ELSEVIER

Computational Statistics & Data Analysis 33 (2000) 79–100

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Diagnostic tools for random effects in the repeated measures growth curve model

P.J. Lindsey *, J.K. Lindsey

Biostatistics, Limburgs Universitair Centrum, Diepenbeek, Belgium

Received 1 April 1999; received in revised form 1 June 1999

Abstract

Growth curve models assuming a normal distribution are often used in repeated measurements applications because of the wide availability of software. In many standard situations, a polynomial in time is fitted to describe the mean profiles under different treatments. The dependence among responses from the same individuals is generally handled by a random effects model, although an auto-regressive structure can often be more appropriate. We consider both, in the context of missing observations. We present diagnostics for two major problems: (1) the forms of the mixing distribution in random effects models, and their influence on inferences about treatment effects, and (2) the randomness of missing observations. To demonstrate the utility of our techniques, we reanalyze data on percentage protein content in milk, often erroneously analyzed as illustrating a dropout phenomenon.

Keywords: AIC; Auto-regression; Continuous AR; Fixed effects; Growth curve; Individual profile plots; Logistic distribution; Missing values; Random effects; Recursive residuals; Repeated measurements; Stable distribution

1. Introduction

The original growth curve model for repeated measurements over time, introduced by Elston and Grizzle (1962), and generalized by Potthoff and Roy (1964), has come to be widely used when the responses can be assumed to be approximately normally

* Corresponding author.
E-mail address: plindsey@luc.ac.be (P.J. Lindsey)

0167-9473/00/$ - see front matter
PII: S 0167-9473(99)00049-3

distributed. It was later popularized by Laird and Ware (1982). This model uses polynomials in time to describe mean profiles, with random coefficients to generate a correlation structure among the repeated observations on each individual. However, as Elston (1964) pointed out, such a covariance matrix depends crucially on the time origin used, making the model difficult to interpret and often unsuitable. For further discussion of the problems with such random coefficients models, see Lindsey (1993, pp. 85–97).

A more appropriate approach to modelling the dependence of responses over time is to introduce some form of auto-regressive structure, leaving any random effects to handle inter-individual heterogeneity. Such models have been discussed by many authors; for a survey, see Lindsey (1993, Chapter 4). One of the most flexible techniques for fitting such models, combining random effects and auto-regression, is the Kalman filter. This allows observations to be unequally spaced in time and can, thus, handle randomly missing values; see, for example, Jones and Ackerson (1990), Jones and Boadi-Boateng (1991), and Jones (1993). We shall use Jones' software in the analyses to follow.

One major drawback of this linear growth curve model is that the mean response is generally taken to vary as a polynomial over time. In most situations, this will be biologically unreasonable, some nonlinear function adapted to the specific situation being more appropriate. Few authors have attempted to accommodate this situation with both random effects and auto-regressive components; see, however, Heitjan (1991a, b) and Lambert (1996). We shall not consider diagnostics to detect this type of problem, although mis-specification in this part of the model will influence conclusions about the suitability of other components of a model.

Many aspects of a repeated measurements model need to be checked when analyzing such data. We can only cover a few here. We shall be particularly interested in the appropriateness of the random effects distribution(s) in describing the heterogeneity found in the data, including ways in which specification of the time variable influences this. We shall also look at what information may be available about whether or not missing values can be assumed to be random, without making any attempt to model them. For the first problem, we shall look at what a fixed effects model can tell us about random effects. For the second, we shall consider the results of different approaches to fitting the auto-regression of the model, as well as looking at logistic regression. We shall study these in the context of residual analysis, and individual profiles.

2. The milk data

To illustrate our procedures, we shall use data that have been discussed several times in the statistical literature (Verbyla and Cullis, 1990; Diggle, 1990; Diggle and Kenward, 1994; Diggle et al., 1994; Little, 1995). They concern an experiment on the effect of three different feeding strategies on the protein content of milk produced by cows over time. The percent protein level was measured on 79 cows weekly during

19 consecutive weeks. The cows were randomly divided into three diet groups kept in separate paddocks: the first 25 cows were assigned to a barley diet, the next 27 to a mixed barley and lupins diet, and the last 27 cows to a lupins diet.

In a number of the analyses, the data were erroneously taken as if all of the cows entered the experiment at the same time, with missing values at the end of a series indicating that the cow dropped out before the experiment terminated. However, the cows actually entered when they calved and the experiment was ended at the same time for all, so that the 'missing values' appear at the beginning (Cullis in discussion of Diggle and Kenward, 1994; Diggle et al., 1994, pp. 5, 100). This is exactly equivalent to the design of a standard survival study with staggered entry and Type I censoring. The only difference might be that cows may have been randomized at the beginning of the experiment instead of at 'entry', that is at calving. Because the special diets were apparently not started then, treatment could not affect the time to calving.

Thus, 'missing values' at the end of the shorter series, due to cows starting late, are ignorable; with this design, it is impossible that 'there is a strong dependence of the dropout probability on the most recently observed measurement' (Diggle et al., 1994, p. 215), although some spurious correlation might be detectable. On the other hand, the response value at any time point can depend on the previous history of the cow, including time of calving (that is, a factor variable indicating the cohort) and time since calving.

There are also a total of 11 missing values within the series for eight cows (three are from one cow, with two values missing in consecutive time periods); we shall look closely at these. Thus, there are a total of 1337 observations available on the 79 cows.

For these data, two time origins are possible: the beginning of the experiment (that is, chronological time) and the moment when a cow begins producing milk after calving (biological time). Separate models are considered for these two aligning methods, as a common model is not possible due to collinearity. Dependence of protein level on the first would indicate the action of external factors common to all cows, including changes in feeding in each paddock over time. We know that the trial was terminated when feed availability declined in the paddocks in which the cows were grazing (Cullis, in discussion of Diggle et al., 1994). In contrast, dependence on biological time would correspond to internal factors specific to a given cow.

An additional factor is that the first three weeks after calving constitute a settling in period. Some authors have ignored this part of the data, either by excluding it or by simply leaving it as an integral part of the whole series. We shall directly model this aspect of each series.

In order to carry out a reasonable analysis of these data, we would also need to know a number of other things that are not available. These include why two fewer cows were assigned to the first group (problems with calving?), the diet of cows before the experiment began, the treatment of cows not entering at the beginning (that is, all cohorts but the first) while they waited, and the reasons why the 11 missing responses were not recorded.

3. Models

In the simplest cases, the growth model takes the $N \times T$ matrix of response values, $\boldsymbol{Y}$ (where N is the number of individuals, and T is the number of time points), to have mean

$$E[\boldsymbol{Y}] = \boldsymbol{XBZ},$$

where $\boldsymbol{X}$ is the $N \times C$ inter-subject design matrix (describing the C diets and cohorts in our example) for the N individuals, while $\boldsymbol{B}$ is a $C \times P$ location parameter matrix (where P is the number of time-varying covariates) and $\boldsymbol{Z}$ is a $P \times T$ matrix of covariates changing with the responses on a unit, most often simply a $P-1$ degree polynomial over the T points in time. This is assumed to describe the mean of a multivariate normal distribution with variance-covariance matrix, $\boldsymbol{\Sigma}$. One important question with which we shall be concerned is how to check for appropriate structuring of this matrix.

Classical first-order auto-regression, or AR(1), based on a multivariate normal distribution, has a covariance matrix

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-3} & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-2} & \rho^{T-3} & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \dots & \rho & 1 \end{pmatrix}$$

for T time periods, where σ^2 is the variance and ρ the auto-correlation.

Often, time-varying covariates, say z, will be available (as will time-constant covariates, but the presence of these does not influence the following discussion so that they will be omitted for simplicity of notation). The regression equation describing how the mean depends on such explanatory variables can be developed in two different ways (Lindsey, 1993, pp. 99–112). We may start with a simple model for the conditional mean, $\mu_{t|t-1}$:

$$\mu_{t|t-1} = \rho y_{t-1} + \sum_{i=0}^{t} \beta_i z_{it}, \tag{1}$$

where t indexes equally spaced points in time, and i the individual. The resulting marginal mean of the multivariate normal distribution, μ_t, is relatively complex:

$$\mu_t = \sum_{k=0}^{t} \rho^{t-k} \sum_{i=0}^{t} \beta_i z_{ik}. \tag{2}$$

This can be called a state dependence model and can be fitted with Eq. (1) by any standard regression software if the first response is ignored, except conditioning on it as a fixed value. (We shall use GLIM4; Francis et al., 1993.) Hence, this is sometimes known, in the time series literature, as constructing a 'conditional' likelihood.

A second approach is to consider a simple marginal model, still with the same multivariate covariance matrix, such that

$$\mu_t = \sum_{i=0}^{t} \beta_i z_{it}. \tag{3}$$

Here, however, the conditional mean is given by

$$\mu_{t|t-1} = \rho \left(y_{t-1} - \sum_{i=0}^{t-1} \beta_i z_{i,t-1} \right) + \sum_{i=0}^{t} \beta_i z_{it}. \tag{4}$$

This can be called the serial correlation model. It involves a nonlinear regression that requires special software for estimation. (We shall use the CARMA software of Jones, 1993, that works in continuous time so that observations could be unequally spaced, with varying times among individuals. SAS Proc Mixed could also have been used.) As can be seen from the formula, it is modelling dependence of the response on the residuals, that are thus assumed to be auto-correlated. Here, the first observation is generally used, with the assumption that it has the appropriate (stationary) marginal distribution, that is, normal with the mean just given and variance, $\sigma^2/(1-\rho^2)$. Thus, this is sometimes called the complete likelihood approach.

Although the second model, with auto-correlated residuals, is widely used in time series analysis, it may be argued (Lindsey, 1992, pp. 119–132; Hendry, 1995) that the existence of auto-correlation among the residuals indicates inappropriate conditioning on explanatory variables. Thus, where possible, the first model, with the simpler conditional regression and uncorrelated residuals, will usually be preferable because it indicates adequate allowance for the previous history of the subjects. If the second model fits better, this is an indication that some pertinent explanatory variables are missing or incorrectly used in the model. The obvious extension, unifying both models, is to use

$$\mu_{t|t-1} = \rho y_{t-1} + \sum_{i=0}^{t-1} \gamma_i z_{i,t-1} + \sum_{i=0}^{t} \beta_i z_{it} \tag{5}$$

as Hendry suggests, instead of imposing the constraint that $\gamma_i = \rho\beta_i$. Again, this model is linear and can be fitted with standard regression software.

If there are no time-varying covariates (z_{it}), the three models collapse to be the same (except perhaps for the way in which the first observation is handled). In the case of our growth curve model, the only time-varying covariates are the terms in time itself, in form of a polynomial. Then, it is easy to show that, again, all three models collapse to be identical: $z_{it} = t^i$ so that the γ_i is the coefficient of $(t-1)^i$ and β_i that of t^i and these parameters are not separately identifiable. In other words, in this particular case, the γ_i and β_i coefficients of Eq. (5) become indistinguishable, as do β_i and $\rho\beta_i$ in Eq. (4), both resulting in Eq. (1). In the same way, the β_i coefficient of t^i in Eq. (3) is not distinguishable from the coefficient of t^i in Eq. (2), the latter being a function of ρ and $\boldsymbol{\beta}$. We shall use this equivalence among the models, and the fitting procedures, to provide extra flexibility in modelling, including the study of missing values.

The second aspect of modelling the covariance matrix concerns inter-subject heterogeneity. This is usually handled by introducing random coefficients, which, in the most general case, can be written as

$$E[\boldsymbol{Y}|\boldsymbol{\Lambda}] = \boldsymbol{XBZ} + \boldsymbol{\Lambda V},$$

where $\boldsymbol{\Lambda}$ is the random effect

$$\Lambda \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{I} \otimes \boldsymbol{\Delta})$$

and $\boldsymbol{Z}$ and $\boldsymbol{V}$ are polynomials in time. In the complete model, this will be combined with the dependence structure of auto-regression discussed above. In what follows, we shall be especially concerned with the commonly used assumption of normality of the random parameters. The multivariate normal mixing distribution for Λ attempts to describe how a coefficient in the model varies among individuals included in the study. In this sense, the need for random coefficients is an admission that important explanatory variables are missing for the question at hand: incomplete information is available about differences among individuals upon which one could condition using covariates. Thus, introducing the random coefficients is a mathematical technique for inducing a multivariate normal distribution with some specific covariance matrix to allow for the unexplained variation.

As in Lindsey and Jones (1997) for Poisson repeated measurements, we shall look at fixed effects models as a diagnostic tool. Such models can be thought of as non-parametric estimations of the mixing distribution whose form can then be studied. As well, fixed effects are often informative in their own right, allowing easy detection of individuals affected in extreme ways by the treatments, something that is often important, for example in drug testing. By studying the individuals detected in this way, one may be able to isolate what variables are missing in the modelling process. Because we shall be using exact small sample direct likelihood methods for inference, asymptotic questions of consistency are irrelevant (although they are for other inference methods such as standard errors).

The inference criterion that we shall use for comparing the models under consideration will be their ability to predict the observed data, that is how probable they make the data. In other words, models will be compared directly through their minimized $-\log$ likelihood. When the numbers of parameters in models differ, they will be penalized by adding the number of estimated parameters, a form of the Akaike information criterion (AIC, see Akaike, 1973; Lindsey and Jones, 1998). Smaller values indicate more preferable models. This criterion allows direct comparisons among models, that are not required to be nested.

4. Preliminary analysis

These data have already been analyzed several times in the literature, so that preliminary analysis using a series of plots, especially individual profiles, need not be presented here. That for biological time origin was presented by Verbyla and Cullis (1990), Diggle (1990, p. 159), and Diggle et al. (1994, p. 54). We now proceed

Table 1
AICs for a number of models fitted to the milk data using continuous AR

	Biological	Chronological	Parameters
Intercept	423.41		2
Linear	418.4	419.2	3
Quadratic	385.4	420.1	4
Cubic	363.6	378.0	5
+ Treatment	310.1	326.0	7
+ AR(1)	12.7	23.6	8
+ Measurement error	−15.0	−11.6	9
+ Random intercept	−14.0	−10.6	10

in fitting what were called complete likelihood models above, using continuous AR (Jones, 1993).

However, we first look at the calving time as a response variable in order to check randomness of cohorts among treatments. An unequal repartition of the calving times might have resulted if randomization occurred at the beginning of the trial and not at calving. The frequencies of the six different calving times can be cross-classified by treatment; the resulting contingency table has a deviance for independence of 1.64 with eight degrees of freedom indicating no difference in calving times with treatment.

We can now apply the continuous AR models (using CARMA) to the full data set. Note that, with Kalman filtering in continuous time, when an observation is missing within a series, conditioning occurs upon the most recent observed response, but with an appropriately lower auto-correlation. The AICs for a hierarchically nested series of models for the evolution of protein content in the milk are shown in Table 1.

We see here that auto-regression and measurement error are required but not a random intercept. The biological time origin, located when each cow begins producing milk, provides a superior fit. More complex auto-regressive and random coefficient structures were also tried; the only improvement found was with an AR(3) for the biological time, reducing the AIC from −15.0 to −20.8. There is no evidence of an interaction between treatments and time; in other words, the mean profiles are parallel. When treatment is removed from the AR(1) models (without random intercept), the AICs rise, respectively, to −9.2 (biological) and −7.6 (chronological). Adding a cohort effect, distinguishing calving times by a factor variable, does not improve either model. Let us look in more detail at some results for the best model with biological time. All of the following graphs are standard output from CARMA. The mean profiles for the three treatments are plotted in Fig. 1.

We see that the barley diet yields the highest protein content, followed by the mixed and then the lupin diets. The estimates (standard errors) of the differences of the latter two with the first are, respectively, −0.101 (0.051) and −0.212 (0.051).

The individual profiles for nine selected cows (numbers 12 and 74 have missing values) are shown in Fig. 2.

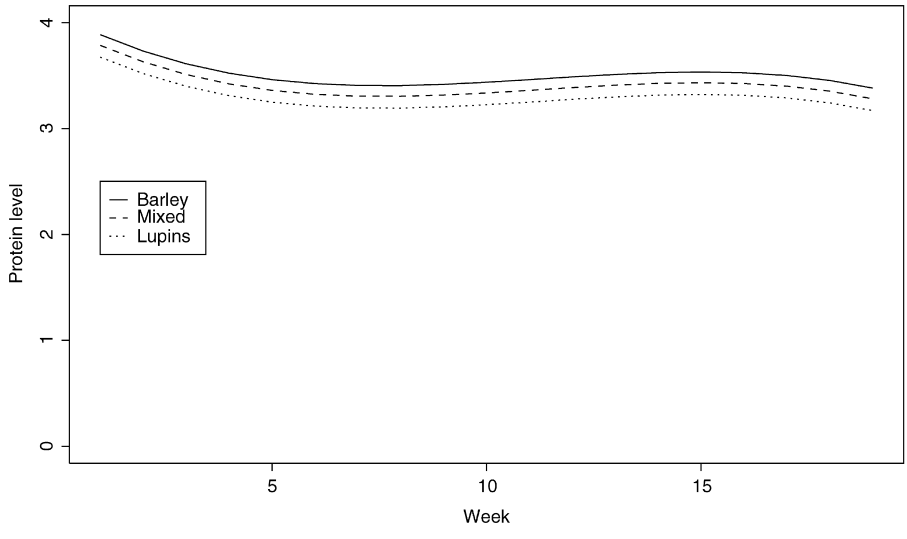

Fig. 1. Mean profiles for the three treatments.

The shorter ones are cows starting later, but aligned to the left because the time origin is biological. Notice the irregularities, as compared to the mean profiles in Fig. 1, due to the auto-regressive effect that takes into account the magnitude of the previous response. The recursive residuals for the same nine cows are presented in Fig. 3.

Except for one somewhat extreme negative residual at the first time point for cow 73, there appear to be no major anomalies. In principle, one would generally be satisfied with this final model. However, we shall now go on to consider some more non-standard diagnostics.

5. Diagnostics

In complex models, such as those for repeated measurements or where dispersion parameters (such as the variance) depend on the covariates, standard linear regression diagnostics such as residual analysis are often of limited use in detecting problems with a model. (For other examples, see Lindsey and Jones, 1997.) The best approach seems to be to try fitting a wider range of models in order to check the assumptions being made.

5.1. Conditional AR models

We now turn to the 'conditional' likelihood model. To do this, we must eliminate the first response of each cow and the response following any missing value in the middle of a sequence (except to condition on them) because we do not have the preceding value for these. We lose a total of 89 observations, the same ones with

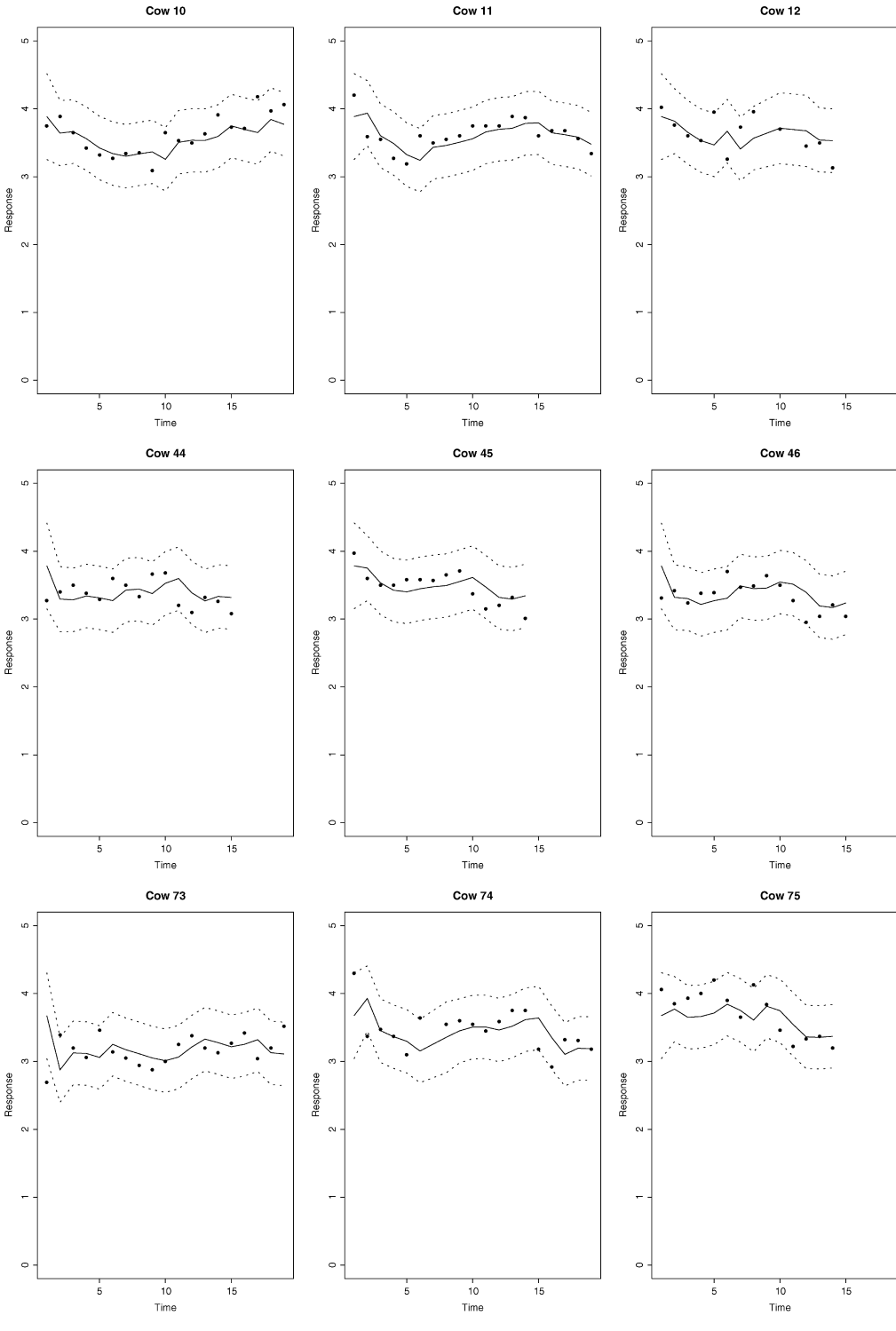

Fig. 2. Individual profiles (solid lines) over biological time or nine selected cows, with the three rows corresponding, respectively, to cows receiving protein, mixed, and lupins. Twice the standard deviations are indicated by broken lines.

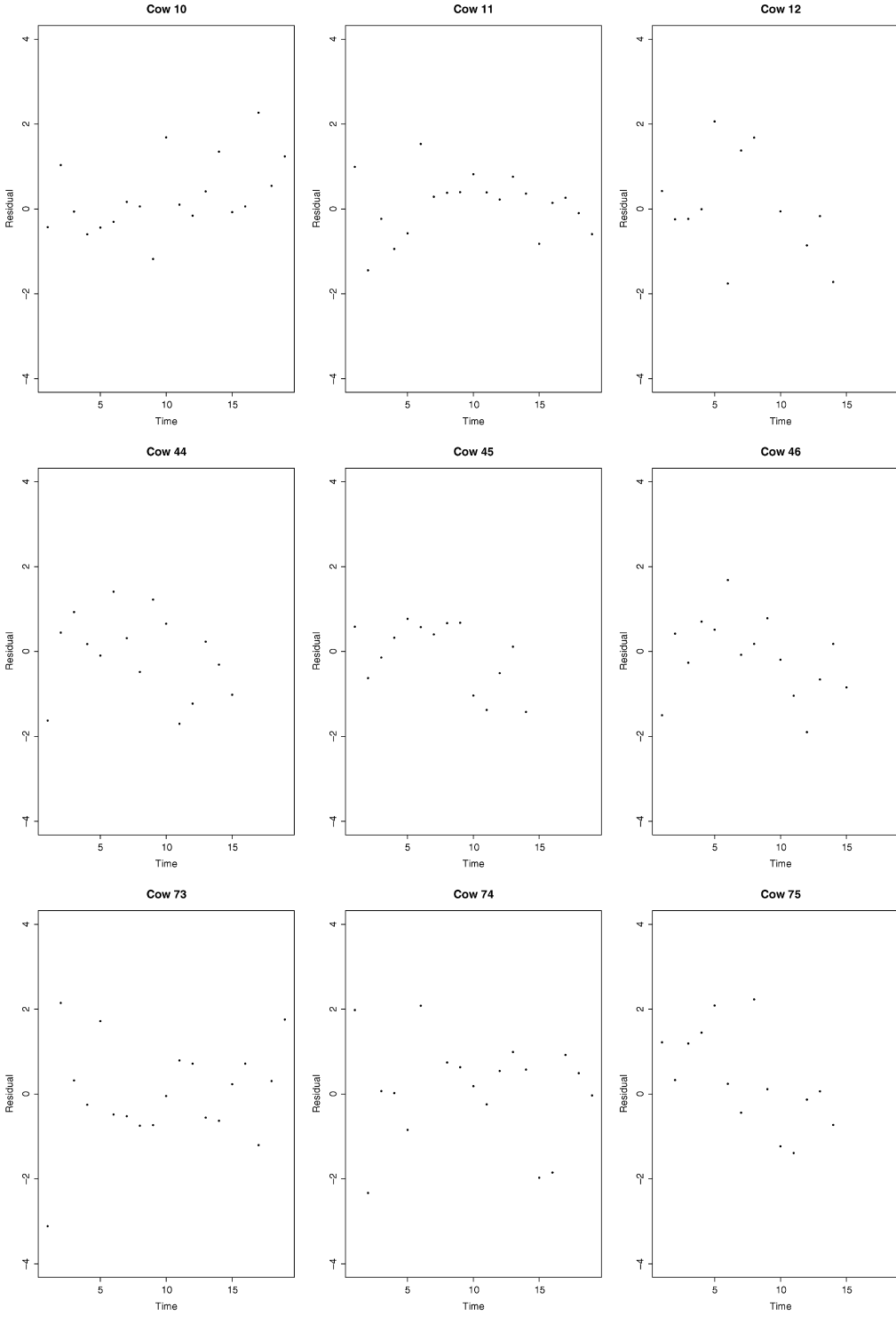

Fig. 3. Recursive residuals for the nine selected cows, with the three rows corresponding, respectively, to cows receiving protein, mixed, and lupins.

Table 2
AICs for the model with a cubic polynomial in time, treatment effect, and an AR(1) (eight parameters) fitted by various methods

	Biological	Chronological
Continuous AR (complete data)	12.7	23.6
Conditional AR (linear regression)	33.3	2.4
Continuous AR (incomplete data)	37.0	18.7

both time origins. Because these responses are fixed at their observed values, we shall count them as additional parameters when penalizing the likelihood in the AIC. In this way, the AICs will be comparable with those previously given for the continuous AR models. Note that this is a rather heavy penalty because of the correlation among the observations.

No software is available to fit measurement error in such conditional models, so we begin from the model above without it, that is, a cubic polynomial in time, the treatment effect, and an AR(1). The AICs for this model are shown in Table 2, fitted as above (first line, from Table 1) and by linear regression (second line).

The only difference between these two models is that the first (continuous AR) takes the first observation for each cow to have a stationary marginal distribution, while the second (linear regression) assumes that it, and any one following a missing value, have fixed values (counting them as additional parameters).

Surprisingly, we find that the model with chronological time origin now fits better, an improvement on either of the equivalent continuous AR models. We can also fit the continuous AR model to the incomplete data. Here, the 89 eliminated responses are not used at all, whereas, with linear regression, they were only used conditionally. (They are still counted as estimated parameters in the penalty for the AIC.) As can be seen in the last line of Table 2, this model fits more poorly, especially with the biological time origin.

These results indicate that, under our model, the eliminated responses may be different than the others, especially with the chronological time origin. The assumption of a stationary distribution at the beginning of each series may not be satisfactory or the 10 observations eliminated after the missing values may be special. With a continuous AR, we can eliminate either the first value for each cow or the response following an intermediate missing value separately. The AICs are given in Table 3.

We see that the first response of each cow, with the chronological time origin, was fitting badly; the model improves when these are removed. This is not surprising if a settling in period is required because this period occurs at various times with respect to when the experiment began. We shall look further at this problem below.

Instead of removing the response following a missing value, we may rather remove that preceding it, also shown in Table 3. We see that this provides a better model than when the response following a missing value is removed, not much worse than the model with the complete data. Thus, the responses immediately before a missing value are suspect. We can investigate this further.

Table 3
AICs for the model with a cubic polynomial in time, treatment effect, and an AR(1) (eight parameters) fitted, using a continuous AR, with various observations removed

	Biological	Chronological
Complete data	12.7	23.6
First and after missing removed	37.0	18.7
First value removed	23.7	5.0
Value after missing removed	25.5	36.8
Value before missing removed	13.4	26.8

We may check the randomness of the responses missing in the middle of the series by constructing a binary missingness indicator and performing logistic regression on the immediately preceding response value (two are consecutive, so that only ten of the 11 can be used). We find that a model with such dependence fits better (AIC, 53.9) than without (AIC, 58.8). Missing values tend to follow high responses, with a logistic regression coefficient (standard error) of 2.996 (0.902). They are, thus, not missing completely at random. Unfortunately, no information is available about why such non-responses occurred so that appropriate models cannot be developed.

We can now check if a cohort effect is necessary in the conditional (linear regression) models. For biological time, the AIC is reduced from 33.3 to 32.3 and, for chronological time, from 2.4 to -4.9 when those cows calving first are distinguished from the others, there being no difference among the other cohorts. This differs from the results for the continuous AR with the complete data above, where no cohort effect was detected. We still do not have a model that fits quite as well as the one with measurement error in the previous section.

5.2. Individual heterogeneity

Let us now examine more closely the individual heterogeneity of the cows. The following models will include a fixed effect for the cows; this is an individual parameter for each animal. The inter-subject design matrix $\boldsymbol{X}$ from the above model will now be of size $N \times N$. When we fit fixed effects models by the linear regression method, we obtain the results shown in Table 4 (two of these fixed effects estimates being aliased with the treatment effects).

Once the quadratic interactions between individual cows and time is introduced into the model, the biological and chronological approaches become identical. The presence of five distinct parameters involving time, including interactions, also allows the model to take into account the five different occasions on which cows were entered into the study in the chronological setting. Note that these models contain a very large number of parameters. But, at the same time, the likelihood is very heavily penalized in the AIC: the penalty is 407 ($=6 + 4 \times 78 + 89$) with 1248 observations (after elimination of 89), much more, for example, than a frequentist likelihood ratio based Chi-squared test at 5%. Nevertheless, the AIC indicates that

Table 4
AICs for a number of fixed effects models, with a cubic polynomial in time, treatment effect, and an AR(1) fitted to the milk data using linear regression. The first column indicates the interaction between individual cows and time included in the model

	Biological	Chronological	Parameters
None	33.3	2.4	8
Intercept	13.6	−43.5	84
+ Linear	−39.7	−82.3	162
+ Quadratic	−128.9		240
+ Cubic	−138.8		318

these models fit very much better than those in Table 1. Perhaps the presence of measurement error in the earlier continuous AR models may have occurred because non-normal mixing distributions were not being detected. We shall look at this in more detail below, but first let us consider one further improvement to the model.

The one variable that we have not yet used is the settling in period. If we allow the intercept and the linear component of time to be different during this period, the AIC is reduced from −138.8 to −147.0 for chronological time origin and to −161.2 for biological time origin. The linear component of the slope of the profile is less negative during these three weeks.

5.3. Distribution of the fixed effects

We can now study more closely the fixed effects estimates for these models. Among other things, this will allow us to see how changing the time origin for all cows, for example by standardizing to set zero time at the average observed time, affects the model, as mentioned in the introduction. (Note that this standardization is a separate issue from that of the appropriate time origin, biological or chronological, for individual cows in this specific data set.) The estimates of these parameters (with the conventional constraint of summation to zero) provide us with information about the heterogeneity among the cows, and hence about possible mixing distributions for a random effects model. One way to study the form of these distributions is to use kernel density estimation. Consider our best model with the biological time origin and settling in effect. The results, when time is standardized to be centred at its average, for the four sets of fixed effects (corresponding to four random coefficients) are shown in Fig. 4.

Let us next look at the fits of several distributions to these estimates, as summarized in the top left panel of Table 5.

For the intercept and the coefficient for linear time, the normal distribution fits best of those tried. For quadratic and cubic time, the logistic distribution (not to be confused with binary logistic regression) yields a better fit. The densities for these two distributions are also plotted in Fig. 4. We see that, except for the logistic distribution applied to the fixed effects interaction with the cubic in time, the estimated curves do not follow the kernel density very well.

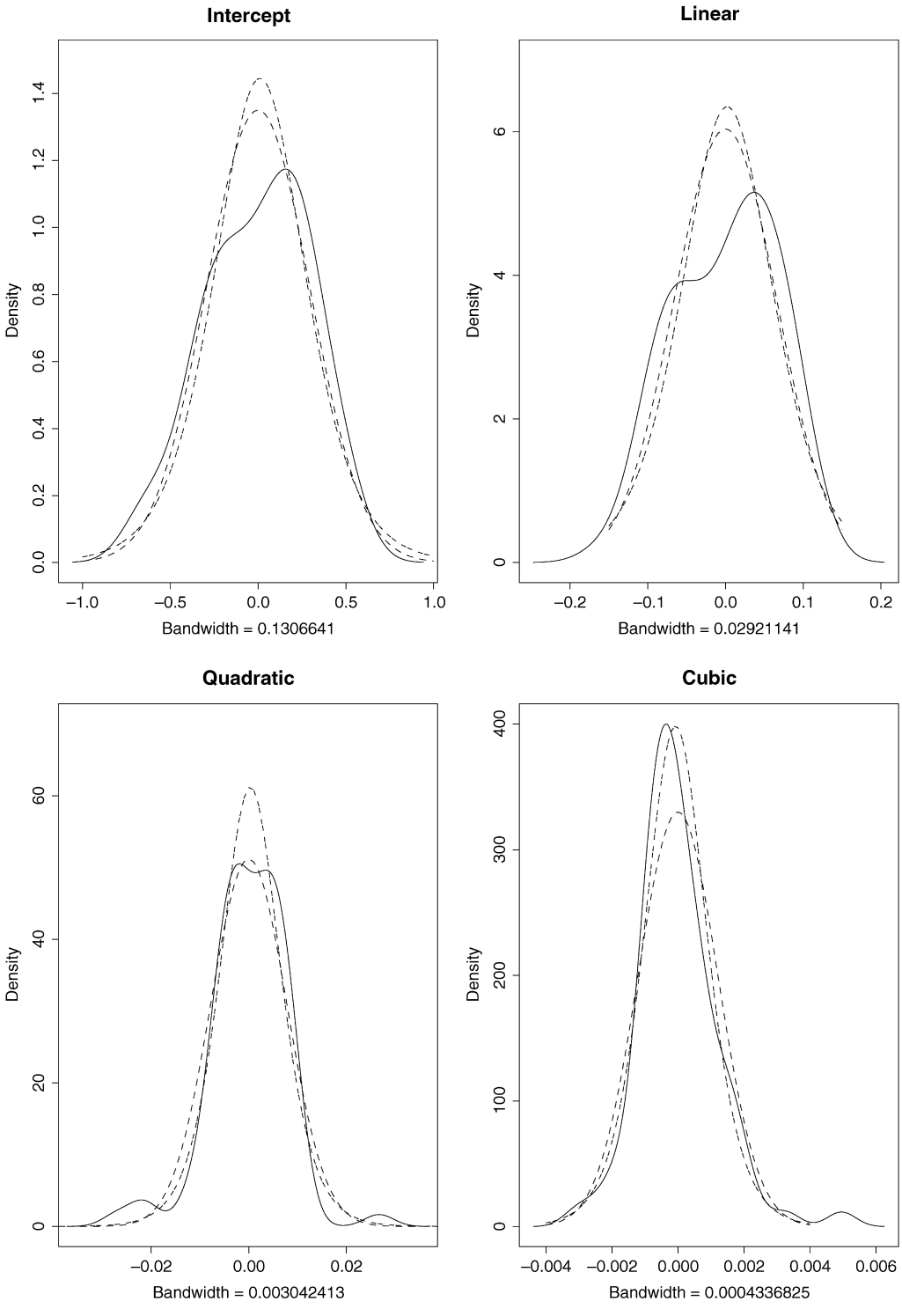

Fig. 4. Density plots for the fixed effects with times centred to have zero mean, for the model with the biological time origin. Solid: kernel density estimate; dashed: normal distribution; double-dashed: logistic distribution.

Table 5
AICs for fitting the normal, logistic, Cauchy, and stable distributions to the estimates of fixed effects for centred and uncentred time for the models with the two time origins

	Centred				Uncentred			
	Normal	Logistic	Cauchy	Stable	Normal	Logistic	Cauchy	Stable
Biological time origin								
Intercept	17.3	19.4	36.1	19.3	93.8	96.2	112.4	95.8
Linear	−101.1	−98.1	−80.2	−99.1	11.1	11.3	24.6	12.4
Quadratic	−269.9	−273.8	−262.9	−274.0	−154.2	−155.5	−144.7	−155.1
Cubic	−417.1	−421.2	−414.9	−420.6	−417.1	−421.2	−414.9	−420.6
Chronological time origin								
Intercept	25.8	28.2	44.5	27.8	165.7	157.6	154.6	153.5
Linear	−132.2	−130.4	−115.5	−130.2	56.4	49.0	47.4	46.5
Quadratic	−254.2	−258.5	−251.7	−257.2	−138.5	−145.2	−143.3	−145.8
Cubic	−425.0	−430.9	−426.6	−430.8	−425.0	−430.9	−426.6	−430.8

Hougaard (1986) has suggested the use of stable distributions in the context of random effects for survival data. We can try them here, as also presented in Table 5. (They are fitted by numerical inversion of the characteristic function; see Lambert and Lindsey, 1999.) Thus, the four parameter α-stable distribution can be used to check the relative goodness of fits of the normal ($\alpha = 2$) and Cauchy ($\alpha = 1$) distributions, where α is the parameter indexing the heaviness of the tails. For the intercept and linear components, they point to the normal distribution, with $\hat{\alpha} = 2$ (but are penalized by two extra parameters, hence the larger AIC). For the quadratic and cubic components, they fit about as well as the logistic distribution, with $\hat{\alpha}$ about 1.6, but indicate that some skew is present.

Now, let us consider the fixed effects estimates when time is not standardized (that is, runs from one to 19). Note that this is simply a different parametrization of the same model, with exactly the same fit (although this would not be true if a random effects model, with some given specific mixing distribution such as the normal, were fitted). The AICs are presented in the top right panel of Table 5. The kernel density estimates are plotted in Fig. 5, along with the fitted stable, normal, logistic, and Cauchy distributions.

(The fixed effects for interaction with the cubic in time are identical in the two parametrizations.) Here, the normal and logistic distributions fit about the same, but it is evident from the graphs that neither does very well.

It is instructive to compare these results to those for the model with the chronological time origin (which fits somewhat more poorly). The AICs are shown in the bottom panel of Table 5. The distributions are quite different, especially for uncentred time, as can be seen when they are plotted in Fig. 6.

Although the logistic distribution and the stable family fit about equally well, neither is capable of representing the kernel density curve very closely. In no case with uncentred mean, is the normal distribution a serious competitor. Here, α of the

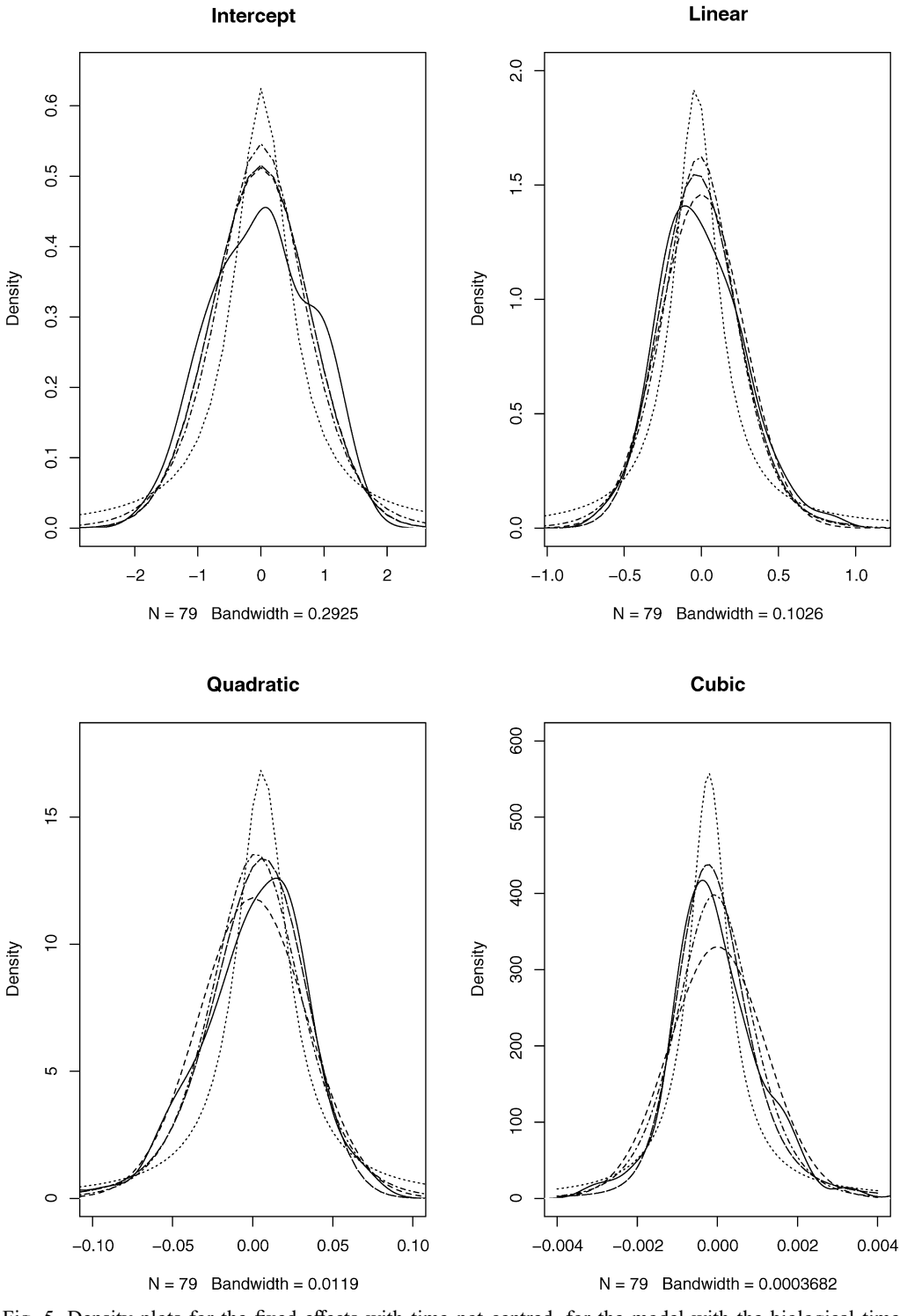

Fig. 5. Density plots for the fixed effects with time not centred, for the model with the biological time origin. Solid: kernel density estimate; short dashed: normal distribution; dot–dashed: logistic distribution; dotted: Cauchy distribution; long dashed: stable distribution.

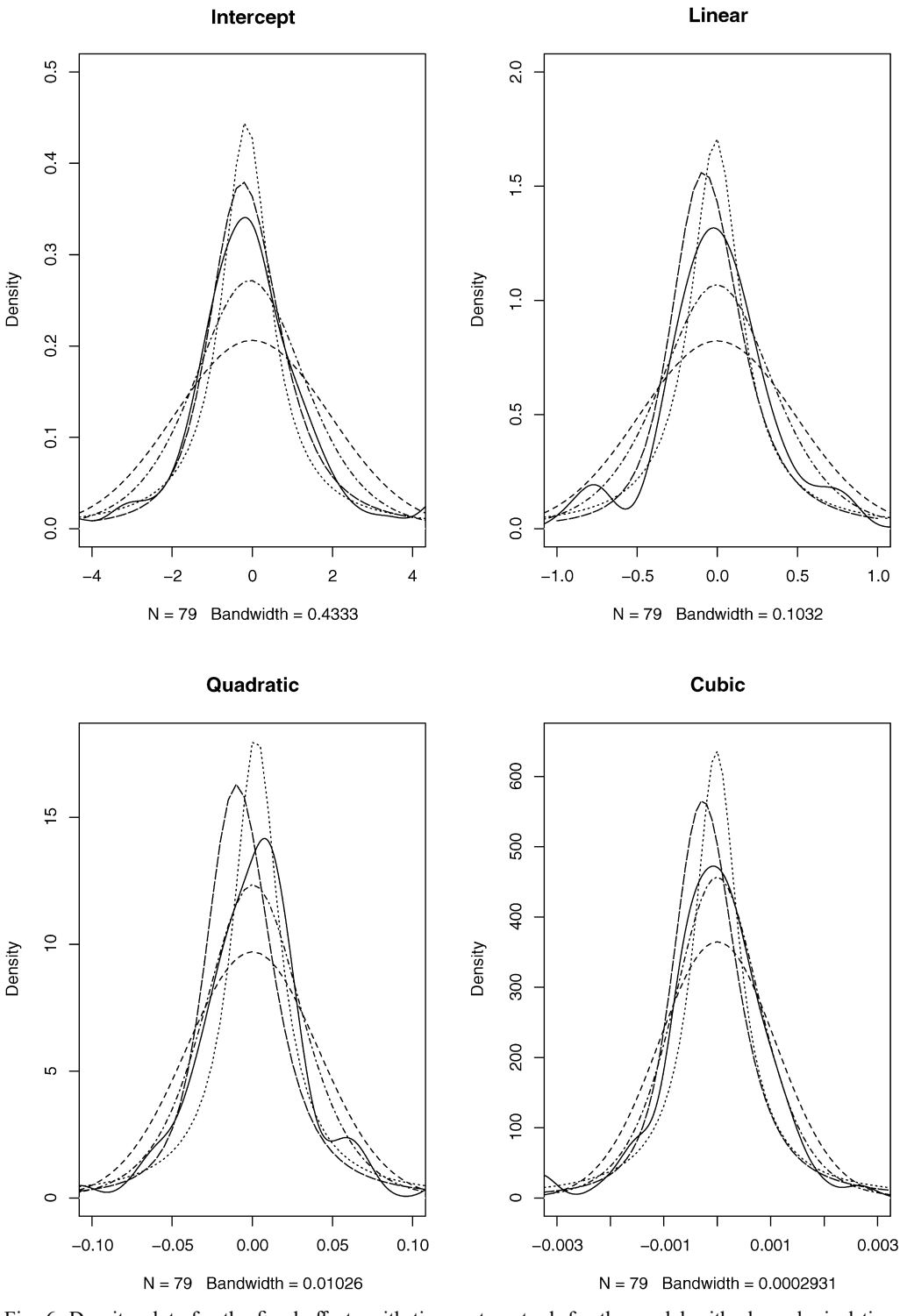

Fig. 6. Density plots for the fixed effects with time not centred, for the model with chronological time origin. Solid: kernel density estimate; short dashed: normal distribution; dot–dashed: logistic distribution; dotted: Cauchy distribution; long dashed: stable distribution.

Table 6
AICs for the full fixed effects model, with a cubic polynomial in time, treatment effect, settling period, and an AR(1) (320 parameters) fitted to the milk data for several conditional distributional assumptions, using linear regression

	Biological	Chronological
Normal	−161.2	−147.0
Gamma	−153.3	−139.0
Inverse Gauss	−143.7	−129.5

stable family is estimated between 1.3 and 1.6, indicating a distribution in between the normal and Cauchy, as can be seen from the graphs.

The need for thicker tailed distributions, such as the stable, may indicate that the structural part of the model fits most cows well, but that there are a few extreme individuals not well accounted for by the model. Note that the need for any non-normal mixing distribution of the random effects means that the induced multivariate distribution is not multivariate normal, even if the conditional distribution of the responses is normal.

We have not succeeded in finding an appropriate parametric mixing distribution for these data. Although not really true for the distributions we have tried for these data, the choice of the mixing distributions in a random effects model can depend critically on the parametrization used (for example, time origin at the beginning or centred). This is related to the warnings of Elston (1964) about the problem that the random effects growth curve model is highly dependent on the coding of time used.

5.4. Other response distributions

As a final step, we can consider other conditional distributions than the normal. Results for the two other standard generalized linear models are presented in Table 6.

Among these possibilities, the normal seems to be the appropriate choice. However, another symmetric distribution, such as the logistic or some member of the stable family might fit better (they do better when fixed effects are not included). Present computing power does not allow us to try this with such a large number of fixed effects parameters.

For the full fixed effects conditional normal model with the biological time origin, but standardized to be centred around its mean, the final parameter estimates (standard errors) for treatment effects are −0.493 (0.098) and −0.341 (0.107) for contrasts of mixed and lupin feed with the first treatment, barley. (These estimates are confounded with the fixed effect for cows so that an AIC for removing them is not available.) After allowing for heterogeneity in the time profile among the cows, the ordering of the treatments has changed markedly as compared to the estimates in the previous section. Note that, with heterogeneous (that is, non-parallel) profiles,

these estimates refer to means (intercepts) at the middle of the experiment; with uncentred time, the differences in intercepts become, respectively 1.102 (0.4473) and 0.526 (0.546), referring to means at the beginning. Exactly, the same problem of interpretation would occur if the equivalent random coefficients for the time profile were included, with time standardized in different ways.

5.5. Cohort effects

In fact, one simpler model, not considered above, does provide some improvement in fit over the original cubic polynomial model with treatment effect, an AR(1), and measurement error, although not as much as the fixed effects models. This model, over chronological time, has treatment differences only for the first cohort (that is, an interaction between treatment and cohort) with an interaction of this cohort with linear time and no settling in effect. (This result is already fairly evident from Fig. 6 of Diggle et al., 1994.) The continuous AR(1) model, with measurement error (there being no evidence for higher-order auto-regression or random effects), has an AIC of -59.9 whereas the linear regression model (without fixed effects or measurement error) has -27.3. The former fits better than the model with only a fixed effect for the intercept (AIC, -43.5) showing that we have accounted for individual differences in average response level, but not for differences in the shapes of individual profiles.

The mean profiles for the first cohort, for the three diets, are similar to those in Fig. 1, except that they start higher (above 4%) and finish lower (at about 3%). The treatment effects (standard errors) are estimated to be -0.174 (0.064) and -0.351 (0.064) for mixed and lupins with respect to barley, considerably larger than those obtained above when all cohorts were grouped together (without fixed effects, thus not taking into account differences in shape of the profiles). On the other hand, the mean profile for all other cohorts, with all diets, is closest to that for the mixed diet of the first cohort. The individual profiles for this model, for the same nine cows as in Fig. 2, are shown in Fig. 7.

This model is somewhat better than the equivalent one with individual time. Apparently, external factors over time may be most important; for some reason, the cows in the paddock calving at the beginning of the experiment were the only ones to benefit from the differences in diet. (Were cows introduced into their paddock at the beginning of the experiment or at calving?) However, such a conclusion must be interpreted cautiously because of the great variability among individual profiles not allowed for in this model.

6. Discussion

The standard diagnostic techniques, such as residual plots, that we presented at the beginning of our analysis, showed us no basic anomalies with the models that we developed there. (We also tried variograms, but they proved to be of little use, perhaps because they are inappropriate when random slopes are present.) And yet, further study, especially by fitting fixed effects models, demonstrate that these models fit the

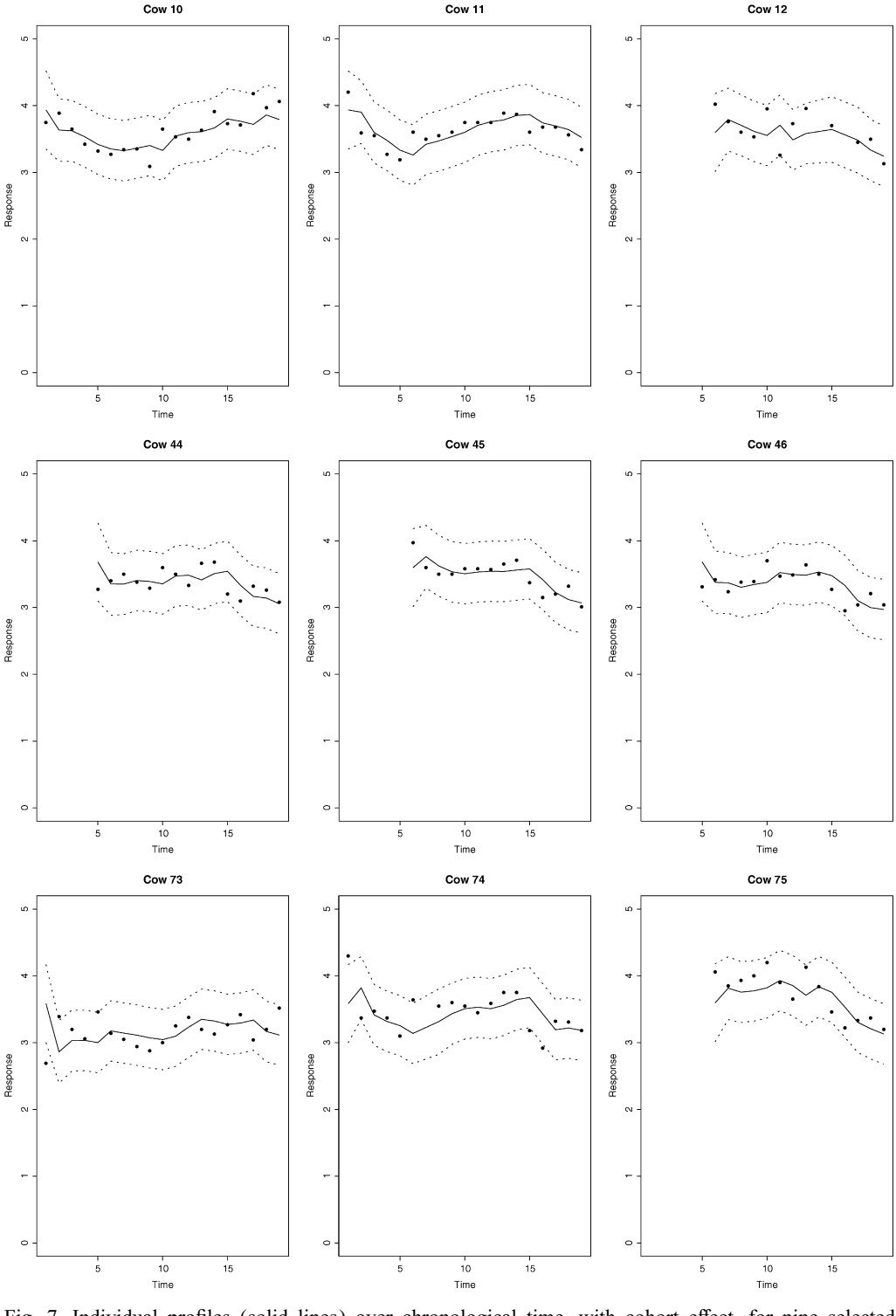

Fig. 7. Individual profiles (solid lines) over chronological time, with cohort effect, for nine selected cows, with the three rows corresponding, respectively, to cows receiving protein, mixed, and lupins. Twice the standard deviations are indicated by broken lines.

data very poorly. There is so much variability among the individual profiles of the cows that no valid inferences appear to be possible about differences in diet.

If software was available to fit random effects models with a variety of mixing distributions, such an indirect procedure of fitting a distribution to the corresponding estimated fixed effects would not be necessary. One could just fit models, much more parsimoniously, with various mixing distributions and compare the AICs. Unfortunately, computers will have to be considerably more powerful before this can become a routine interactive task, as was fitting all of the above models. However, even when this becomes possible, care will still have to be taken about the parametrization of the regression part of the model in so far as it affects these distributions. As we have seen, if the response profiles over time cannot be assumed to be parallel for all individuals, because either random coefficients or interactions of time with fixed effects are included, the interpretation of any treatment effects is extremely difficult. Such models imply that individual response differences among treatments are varying over time.

Random effects models are now widely used because they can easily be fitted in many software packages. However, as mentioned above, the need for random coefficients can point to a failure in the data collection, and subsequent modelling, processes. If random effects, especially random coefficients, are required, either information is lacking about the reasons for heterogeneity among individuals or other aspects of the model are poorly specified. Subtle and seemingly unimportant changes in the model structure can vastly change the form of the mixing distributions required to represent the random effects and can make interpretation of treatment differences impossible. In this particular case, the improvements brought to the model by the intercept, linear, quadratic, and cubic fixed effects implies that an important time-varying covariate (different for each individual) may be missing. This was not detected by simply fitting normal random effect models. Thus, such fixed effects models can be useful to assess the reliability and fit of normal random effect models.

Acknowledgements

A version of CARMA, supplied as Fortran code to the second author in 1990 by Richard Jones whom we thank, was used to produce the results included in this paper. This program was applied by means of a user-friendly front end constructed in R (Ihaka and Gentleman, 1996), a fast S-Plus clone freely available under the GNU licence, which we thank Robert Gentleman and Ross Ihaka for developing. It is available in the R public library called `growth` at www.luc.ac.be/~jlindsey/rcode.html.

Philippe Lambert, Nick Longford, and Geert Molenberghs provided many helpful comments on an earlier version of the paper.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In Petrov, B.N. Csàki, F. (Eds.), Second International Symposium on Inference Theory, Akadémiai Kiadó, Budapest: pp. 267–281.

Diggle, P.J., 1990. Time Series. A Biostatistical Introduction. Oxford University Press, Oxford.

Diggle, P.J., Kenward, M.G., 1994. Informative drop-out in longitudinal data analysis. Appl. Statist. 43, 49–93.
Diggle, P.J., Liang, K.Y., Zeger, S.L., 1994. The Analysis of Longitudinal Data. Oxford University Press, Oxford.
Elston, R.C., 1964. On estimating time-response curves. Biometrics 20, 643–647.
Elston, R.C., Grizzle, J.F., 1962. Estimation of time response curves and their confidence bands. Biometrics 18, 148–159.
Francis, B., Green, M., Payne, C., 1993. Glim 4: The Statistical System for Generalized Linear Interactive Modelling. Oxford University Press, Oxford.
Heitjan, D.F., 1991aa. Generalized Norton–Simon models of tumour growth. Statist. Med. 10, 1075–1088.
Heitjan, D.F., 1991bb. Nonlinear modeling of serial immmunologic data: a case study. J. Amer. Statist. Assoc. 86, 891–898.
Hendry, D., 1995. Dynamic Econometrics. Oxford University Press, Oxford.
Hougaard, P., 1986. Survival models for heterogeneous populations derived from stable distributions. Biometrika 73, 387–396 (correction, 75, 395).
Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. J. Comput. Graphics Statist. 5, 299–314.
Jones, R.H., 1993. Longitudinal Data Analysis with Serial Correlation: A State-space Approach. Chapman & Hall, London.
Jones, R.H., Ackerson, L.M., 1990. Serial correlation in unequally spaced longitudinal data. Biometrika 77, 721–731.
Jones, R.H., Boadi-Boateng, F., 1991. Unequally spaced longitudinal data with AR(1) serial correlation. Biometrics 47, 161–175.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.
Lambert, P., 1996. Modelling irregularly sample profiles of non-negative dog triglyceride responses under different distributional assumptions. Statist. Med. 15, 1695–1708.
Lambert, P., Lindsey, J.K., 1999. Generalized regression models for heavy-tailed processes based on non-symmetric stable distributions: a likelihood approach. Appl. Statist. 48, 409–424.
Lindsey, J.K., 1992. The Analysis of Stochastic Processes Using GLIM. Springer, Berlin.
Lindsey, J.K., 1993. Models for Repeated Measurements. Oxford University Press, Oxford.
Lindsey, J.K., Jones, B., 1997. Treatment–patient interactions for diagnostics of crossover trials. Statist. Med. 16, 1955–1964.
Lindsey, J.K., Jones, B., 1998. Choosing among generalized linear models applied to medical data. Statist. Med. 16, 59–68.
Little, R.J.A., 1995. Modeling the drop-out mechanism in repeated-measures studies. J. Amer. Statist. Assoc. 90, 1112–1121.
Potthoff, R.F., Roy, S.N., 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51, 313–326.
Verbyla, A.P., Cullis, B.R., 1990. Modelling in repeated measures experiments. Appl. Statist. 39, 341–356.