION

Feed efficiency is a very important economic trait in swine production. For production purposes, feed efficiency is usually defined as the ratio of average daily body weight gain (ADG) to average daily feed intake (ADFI). In recent

years, the swine industry has increasingly used electronic feeders in recording individual daily feed intake (DFI) and body weight (BW) on group-housed pigs to improve feed efficiency. Measurements of DFI and BW are typically longitudinal along the growth trajectory of pigs.

Longitudinal measurements of DFI and BW for a given pig tend to be correlated. One simple approach to deal with longitudinal measurements is to reduce the longitudinal measures to a single summary for each animal and then analyze each summary variable. For example, longitudinal DFI records are often summarized as ADFI for the further analysis. Diggle et al. (2002) referred to this as two-stage analysis. Random regression models (Schaeffer and Dekkers, 1994) are other suitable option for analysis of longitudinal data on DFI and BW. Such models use data from all pigs simultaneously and allow estimation of individual and population curves. Schaeffer (2004) presented a thorough review on the application of random regression (RR) models in animal breeding. As one of the first applications to data other than milk production in cattle, Andersen and Pedersen (1996) applied RR models to analyze growth and food intake curves for pigs. Non-linear

mixed models, e.g., the Logistic and Gompertz models, are another option for analysis of DFI and BW data. Whittemore et al. (1988) used the Gompertz function to model body weight of pigs over time on a pig-by-pig basis. Ratkowsky (1990) presented a thorough review of commonly used non-linear regression models (e.g., the Logistic and Gompertz models) and their statistical properties.

In practice, longitudinal measurements of DFI and BW data are often missing for substantial parts of growth period, including at the beginning and end. Major missing data come from switching of pigs between electronic and commercial feeders to enlarge the test capacity because of high expense of electronic feeders (Von Felde et al., 1996; Eissen et al., 1999; Schulze et al., 2001; Casey, 2003), and data errors and malfunction of electronic feeders (Eissen et al., 1998; Casey et al., 2005). Several studies have shown that missing DFI information on different parts of the growth period had a limited effect on the accuracy of evaluating ADFI (Eissen et al., 1999; Casey, 2003). However, there have been few studies on the effect of missing data on evaluation of feed intake and body weight curves, which requires sophisticated statistical models to inter- and extrapolate DFI and BW curves for individual pigs. Thus, the first objective of this study was to develop and compare RR models and non-linear mixed models for analysis of DFI and BW in pigs with substantial missing data in order to identify the best model to predict DFI and BW curves for individual pigs. Data used for this study are from a selection experiment in Yorkshire pigs for reduced residual feed intake (RFI) at Iowa State University (Cai et al., 2008). The selection experiment consists of a line selected for lower residual feed intake (LRFI) for 5 generations and a randomly selected control line (CTRL). Thus, the second objective of this study was to evaluate the effect of selection for reduced RFI on DFI and BW population curves. These studies will bring opportunities for swine industry to directly select growth and feed intake curves to improve feed efficiency.

MATERIALS AND METHODS

Experimental design and data collection

Pigs from the 5th generation of the LRFI and CTRL lines of the residual feed intake selection experiment conducted at Iowa State University were used in this study. All procedures with pigs were approved by the Iowa State University Institutional Animal Care and Use Committee. The protocol of the selection experiment was specified in detail by Cai et al. (2008). Selection was based on estimated breeding value (EBV) for RFI. The data used in this study follow a randomized complete block design with line (LRFI vs. CTRL) as the investigating factor and pen as the block factor. A total of 192 boars from the first parity of generation 5 were put into 12 pens at ~90 d of age and ~40 kg of body weight for evaluation of growth and feed intake. Sixteen boars from the LRFI and CTRL lines were assigned to each pen by body weight and age, balancing to the extent possible across line within pen. Pigs that got sick or died were removed from their pens and pigs were taken off test on an individual basis when they reached 115 kg of body weight. When only three pigs were left in a pen, they were all taken off test, resulting in some lighter off-test body weights. A total of 151 pigs, 64 LRFI and 87 CTRL line boars with off-test body weight greater than 102 kg were used for analysis.

The experiment was replicated using 192 gilts from the second parity of generation 5. The same boars and sows that produced parity 1 of generation 5 were used to produce these gilts with the same mating design as parity 1. Besides gender, the only difference between the protocols from these two replicated experiments was that in order to get sufficient numbers of gilts for slaughter, gilts were off-tested in three groups instead of on an individual basis. A total of 162 pigs, 75 LRFI and 87 CTRL line gilts, with off-test body weight greater than 102 kg were used for analysis.

Six of the twelve pens were equipped with one single-space FIRE® feeders for feed intake recording. To allow all pigs to obtain feed intake data, pens were switched every 2 wk. Alternate pens were in the same room and had feeding equipment equivalent to the FIRE® feeders so as not to induce an acclimation period. The feed intake data from the day of switching were not used. Body weights were measured bi-weekly. Longitudinal measurements of DFI and BW data on these pigs were from ~3 to ~8 months of age. The average number of measurements of BW and DFI per pig is shown in Table 1.

Model selection and statistical analysis

Data from the two parities were analyzed separately using random regression and non-linear mixed model analyses. For comparison, data were also analyzed using simple linear and quadratic models fitted on a pig-by-pig basis. The models used and the process of model selection

Table 1. Frequency of measurements on daily feed intake and body weight per pig

Number of measurements	Body weight				Daily feed intake			
	7	8	9	10	25-50	51-60	61-70	71-85
Number of boars	36	50	36	29	21	52	55	23
Number of gilts	41	2	77	42	91	35	28	8

will be described in the following.

Simple individual pig models : Cai et al. (2008) fitted simple quadratic and linear regressions of DFI and BW against days of age for each pig separately. For the purpose of comparison, these two simple regression models on a pig-by-pig basis were also fitted in this study by the MIXED procedure of SAS (SAS Institute, 2008).

Random regression models : Let Y_{hijk} denote either BW or DFI at k days of age (from 64 to 230 days for boars and from 80 to 253 days for gilts) for pig j (j = 1,2,…,151 for boars and j = 1,2,…,162 for gilts) of line i (i = 1,2; 1 is LRFI and 2 is CTRL) raised in pen h (h = 1,2,…,12). For numerical reasons, age was adjusted as t_k = (age-90)/100, where 90 is the average on-test age (days). Random regression models with different-order polynomials of age as fixed and random effects, and with homogeneous residual variance, were fitted for both DFI and BW using the MIXED procedure of SAS (SAS Institute, 2008). Taking the model with quadratic polynomials of age as fixed and random effects as an example, the model can be denoted as:

$$y_{hijk} = Pen_h + \beta_{0i} + \beta_{1i}t_k + \beta_{2i}t_k^2 + \gamma_{0j} + \gamma_{1j}t_k + \gamma_{2j}t_k^2 + \varepsilon_{jk},$$

where $\beta_{0i} + \beta_{1i}t_k + \beta_{2i}t_k^2$ are the fixed effects representing the population curve; $\gamma_{0j} + \gamma_{1j}t_k + \gamma_{2j}t_k^2$ are the random effects representing the individual pig curve; *Pen* is the fixed block effect to account for systematic difference between pens and feeding stations. The distribution assumptions for the random effects were multivariate normal:

$$\gamma_j = \begin{pmatrix} \gamma_{0j} \\ \gamma_{1j} \\ \gamma_{2j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \right),$$

which was independent of residuals $\varepsilon_{jk} \sim N(0, \sigma_e^2)$. The design vector of random coefficients for the kth observation on the jth pig was $Z_{jk} = (1, t_k, t_k^2)$. The method of restricted maximum likelihood (REML) was used to estimate the variance components. Based on this model, the variance of the response variable Y changes with age as: $Var(Y_{hijk}) = Z_{jk}Var(\gamma_j)Z_{jk}' + Var(\varepsilon_{jk})$.

In matrix form, the model is written as Y = Xβ+Zγ+ε with $E(Y) = X\beta$ and $Var(Y) = V = ZGZ' + R$, where γ~N(0,G) and ε~N(0,R). Estimates of β and γ can be written as $\hat{\beta} = (X'\hat{V}^{-1}X)^- X'\hat{V}^{-1}y$ and $\hat{\gamma} = \hat{G}Z'\hat{V}^{-1}(y - X\hat{\beta})$, respectively. The variance of the response at different ages based on the RR model can be estimated by substituting the estimated variance components for G and R in the above equation.

A total of 20 different RR models were fitted by varying the order of fixed and random polynomials. Fixed-effect polynomials of age were fitted up to the 5th order, while random-effect polynomials of age were fitted up to the highest order polynomials of the fixed effects in the model. That is, if the fixed effect in the model was a quadratic polynomial, three different random-effect polynomials were fitted: i) intercept only; ii) intercept and linear term of age; iii) intercept, linear and quadratic term of age. The same set of 20 linear mixed models were also fitted with heterogeneous residual variances by month of age, allowing for a different residual variance for each of 6 months of age in view of possible different variation in different periods of growth. Six different periods (each of 6 months of age) were chosen to balance between the numbers of residual variance parameters to estimate and the numbers of available observations within each period for estimating them. Residual variances in different periods were assumed independent of each other. This is the same specification as models with homogeneous residual variances because the covariance between the observations had already been accounted by the covariance between random coefficients.

Non-linear mixed models : Logistic and Gompertz non-linear mixed models were fitted to the DFI and BW data using the NLMIXED procedure of SAS (SAS Institute, 2008). The random effects were estimated by the empirical Bayes method. Using the adaptive Gaussian quadrature method to do integral approximations, the NLMIXED procedure maximized the approximate likelihood integrated over the random effects. The number of quadrature points was set to 50 and the dual quasi-Newton algorithm was used as the optimization technique. Model details are described below.

To avoid convergence problems, the DFI and BW data were pre-adjusted for the effect of pen (12 pens in this study) before non-linear-model analysis, using the estimates from the selected RR models for DFI and BW, which were those with quadratic polynomials of age as both fixed and random effects. Then, for BW, the Logistic and Gompertz non-linear mixed models were fitted using the following non-linear mixed model equations: $Y_{ijk} = \alpha_{1ij}/(1+\exp(-(age_k - \alpha_{2i})/\alpha_{3ij})) + \varepsilon_{ijk}$ for the Logistic model and $Y_{ijk} = \alpha_{1ij} \times \exp(-\exp(-(age_k - \alpha_{2i})/\alpha_{3ij})) + \varepsilon_{ijk}$ for the Gompertz model. For both models, α_{1ij} represents the mature body weight for pig *j* in line *i*; α_{2i} represents the fixed inflection point (number of days) for line *i*; α_{3ij} represents the decay parameter for pig *j* in line *i*. Only two random effects α_{1ij} and α_{3ij} were fitted because of convergence problems when more random effects were fitted. Distribution assumptions for the random effects were multivariate normal:

$$\alpha = \begin{pmatrix} \alpha_{1ij} \\ \alpha_{3ij} \end{pmatrix} \sim iid\ N\left(\begin{pmatrix} \alpha_{1i} \\ \alpha_{3i} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right) ,$$

which were independent of $\varepsilon_{ijk} \sim N(0,\sigma_e^2)$. Because the random effects enter both the Logistic and Gompertz mixed models non-linearly, the variance of the response for the kth observation on the jth pig at different ages was derived by the delta method (Casella and Berger, 2002).

For DFI, the same Logistic and Gompertz non-linear mixed models were fitted as for BW, except that only one random effect α_{1ij} was fitted for DFI because of convergence problems. Therefore, for both the Logistic and Gompertz non-linear models of DFI, α_{3i} represents the fixed decay parameter across pigs for line i, in contrast to α_{3ij} for BW, which represented the decay parameter for pig j in line i. The distribution assumption for the random effect was $\alpha_{1ij} \sim N(\alpha_{1i},\sigma_{11})$, independent of $\varepsilon_{ijk} \sim N(0,\sigma_e^2)$.

Model selection : Model comparison and selection was based on statistics of predicted residual sum of squares (PRESS) because prediction is the most important focus here. The basic concept of the PRESS statistic is to fit the model to a subset of the data, use the resulting estimates to predict observations in the rest of the data, and compute the sum of squares of predicted residuals. A smaller PRESS indicates a model with better predictability. The PRESS residuals given by the MIXED procedure of SAS (SAS Institute, 2008) are marginal PRESS residuals, i.e., they are not conditional on random effects but calculated as $\hat{\varepsilon}_{m(-m)} = y_m - X_m \hat{\beta}_{(-m)}$, with notations explained below. However, to evaluate the predictability of the model reasonably, PRESS residuals should be conditional on random effects. Therefore, a macro was written to compute conditional PRESS residuals as described in the following.

The data from each pig (observed feed intake on each day for DFI data and observed bi-weekly body weight for BW data) were randomly divided into 9 parts. To compute the PRESS, each time one part of the data (y_m, with m = 1 to 9) was set aside and the model was built based on the other 8 parts of the data ($y_{(-m)}$). The estimate of β obtained from data $y_{(-m)}$ will be denoted by $\hat{\beta}_{(-m)}$. Prediction of the part of data y_m based on data $y_{(-m)}$ was $\hat{y}_{m(-m)} = X_m \hat{\beta}_{(-m)} + \hat{C}_{m(-m)} \hat{V}_{(-m)}^{-1}(y_{(-m)} - X_{(-m)}\hat{\beta}_{(-m)})$, where $\hat{C}_{m(-m)}$ is the estimates of the model-based covariance matrix between y_m and $y_{(-m)}$, and $\hat{V}_{(-m)}$ is the estimates of the model-based variance matrix of $y_{(-m)}$. The conditional PRESS residual was computed as $\hat{\varepsilon}_{m(-m)} = y_m - \hat{y}_{m(-m)}$. This procedure was used for prediction of each of the 9

subsets and the PRESS statistics was computed as $PRESS = \sum_{m=1}^{9} \hat{\varepsilon}'_{m(-m)}\hat{\varepsilon}_{m(-m)}$. A similar procedure was used to compute the PRESS statistic for the non-linear mixed models.

The forecast ability of the models to account for missing data at the beginning (90 to 120 days old) and end (181 to 210 days old) of the test period was also evaluated. For this evaluation, data from ages younger than 121 days for one pig were set aside each time and the model was built based on the remaining data for that pig and all data for all other pigs. The conditional PRESS residuals were then calculated for the data from 90 to 120 days of age for that pig by the above method. After repeating this one-pig-at-a-time for all pigs, the PRESS statistics were summarized for all pigs from 90 to 120 days old. Similarly, the forecast ability of the model at the end was evaluated by setting aside data with age older than 180 days for one pig at a time. Because these approaches are computationally intensive, only quadratic and cubic polynomial RR models and the Gompertz non-linear mixed model were evaluated, along with the simple quadratic and linear regression on age on a pig-by-pig basis for DFI and BW.

RESULTS

Model selection on random regression models

Figure 1 shows the PRESS statistics for DFI and BW on gilts and boars from 40 different RR models with different-order polynomials of age as fixed and random effects, and with homogeneous or heterogeneous residual variance by month of age. The PRESS statistics are expressed as a percentage of the PRESS statistics from using simple quadratic and linear regression on age on a pig-by-pig basis. For both boars and gilts, RR models with at least quadratic-order polynomials for random effects for DFI and at least linear order for BW had smaller PRESS statistics than the individual pig models. Heterogeneous residual variance models had PRESS statistics that were similar to those from homogeneous residual variance models for both sexes and both traits.

Table 2 shows that the quadratic and cubic polynomial RR models decreased PRESS statistics dramatically for DFI and BW at the beginning (90 to 120 days old) and end (181 to 210 days old) of the test period compared with the individual pig models. This indicates that RR models have a much better forecast ability than the individual pig models at the beginning and end of the test period. Table 2 also shows that, in most cases, cubic polynomial RR models had smaller PRESS statistics than quadratic polynomial RR models for both DFI and BW. However, for DFI at the end of the test period for boars, quadratic RR models had smaller PRESS than cubic RR models.
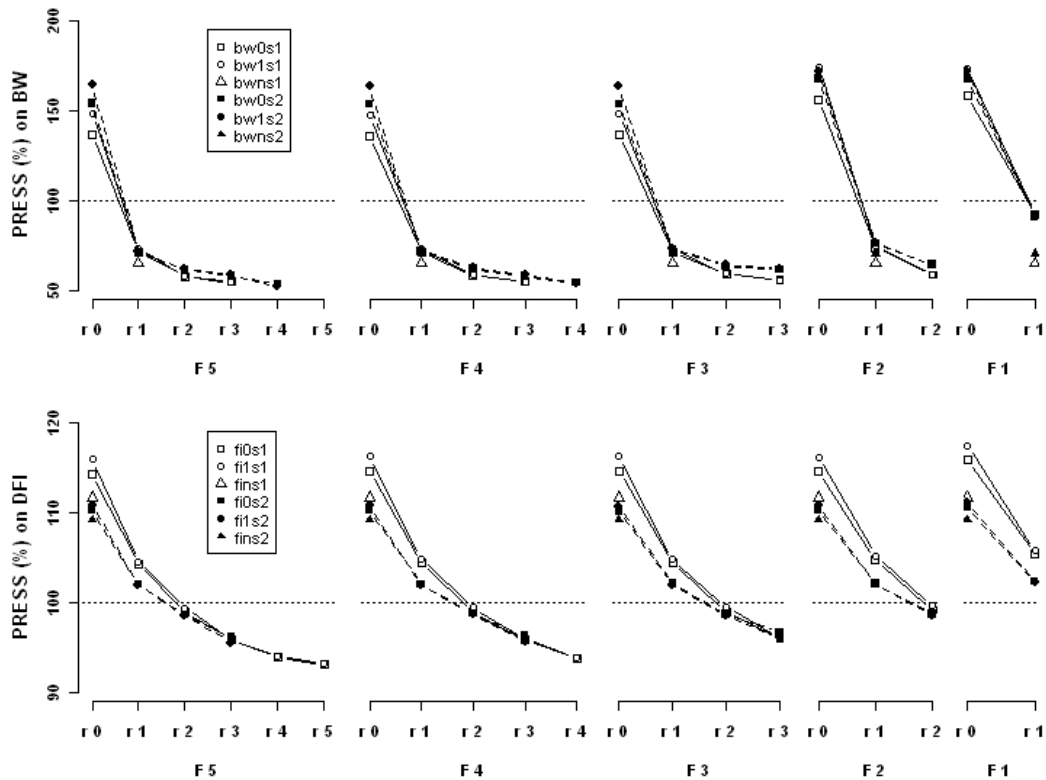
DFI: daily feed intake

BW: body weight

F: the fixed effects of polynomials of age for linear mixed models, e.g., F2 representing intercept, linear and quadratic term of age as the fixed effect

r: the random effects of polynomials of age for linear mixed models, e.g., r2 representing intercept, linear and quadratic term of age as the random effect

bw0s1 and bw0s2: linear mixed models with homogeneous residual variance for BW for boars and gilts, respectively

bw1s1 and bw1s2: linear mixed models with heterogeneous residual variance for BW for boars and gilts, respectively

bwns1 and bwns2: Gompertz non-linear mixed models for BW for boars and gilts, respectively

fi0s1 and fi0s2: linear mixed models with homogeneous residual variance for DFI for boars and gilts, respectively

fi1s1 and fi1s2: linear mixed models with heterogeneous residual variance for DFI for boars and gilts, respectively

fins1 and fins2: Gompertz non-linear mixed models for DFI for boars and gilts, respectively

The PRESS statistics are relative percent of PRESS statistics from the models using simple quadratic and linear regression on age on a pig-by-pig basis for DFI and BW.

Figure 1. PRESS statistics for random regression models.

For all models and both traits, the PRESS statistics decreased with increasing order of polynomials of age as random effects but at a decreasing rate (Figure 1). With quadratic polynomials for both fixed and random effects, residuals checking showed no clear trend of residuals along fitted values but obvious unequal residual variance for DFI. When models with heterogeneous residual variance by month of age were fitted for DFI, the unequal residual variance was much improved. This indicates that the heterogeneous residual variance models behaved better for DFI. Based on these results, quadratic polynomial random regression models were identified to be "best" for both DFI and BW, but with heterogeneous residual variance for DFI

and homogeneous residual variance for BW. These models had the smallest possible order based on both PRESS and residual diagnostics, although they did not have the smallest PRESS among all evaluated polynomials.

Compared to the individual pig models of simple quadratic and linear regression on age, for predictability over the whole test period, the selected quadratic polynomial RR models decreased PRESS by 1% and 2% for DFI for boars and gilts and by 42% and 36% for BW for boars and gilts. For forecast ability at the beginning and end of the test period, the selected quadratic polynomial RR models decreased PRESS from 41% to 87% for DFI and from 26% to 75% for BW compared to the individual pig

Table 2. Predicted residual sum of squares for both daily feed intake and body weight from growth period of 90 to 120 days old and 181 to 210 days old

Sex	Model			PRESS (%) on BW[6]		PRESS (%) on DFI[6]	
	Fixed[1]	Random[2]	Residual[3]	90-120 d	181-210 d	90-120 d	181-210 d
Boars	quadratic	quadratic	homo	25	74	13	59
	cubic	cubic	homo	23	64	13	67
	quadratic	quadratic	hetero	24	78	14	59
	cubic	cubic	hetero	23	65	13	68
	Gompertz with unadjusted pen[4]			30	76	14	47
	Gompertz with adjusted pen[5]			29	76	14	46
Gilts	quadratic	quadratic	homo	40	67	26	29
	cubic	cubic	homo	35	63	23	28
	quadratic	quadratic	hetero	41	70	27	29
	cubic	cubic	hetero	35	65	23	28
	Gompertz with unadjusted pen			44	69	28	29
	Gompertz with adjusted pen			43	68	27	29

[1] Fixed = Fixed effect in the model with quadratic representing quadratic polynomials of age and cubic representing cubic polynomials of age

[2] Random = Random effect in the model with quadratic representing quadratic polynomials of age and cubic representing cubic polynomials of age

[3] Residual = The type of residual variance with homo representing homogeneous residual variance and hetero representing heterogeneous residual variance

[4] Gompertz with unadjusted pen = Gompertz nonlinear mixed model without pen effect pre-adjusted out for daily feed intake and body weight

[5] Gompertz with adjusted pen = Gompertz nonlinear mixed model with pen effect pre-adjusted out for daily feed intake and body weight based on the selected quadratic random regression models

[6] The Predicted residual sum of squares (PRESS) statistics are relative percent of PRESS statistics from the models using simple quadratic and linear regression on age on a pig-by-pig basis for daily feed intake (DFI) and body weight (BW).

models. Thus, compared to the individual pig models, the RR models made prediction of an individual pig's DFI and BW curves more robust and accurate, especially at the beginning and end of the growth period.

Comparison of random regression models and non-linear mixed models

The Gompertz model forecasted better than individual pig models at the beginning (90 to 120 d) and end (181 to 210 d) of the test period for both traits (Table 2). The Gompertz model also predicted better than RR models with the same number of random effects, but not as good as RR models with higher order polynomials (Figure 1). Compared with the selected quadratic polynomial RR models, the Gompertz model had slightly poorer forecast ability at the beginning and end of the test period for BW (Table 2). However, the Gompertz model had comparable forecast ability to the selected RR models for DFI, especially for boars at the end of the test period (Table 2). Table 2 also shows that the forecast abilities of the Gompertz model with or without pre-adjustment for pen effect were very similar. The predictability of the Logistic model was similar to that of the Gompertz model (results not shown).
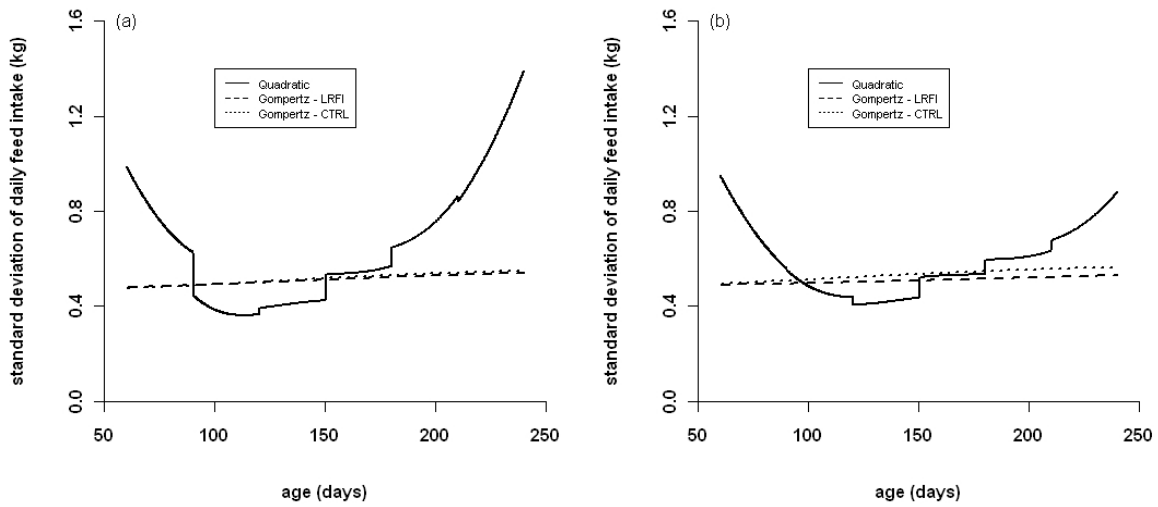
Estimated standard deviations

Estimated phenotypic standard deviations for DFI based on the Gompertz model increased slowly along the growth period for both sexes (Figure 2). Estimated standard deviations for DFI from the quadratic polynomial RR model were close to those for the Gompertz model from ~90 to ~180 days but increased sharply outside that range (Figure 2). Estimated standard deviations for BW from the quadratic polynomial RR and the Gompertz models had a similar increasing trend (Figure 3).

Estimated biological parameters of the Gompertz model

Based on the Gompertz model, LRFI boars had slightly lower mature feed intake (2.93 vs. 2.97 kg) and an earlier inflection point (80 vs. 84 d) for DFI than CTRL boars but differences were not significant (p>0.1) (Table 3). Boars from the LRFI line, however, had a significantly (p = 0.06) greater decay parameter (87 vs. 66 d) for DFI. Boars from the LRFI line had a significantly (p = 0.03) lower mature body weight (263 vs. 296 kg) and a significantly (p = 0.08) earlier inflection point (180 vs. 192 d) for BW than CTRL boars (Table 3). The decay parameter for BW was lower for LRFI boars (127 vs. 134 d) but not significant (p>0.1). However, compared with CTRL gilts, LRFI gilts had a very significantly greater decay parameter for DFI (163 vs. 84 d with p<0.001), a greater mature feed intake (3.31 vs. 2.66 kg with p = 0.02) and a later inflection point (86 vs. 57 d with p = 0.04) (Table 3). Gilts from the LRFI line also had a significantly (p = 0.046) higher mature body weight (296 vs. 266 kg), a significantly (p<0.001) later inflection point (215 vs. 185 d), and a significantly (p<0.001) greater decay

Quadratic: quadratic polynomial random regression model

Gompertz – LRFI: Gompertz non-linear mixed model for lower residual feed intake line

Gompertz – CTRL: Gompertz non-linear mixed model for control line

Figure 2. Estimated standard deviations of daily feed intake for (a) boars and (b) gilts.
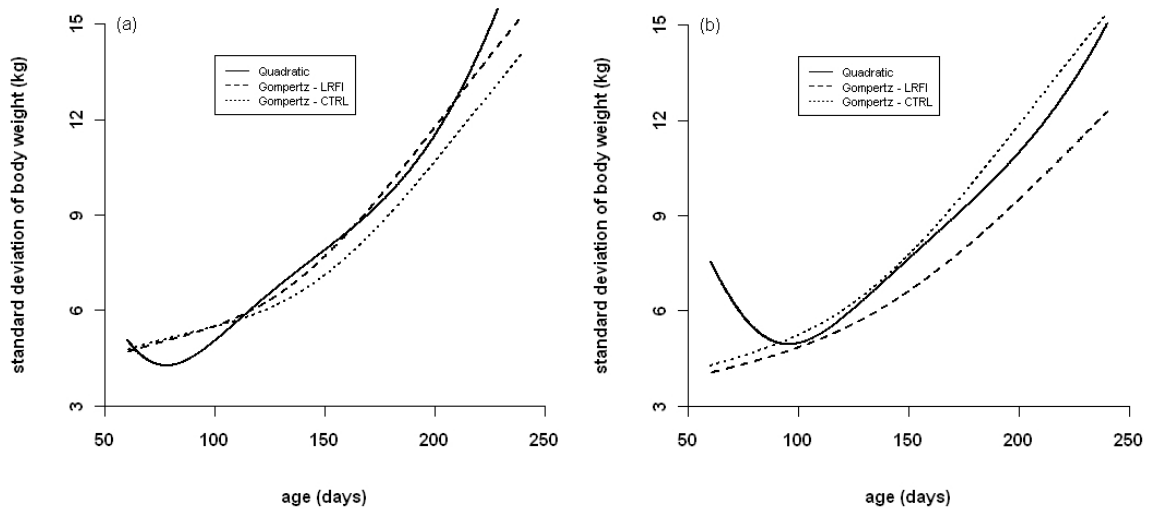
Quadratic: quadratic polynomial random regression model

Gompertz-LRFI: Gompertz non-linear mixed model for lower residual feed intake line

Gompertz-CTRL: Gompertz non-linear mixed model for control line

Figure 3. Estimated standard deviations of body weight for (a) boars and (b) gilts.

Table 3. Estimated parameters in the Gompertz non-linear mixed model for daily feed intake and body weight

Trait	Sex	α_1 (kg)[3]		α_2 (days)[4]		α_3 (days)[5]	
		LRFI[6]	CTRL[7]	LRFI	CTRL	LRFI	CTRL
DFI[1]	Boars	2.93 ± 0.17^{NS}	2.97 ± 0.09^{NS}	80 ± 4^{NS}	84 ± 1^{NS}	87 ± 11^a	66 ± 5^a
	Gilts	$3.31\pm0.29^*$	$2.66\pm0.16^*$	$86\pm13^*$	$57\pm3^*$	$163\pm23^{***}$	$84\pm15^{***}$
BW[2]	Boars	$263\pm11^*$	$296\pm12^*$	180 ± 5^a	192 ± 5^a	127 ± 4^{NS}	134 ± 4^{NS}
	Gilts	$296\pm14^*$	$266\pm10^*$	$215\pm7^{***}$	$185\pm5^{***}$	$168\pm6^{***}$	$136\pm4^{***}$

[1] DFI = Daily feed intake. [2] BW = Body weight. [3] α_1 = Mature body weight or mature daily feed intake.
[4] α_2 = Inflection point for both DFI and BW. [5] α_3 = Decay parameter for both DFI and BW.
[6] LRFI = The line selected for lower residual feed intake. [7] CTRL = The randomly selected control line.
*** $p<0.001$; ** $p<0.01$; * $p<0.05$; [a] $p<0.10$; [NS] $p>0.10$.

parameter (168 vs. 136 d) for BW than CTRL gilts. Knap (2000) summarized previous estimates of mature body weights and the associated Gompertz growth rate parameters (equal to 1/decay parameter in this study) for growing pigs of eight genotypes. Estimates obtained from the current study are within the range summarized in Figure 3 of Knap (2000), which were ~180 to ~320 kg for mature body weight and ~0.005 to ~0.016 d⁻¹ for the Gompertz growth rate parameter (corresponding to ~200 to ~63 days for the decay parameter in this study).

Estimated population curves

Population curves for DFI (Figure 4) and BW (Figure 5) were based on estimated coefficients from the quadratic polynomial RR model and the Gompertz model. Selection for reduced RFI has led to a lower population curve for DFI for the LRFI than the CTRL line (Figure 4). Line differences (CTRL-LRFI) for DFI were small at the

beginning (~90 d) and became larger later in the growing period. Population curves for DFI for boars from the quadratic polynomial RR were similar to those from the Gompertz model, except that curves from the RR model bended faster in the later parts of the growth period (Figure 4a). Population curves for DFI for gilts from the quadratic polynomial RR were higher than those from the Gompertz model (Figure 4b). Population curves for BW from the quadratic polynomial RR and the Gompertz model were similar (Figure 5). Selection for reduced RFI tended to lead to lower body weight for the LRFI than the CTRL line, especially during the later stages of the growth period.

DISCUSSION

Random regression models

Polynomial RR models resulted in more robust and accurate predictions of an individual pig's DFI and BW
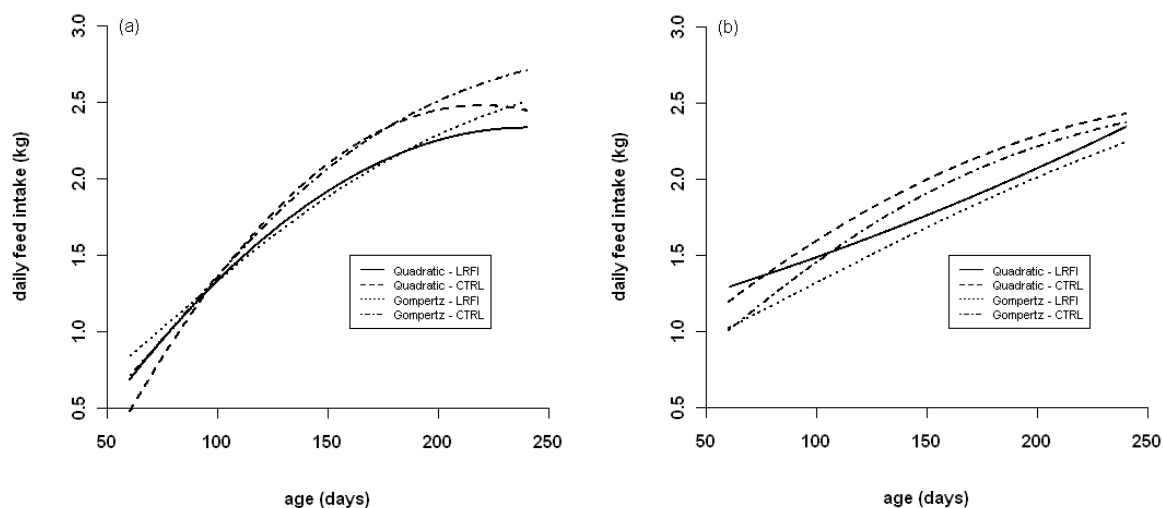
Quadratic – LRFI: quadratic polynomial random regression model for lower residual feed intake line

Quadratic – CTRL: quadratic polynomial random regression model for control line

Gompertz – LRFI: Gompertz non-linear mixed model for lower residual feed intake line

Gompertz – CTRL: Gompertz non-linear mixed model for control line

Figure 4. Estimated population curves of daily feed intake for (a) boars and (b) gilts.

Quadratic – LRFI: quadratic polynomial random regression model for lower residual feed intake line
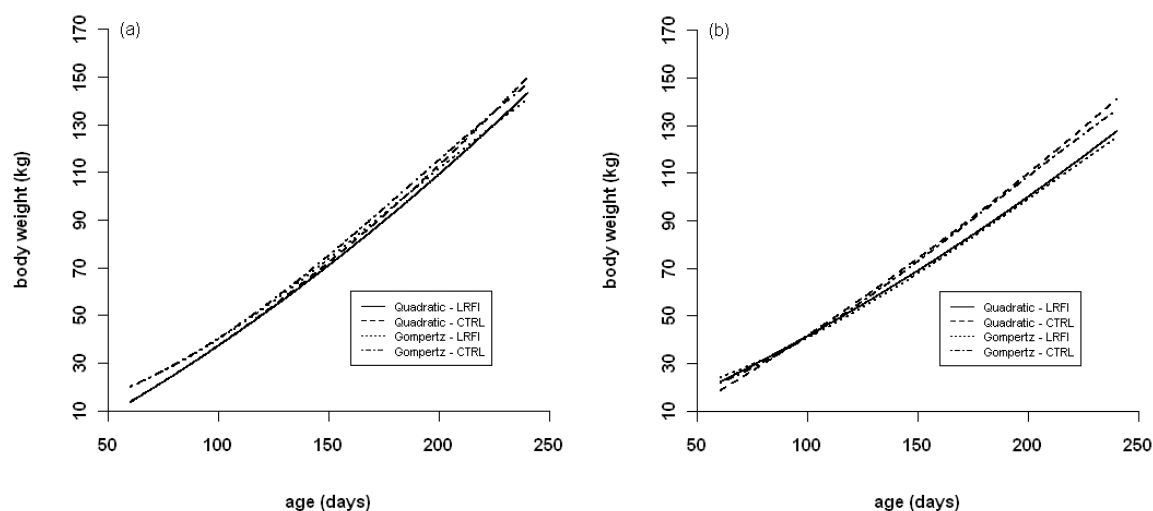
Quadratic – CTRL: quadratic polynomial random regression model for control line

Gompertz – LRFI: Gompertz non-linear mixed model for lower residual feed intake line

Gompertz – CTRL: Gompertz non-linear mixed model for control line

Figure 5. Estimated population curves of body weight for (a) boars and (b) gilts.

curves than individual pig models. Random regression models are a compromise between estimates based only on individual pig's data and an overall estimate across all pigs. If a pig has outlier DFI and BW data, RR models will pull predictions toward the population curve. In addition, RR models allow estimation of population curves for different lines.

Care must be taken when using polynomial RR for data extrapolation because estimated variances for DFI from the RR model became erratic outside the range of the majority of data points (Figure 2). In addition, Lindsey (2001) warned "Care should be taken not to use polynomials of too high an order, usually not more than quadratic, because otherwise the model will be inherently unstable in replications of the data". Typically, the higher the order of polynomials, the more dangerous data extrapolation becomes.

Based on decreases in PRESS, gains in accuracy from RR models were much greater for BW than DFI trait. Compared to the individual pig models, the selected quadratic polynomial RR models only decreased PRESS by 1% and 2% for DFI for boars and gilts but decreased PRESS by 42% and 36% for BW for boars and gilts. One possible reason for the smaller gain in PRESS for DFI than BW is that DFI data are much noisier than BW data. The signal-to-noise ratio, i.e. the ratio of the mean to the standard deviation of a measurement, was ~3 for DFI and ~8 for BW on average over the test period. Thus, RR models may improve predictive accuracy more for data with a higher signal-to-noise ratio. The other possible reason was

that BW is a cumulative trait but DFI is not. Typically, longitudinal measurements for a cumulative trait are more inter-correlated than for a non-cumulative trait. This stronger inter-correlation can be used by RR models to improve data prediction. Another possible reason is fewer longitudinal measurements per pig for BW than for DFI (Table 1). The RR models, which use all pigs' data simultaneously, may gain more accuracy from the data with less information for each pig.

In this study, pen was included as a fixed effect, whereas in previous work (Cai et al., 2008), we fitted pen as a random effect. Cai et al. (2008) implemented a two-stage analysis method (Diggle et al., 2002) for longitudinal measurements of DFI and BW in the previous generations of this experiment. In the first stage, simple quadratic and linear regressions of DFI and BW were fitted for each pig to get a single summary of ADFI and ADG for that pig. Then, these summary variables were analyzed in the second stage. Pen was treated as the random effect when analyzing the summary variables of ADFI and ADG in the second stage because investigation of the variation of pens or feeding stations was one of interest in that study. The RR models of this study can be viewed as an extension of simple regression models of DFI and BW for individual pigs in the first stage. This hierarchical setting of RR models allows each pig within a pen to have its own regression coefficients, which are randomly deviated from their line (LRFI and CTRL) means. In this setting, it makes more sense to fit pen as the fixed block effect to account for systematic difference between pens and feeding stations.

Non-linear mixed models

Sandland and McGilchrist (1979) mentioned that polynomials may provide adequate descriptions of the observed data, but they provide little understanding of the biological data-generating mechanisms. In this study, it is hard to interpret the biological meaning of the regression coefficients from the RR models. In contrast, non-linear models could represent biological data-generating mechanisms and their parameters usually have biological interpretations. As a result, non-linear models are more suitable for data extrapolation. In addition, non-linear models usually need fewer parameters than corresponding linear models for an equal fit to the data (Lindsey, 2001). In this study, the Gompertz non-linear model predicted better than RR models with the same number of random effects.

However, it is difficult for non-linear mixed models to handle complex experimental designs. In this study, both DFI and BW were pre-adjusted for pen effects before analyses by Gompertz and Logistic models to avoid convergence problems. Second, it is difficult to optimize non-linear models with multiple random effects. For example, when Gompertz and Logistic models were fitted to BW with three random effects or to DFI with two random effects, convergence problems occurred. This is the main reason why Andersen and Pedersen (1996) chose simpler linear models instead of non-linear models such as Gompertz and Logistic models.

Generally speaking, if the purpose of the model is data interpolation, and not about understanding the biological data-generating mechanism, polynomial RR models should be used because they are computationally simpler, more flexible, and easier to optimize than non-linear models. However, if the purposes of the model are to explore the biological data-generating mechanism and to do data extrapolation, the non-linear model would be a better option.

Pigs, as food animals, are usually slaughtered below the attainment of half of mature size (Kyriazakis and Whittemore, 2006). Correspondingly, pig breeding programs rarely collected growth performance and feed intake data through the whole growth period until maturity. In this study, pigs were measured repeatedly until they reached about 115 kg BW, which is far below maturity. Many pigs may not even arrive at the decelerating growth period. As a result, the three parameters of the Gompertz model were estimated with limited precision. This may also be one of possible reasons that high-order polynomial RR models predicted better than the Gompertz model.

In this study, we also found that differences between the LRFI and CTRL lines in estimates of coefficients from the Gompertz model for DFI and BW were not consistent between boars and gilts. In addition to data from boars and gilts being obtained at different times and seasons, one possible reason for this inconsistency may be that the experimental protocol for off-testing differed between boars and gilts. Gilts were off-tested in three groups but boars were off-tested on an individual basis, which induced a shorter test period for most gilts. This may lead to less accuracy for estimation of parameters of the Gompertz model for gilts. For DFI, standard errors of estimated coefficients from the Gompertz model for gilts were larger than for boars.

Effect of selection for reduced RFI on DFI and BW curves

In the lines used in this study, selection was based on estimated breeding value for RFI, with component traits of feed intake and growth averaged over the test period. Cai et al. (2008) reported that after four generations of selection, boars of the LRFI line consumed 202 g/d less feed and gained 39 g/d less weight than the CTRL line on average over the test period. This study showed that after five generations of selection for reduced RFI, the LRFI line had a lower population curve for DFI and BW than the CTRL line, especially towards the end of the growth period (Figures 4 and 5). This demonstrated that the difference in feed intake and growth between the LRFI and CTRL lines mostly comes from the late growth period. Bermejo et al. (2003) found that selection on average feed intake over the whole test period led to an increase of feed intake mainly in the second half of the test, which is similar with this study. The lower feed intake and body weight curves because of selection for reduced RFI also indicates that it is possible for the pig breeding industry to optimize growth and feed intake curves by selection.

REFERENCES

Andersen, S. and B. Pedersen. 1996. Growth and food intake curves for group-housed gilts and castrated male pigs. Anim. Sci. 63:457-464.

Cai, W., D. S. Casey and J. C. M. Dekkers. 2008. Selection response and genetic parameters for residual feed intake in yorkshire swine. J. Anim Sci. 86:287-298.

Casella, G. and R. L. Berger. 2002. Statistical inference, second edition. Duxbury Press, Pacific Grove, California.

Casey, D. S. 2003. The use of electronic feeders in genetic improvement programs for swine, PhD Diss. Iowa State Univ., Ames.

Casey, D. S., H. S. Stern and J. C. M. Dekkers. 2005. Identification of errors and factors associated with errors in data from electronic swine feeders. J. Anim. Sci. 83:969-982.

Diggle, P. J., P. J. Heagerty, K. Y. Liang and S. L. Zeger. 2002. Analysis of longitudinal data (second ed). 2nd ed. Oxford University Press, New York.

Eissen, J. J., A. G. de Haan and E. Kanis. 1999. Effect of missing data on the estimate of average daily feed intake of growing pigs. J. Anim. Sci. 77:1372-1378.

Eissen, J. J., E. Kanis and J. W. M. Merks. 1998. Algorithms for

identifying errors in individual feed intake data of growing pigs in group-housing. Appl. Eng. Agric. 14:667-673.

Knap, P. W. 2000. Time trends of gompertz growth parameters in "Meat-type" pigs. Anim. Sci. 70: 39-49.

Kyriazakis, I. and C. T. Whittemore. 2006. Whittemore's science and practice of pig production. 3rd ed. Blackwell Publishing Ltd, Oxford, UK.

Lindsey, J. K. 2001. Families of nonlinear regression functions. Page 37-47 in nonlinear models in medical statistics (Ed. J. K. Lindsey). Oxford University Press, New York, USA.

Lorenzo Bermejo, J., R. Roehe, G. Rave and E. Kalm. 2003. Comparison of linear and nonlinear functions and covariance structures to estimate feed intake pattern in growing pigs. Livest. Prod. Sci. 82:15-26.

Ratkowsky, D. A. 1990. Handbook of nonlinear regression models. Marcel Dekker, Inc., New York.

Sandland, R. L. and C. A. McGilchrist. 1979. Stochastic growth curve analysis. Biometrics 35:255-271.

SAS Institute. 2008. Sas online documentation, version 9.1.3. SAS Institute Inc., Cary, NC.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35-45.

Schaeffer, L. R. and J. C. M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. In: Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Ont., Canada, 18:443-446.

Schulze, V., R. Roehe, H. Looft and E. Kalm. 2001. Effects of continuous and periodic feeding by electronic feeders on accuracy of measuring feed intake information and their genetic association with growth performances. J. Anim. Breed. Genet. 118:403-416.

Von Felde, A., R. Roehe, H. Looft and E. Kalm. 1996. Genetic association between feed intake and feed intake behaviour at different stages of growth of group-housed boars. Livest. Prod. Sci. 47:11-22.

Whittemore, C. T., J. B. Tullis and G. C. Emmans. 1988. Protein growth in pigs. Anim. Prod. 46:437-445.