ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

JOINT MEAN-COVARIANCE MODELS WITH APPLICATIONS TO LONGITUDINAL DATA: UNCONSTRAINED PARAMETERISATION

By

Athanasia G. Georgountzou

A THESIS

Submitted to the Department of Statistics

of the Athens University of Economics and Business

in partial fulfilment of the requirements for

the degree of Master of Science in Statistics

Athens, Greece

November 2008

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΑΠΟ ΚΟΙΝΟΥ ΜΟΝΤΕΛΑ ΜΕΣΟΥ-ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΕ ΔΙΑΜΗΚΗ ΔΕΔΟΜΕΝΑ:

Αθανασία Γ. Γεωργούντζου

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής

του Οικονομικού Πανεπιστημίου Αθηνών

ως μέρος των απαιτήσεων για την απόκτηση

Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα

Νοέμβριος 2008

DEDICATION

I dedicate this thesis to my family: George, Panayotis and Violetta

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Vasilios Vasdekis for his help and guidance during the writing of this dissertation.
I would also like to thank my friend Vassia Hatzilouizou for her joyful help in technical issues.

VITA

I was born in Athens in 1984. I grew up in Kalamata and return to Athens to study Statistics and Actuarial Science at the University of Piraeus. Since then I am trying to improve my knowledge in statistics ending in this thesis.

CHAPTER 1

INTRODUCTION

Longitudinal data is a special type of multivariate data that occur when counts, measurements or categorical responses are obtained from one or more experimental units or subjects through time. A number of various approaches for representing longitudinal data in terms of a statistical model have been developed.

The data used in this thesis are from Kenward (1987) that reports an experiment in which cattle were assigned randomly to two treatment groups A and B, and their weights were recorded to study the effect of treatments on intestinal parasites. Thirty animals received treatment A and another 30 received treatment B. They are weighed 11 times over a 133-day period; the first 10 measurements on each animal were made at two-week intervals and the final measurement was made one week later. The measurement times were common across animals and were rescaled to $t=1,2,\ldots,10,11$. No observation was missing so this is a balanced longitudinal dataset.

Our goal is to find which model among these presented in this thesis fits best to the two different groups of our data. We compare among joint mean- covariance models (Pourahmadi, 1999) and mixed effects models.

CHAPTER 2

THEORY

2.1 Joint mean- covariance model

2.1.1 Cholesky Decomposition

Let our data $\mathbf{Y}_i$ (i=1,2,…,30) where $\mathbf{Y}_i \sim N(\mu,\Sigma)$ is the vector of the weight of the i-th cattle through time. We parameterize the covariance matrix Σ in terms of covariates. We are dropping the subscript i for avoiding confusions and we will use subscript t where Y_t is the weight of the cattle on the t-th moment (t=1,2,,,T=11). A link now that connects linear least predictors of Y_t, based on its predecessors, with the covariance matrix is the Cholesky Decomposition. Specifically, the modified Cholesky Decomposition of Σ^{-1}, not Σ, provides us an unconstrained parameterization of covariance (Pourahmadi, 1999). Since Σ^{-1} is the canonical covariance parameter of a multivariate normal distribution, modeling its unconstrained parameters as a linear combination of covariates is in agreement with the approach of generalized linear models.

We assume the following linear predictor of Y_t

$$\widehat{Y}_t = \mu_t + \sum_{j=1}^{t-1} \phi_{t,j}(Y_j - \mu_j) \Rightarrow Y_t = \mu_t + \sum_{j=1}^{t-1} \phi_{t,j}(Y_j - \mu_j) + e_t \Rightarrow$$

$$Y_t - \mu_t - \sum_{j=1}^{t-1} \phi_{t,j}(Y_j - \mu_j) = e_t \text{ where } e_t \sim N(0, \sigma_t^2)$$

Supposing z_t are the centred observations where $z_t = Y_t - \mu_t$

$$Y_t - \mu_t = z_t \Rightarrow$$

$$e_t = z_t - \sum_{j=1}^{t-1} \phi_{t,j} z_j \Rightarrow$$

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_T \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & . & . & 0 & 0 \\ -\phi_{2,1} & 1 & 0 & . & . & . & 0 \\ -\phi_{3,1} & -\phi_{3,2} & 1 & 0 & . & . & . \\ . & . & -\phi_{4,.3} & . & 0 & . & . \\ . & . & . & . & . & 0 & 0 \\ . & . & . & . & -\phi_{\mathrm{T}-1,n-1} & 1. & 0 \\ -\phi_{\mathrm{T},1} & . & . & . & -\phi_{\mathrm{T},n-1} & -\phi_{\mathrm{T},n} & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ . \\ . \\ . \\ z_{\mathrm{T}} \end{pmatrix} \Rightarrow e = TZ \quad \text{or} \quad e = \mathrm{T}\,\tilde{y}$$

where $\tilde{y} = y - \mu$

Also since prediction errors e_t are uncorrelated D=cov(e) is a diagonal matrix

$$\mathrm{D} = \mathrm{cov(e)} = \begin{pmatrix} \sigma_1^2 & 0 & . & . & 0 & 0 \\ 0 & \sigma_2^2 & 0 & . & . & 0 \\ . & 0 & . & 0 & . & . \\ . & . & 0 & . & 0 & . \\ . & . & . & 0 & . & 0 \\ 0 & 0 & . & . & 0 & \sigma_{\mathrm{T}}^2 \end{pmatrix} = \mathrm{T}\mathrm{cov}(\tilde{y})\mathrm{T}' = \mathrm{T}\,\mathrm{cov}(y)T' = \mathrm{T}\Sigma\mathrm{T}' \Rightarrow$$

D= cov(e) = $\mathrm{T}\Sigma\mathrm{T}'$

The Cholesky Decomposition gives us the matrix T which bellow-diagonal entries are the negatives of the coefficients of $\hat{Y}_t = \mu_t + \sum_{j=1}^{t-1} \phi_{t,j} (\mathrm{Y}_j - \mu_j)$, the linear least-squares predictor of Y_t based on its predecessors Y_{t-1}, … , Y_1 . It also provides us matrix D which diagonal entries are the prediction error variances $\sigma_t^2 = \mathrm{var}(\mathrm{Y}_t - \hat{Y}_t)$, for $1 \le t \le \mathrm{T}$.

Consequently, since the above entries of T and D have statistical meaning, we can trade in the $\frac{1}{2}$n(n+1) constrained and hard-to-model parameters of Σ for the $\frac{1}{2}$n(n+1) unconstrained and interpretable parameters $\varphi_{t,j}$,log σ_t^2, for $1 \le t \le \mathrm{T}$ and $1 \le j \le t-1$. We refer to the new parameters $\varphi_{t,j}$'s and σ_t^2's as the generalized autoregressive parameters and the innovation variances of Σ or Y.

2.1.2 Linear mean- covariance model

In order to make the most of the results given from the Cholesky Decomposition we use a joint mean-covariance model composed of three submodels describing the mean, variance and dependence of a random vector. Since $\varphi_{t,j}$ and $\log\sigma_t^2$ as we have defined them above are unconstrained, they are modeled in terms of covariates. To this end, for t=1,…,T and j=1,..,t-1, we consider the below three models

$$\mu_t = m(x_t,\beta), \quad \log\sigma_t^2=u(z_t,\lambda), \quad \varphi_{t,j}=d(z_{t,j},\gamma)$$

where m(.,.), u(.,.), d(.,.) are functions, x_t, z_t, $z_{t,j}$ are p x 1, q_1 x 1, q_2 x 1 vectors of covariates, and $\beta=(\beta_1,\ldots,\beta_p)'$, $\lambda=(\lambda_1,\ldots,\lambda_{q_1})'$ and $\gamma=(\gamma_1,\ldots,\gamma_{q_2})'$are parameters corresponding to the mean, variance and dependence, respectively.

When m(.,.), u(.,.) and d(.,.) are linear functions of their parameters we refer to such a model as a linear mean-covariance model.

$$h(\Sigma)=(\varphi_2',\ldots,\varphi_n', \log\sigma_1^2,\ldots, \log\sigma_n^2)'=Z_\alpha,$$

is the link function (McCullagh & Nelder, 1989, p.27) for a linear mean-covariance model. In order for this linear framework to be useful for our longitudinal data analysis, following Diggle, Liange & Zeger (1994, p.16), we use the following notation for the data, parameters and covariates:

$$Y=(Y_1',\ldots,Y_m')', \quad \mu=(\mu_1',\ldots,\mu_m')', \quad \Sigma=\text{block diag}(\Sigma_1,\ldots,\Sigma_m),$$

$$X=(X_1',\ldots,X_m')', \quad \mu_i=X_i\beta_i, \quad Z=\text{block diag}(Z_1,\ldots,Z_m), \quad h(\Sigma_i)=Z_i\alpha,$$

where now the subscript i refers to the *i*th subject (cattle) of our data.

Additionally we can write h(Σ) as a symmetric matrix Θ, where its main diagonal is the logarithm of the diagonal entries of D, its first subdiagonal, i.e. the lag-one regression coefficients, is the first subdiagonal of T and so forth. Since its entries are merely rearrangements of those of Z_α, we have

$$\Theta=\sum_{j=1}^{q}\alpha_j U_j,$$

where the U_j's are symmetric covariate matrices.

2.1.3 Examples

Example1. (a) (Pinheiro & Bates, 1996). For this Σ, its 6x1 vector of covariance predictors $h(\Sigma)$ is computed using

$$\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5 & 5 \\ 1 & 5 & 14 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix} = LDL' ,$$

$$T = L^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, \quad D = \mathrm{diag}(1,4,9), \;\; h(\Sigma) = (1,0,1,0,\log 4,\log 9)' .$$

It follows that $Y_1=\varepsilon_1$, $Y_t=Y_{t-1}+\varepsilon_t$, for t=2,3.

(b) Given $h(\Sigma) = (3, -1.5, -1, 0, -1, 2)'$ for an unknown 3x3 matrix Σ, the matrix is recovered by first constructing T and D :

$$T = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1.5 & 1 & 1 \end{pmatrix}, \quad D = \mathrm{diag}(1, e^{-1}, e^{2}).$$

Then one computes Σ.

Example 2. (a) For n=2 and q=1 and an arbitrary covariance covariate $Z=(z_1, z_2, z_3)'$, the generalised linear model amounts to the following reparameterisation of Σ:

$$\Sigma = e^{z_2 a} \begin{pmatrix} 1 & z_1 a \\ z_1 a & z_1^2 a^2 + e^{(z_1 - z_2)a} \end{pmatrix},$$

containing only one unconstrained parameter α. The parameterised correlation coefficient between Y_1 and Y_2 is given by $z_1 a \left[z_1^2 a^2 + \exp\{(z_3 - z_2)a\} \right]^{-\frac{1}{2}}$, which approaches ± 1 when $z_3 - z_2$ approaches $-\infty$. In a longitudinal study with two measurements made on a subject at times

$t_1 < t_2$, the choice of $z_1 = t_1$,

$z_2 = -(t_2 - t_1)$ and $z_3 = -(t_2 - t_1)^2$ leads to a covariance matrix with many desirable decay properties when t_1 and t_2 grow far apart.

(b) For n=2 and q=2, $\alpha = (\alpha_1, \alpha_2)'$ and

$$Z = \begin{pmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \end{pmatrix}' ,$$

the linear model for the unconstrained entries of $h(\Sigma)=Z_\alpha$ can be solved using $h(\Sigma)$ to express entries of Σ in terms of the new covariance parameters α_1 and α_2 and the explanatory variables in Z as follows :

$$\sigma_{11} = e^{\alpha_1 + z_2\alpha_2}, \quad \sigma_{21} = (\alpha_1 + z_1\alpha_2)e^{\alpha_1 + z_2 a_2}, \quad \sigma_{22} = e^{\alpha_1 + z_3\alpha_2} + (\alpha_1 + z_1\alpha_2)^2 e^{\alpha_1 + z_2 a_2}.$$

(c) The alternative representation of $h(\Sigma)$ above is

$$\Theta = \begin{pmatrix} \log\sigma_1^2 & \phi_{12} \\ \phi_{12} & \log\sigma_2^2 \end{pmatrix} = a_1 J + a_2 \begin{pmatrix} z_2 & z_1 \\ z_1 & z_3 \end{pmatrix} = \alpha_1 U_1 + a_2 U_2,$$

where U_1=J is the 2 x 2 matrix of 1's and the choice for U_2 is obvious.

Next, we highlight recognizable features of regressograms for AR(1) and compound symmetry.

Example 3. The covariance matrix of an AR(1) model is given by $\Sigma=\sigma^2\left(\rho^{|i-j|}\right)_{i,j=1}^{n}$, for $|\rho|<1$ and $\sigma^2>0$. It follows that, for $t\geq 2$, $\varphi_t=(0,\ldots,0,\rho)'$, $\sigma_t^2=\sigma^2$, and $\sigma_1^2=\sigma^2\left(1-\rho^2\right)^{-1}$, so that, assuming that $\mu=0$, $Y_1,\ldots,Y_n$ satisfy $Y_1=\varepsilon_1$ and $Y_t=\rho Y_{t-1}+\varepsilon_t$, for $2\leq t\leq n$.

Here, only the lag-one generalised autoregressive parameters are nonzero and σ_1^2 is a nonlinear function of ρ. Thus, the theoretical regressograms for AR(1) and more generally for AR(p) models are simpler to recognise; they drop off to zero for lags j>p and σ_t^2 is constant for t>p.

2.1.4 Estimation - The likelihood function

We will compute the multivariate normal likelihood function, which has three distinct representations corresponding to the three sets of submodels when the observations $Y_i \sim N(\mu, \Sigma)$, for i=1,2,…,m, are independent.

We define $r_i = y_i - \mu_i = (r_{i,t})_{i=1}^{m}$, we obtain T$r_i$= $r_i - \widehat{r}_i$, for i= 1,2,…,m, where $\widehat{r}_{i,t}$ is the best linear predictor of $r_{i,t}$ based on its predecessors $r_{i,j}$, for

1≤j≤t-1. In the following, we also need r(t) = $(r_{i,t})_{i=1}^{m}$ which is the vector of the centred observations made on the *t*th occasion on all *m* subjects.

The quadratic form Q in the exponent of the likelihood function can be written

$$Q = \sum_{i=1}^{m} (y_i - \mu_i)' \Sigma^{-1} (y_i - \mu_i) = \sum_{i=1}^{m} r_i' \mathrm{T}' D^{-1} \mathrm{T} r_i$$

$$= \sum_{i=1}^{m} (r_i - \widehat{r}_i)' D^{-1} (r_i - \widehat{r}_i) = \sum_{i=1}^{m} \sum_{t=1}^{n} \frac{(r_{i,t} - \widehat{r}_{i,t})^2}{\sigma_t^2} = \sum_{t=1}^{n} \frac{RSS_t}{\sigma_t^2},$$

where

$$RSS_t = \sum_{i=1}^{m} (r_{i,t} - \widehat{r}_{i,t})^2$$

is the residual sum of squares for the analysis of covariance of r(t) with r(t-1),…,r(1) as covariates (Kenward, 1987).

When we assume a linear mean-covariance model

$h(\Sigma) = (\varphi_2', \ldots, \varphi_n', \log\sigma_1^2, \ldots, \log\sigma_n^2)' = Z_\alpha$

it is evident that RSS_t and hence Q are quadratic functions of the correlation parameters γ:

$$\mathrm{RSS}_t = \sum_{i=1}^{m} \left(r_{i,t} - \sum_{j=1}^{t-1} \phi_{t,j} r_{i,j} \right)^2 = \sum_{i=1}^{m} \left\{ r_{i,t} - \left(\sum_{j=1}^{t-1} z_t' r_{i,j} \right) \gamma \right\}^2 = \sum_{i=1}^{m} \{ r_{i,t} - z'(i,t)\gamma \}^2,$$

$$\mathrm{Q} = \sum_{i=1}^{m} \sum_{t=1}^{n} \sigma_t^{-2} \{ r_{i,t} - z'(i,t)\gamma \}^2 = \sum_{i=1}^{m} \{ r_i - Z(i)\gamma \}' D^{-1} \{ r_i - Z(i)\gamma \},$$

where

$$z(i,t)=\sum_{j=1}^{t-1} z_{t,j} r_{i,j},\ \mathrm{Z(i)} = (\mathrm{z(i,1)},\ldots,\mathrm{z(i,n)})',$$

are respectively q_2x1 and nxq_2 matrices.

The loglikelihood L(β,λ,γ;Y), up to the additive constant $mnlog2\pi$, satisfies

$$-2L(\beta,\lambda,\gamma;Y)=m\log|\Sigma|+\sum_{i=1}^{m}(y_i-X_i\beta)'\Sigma^{-1}(y_i-X_i\beta)=m\sum_{t=1}^{n}\log\sigma_t^2+\sum_{t=1}^{n}\frac{RSS_t}{\sigma_t^2}=$$

$$=m\sum_{t=1}^{n+}\log\sigma_t^2+\sum_{i=1}^{m}\{r_i-Z(i)\gamma\}'D^{-1}\{r_i-Z(i)\gamma\}$$

If we use the above representations of the likelihood, the score vector and the Fisher expected information can be computed. Also a three-stage estimation procedure can be developed as involving three sub-models for the mean, variance and correlation (Smyth, 1989; Verbyla, 1993). For given (λ,γ) or Σ, the first equation defines the mean model with y_i as its response; for given β and γ, the second identity is viewed as the variance model with RSS_t as response; and, for given β and λ, the third identity can be viewed as the correlation model with r_i as response. In this thesis we used the Matlab package and specifically the simplex algorithm to minimize the -2loglikelihood function.

2.1.5 Hypothesis Testing

To compare the fits rigorously, it is standard to rely on penalised likelihood criteria such AIC and BIC. We use the BIC criteria which in our context of covariance model selection is defined as

$$\mathrm{BIC} = -\frac{2}{m}L+(p+q_1+q_2)\frac{\log m}{m},$$

(Pan and Mackenzie, 2003) where m is the sample size, L is the maximised loglikelihood for the joint mean-covariance model and $p+q_1+q_2$ is the number of parameters in the 3 associated submodels. A smaller value of BIC is associated with a better fitting model. As we can see from the definition of the BIC criteria presented above a large value of the maximized loglikelihood will give a small value of BIC. On the contrary a large amount of parameters will give a big value of BIC.

2.2 Normal Mixed model

Mixed models analysis is a type of statistical analysis based on linear regression models. In particular, it applies to research involving factor whose levels can be controlled by the researcher (fixed) as well as factors whose levels are beyond the researcher's control (random effects).

Here, the mixed model is defined using a general matrix notation which provides a compact means to specify all types of mixed model. We start by defining the fixed effects model and then extend this notation to encompass the mixed model.

2.2.1 The fixed effects model

All fixed effects model can be specified in the general form

$$y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_p x_{ip} + e_i$$

$$\mathrm{var}(e_i) = \sigma^2$$

This model fits $p+1$ fixed effects parameters, α_1 to α_p, and an intercept term, μ. If there are n observations, then these may be written as

$$y_1 = \mu + \alpha_1 x_{11} + \alpha_2 x_{12} + \ldots + \alpha_p x_{1p} + e_1$$

$$y_2 = \mu + \alpha_1 x_{21} + \alpha_2 x_{22} + \ldots + \alpha_p x_{2p} + e_2$$

.
.
.

$$y_n = \mu + \alpha_1 x_{n1} + \alpha_2 x_{n2} + \ldots + \alpha_p x_{np} + e_n$$

$$\mathrm{var}(e_1) = \sigma^2,$$

.
.
.

$$\mathrm{var}(e_n) = \sigma^2.$$

This can be expressed more concisely in matrix notation as

$\mathbf{y} = \mathbf{X\alpha} + \mathbf{e}$,

$\mathbf{V} = \mathrm{var}(\mathbf{y}) = \sigma^2\,\mathbf{I}$

where

$\mathbf{y} = (y_1, y_2, y_3, \ldots, y_n)'$ = observed values,

$\boldsymbol{\alpha} = (\mu, \alpha_1, \alpha_{2,\ldots,}\, \alpha_p)'$ = fixed effects parameters,

$\mathbf{e} = (e_1, e_2, e_3, \ldots, e_n)'$ = residuals,

σ^2 = residual variance,

$\mathbf{I}$ = n x n identity matrix

The parameters in $\boldsymbol{\alpha}$ may encompass several variables. Those could be qualitative or categorical variables and we will refer to such effects as categorical effects. They are also sometimes referred to as factor effects. More generally, categorical effects are those where observations will belong to one of several classes. There may also be several covariate effects (such as time or baseline measurement) contained in **a.** These relate to variables which are measured in a quantitative scale. Several parameters may be required to model categorical effects, but just one parameter is needed to model a covariate effect.

X is known as the design matrix and has the dimension n x p (i.e. n rows and p columns). It specifies values of fixed effects corresponding to each parameter for each observation. For categorical effects the values of zero and one are used to denote the absence and presence of effect categories, and for covariate effects the variables themselves are used in **X.**

V is a matrix containing the variances and covariances of the observations. In the usual fixed effects model, variances for all observations are equal and no observations are correlated. Thus **V** is simply $\sigma^2\mathbf{I}$.

2.2.2 The mixed model

Extend our fixed effects model to incorporate random effect, the mixed model may be specified as

$$y_i = \mu + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_p x_{ip} + \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_q z_{iq} + e_i$$

for a model fitting p fixed effect parameters and q random effect (or coefficient) parameters. Random effects are assumed to follow a distribution, whereas fixed effects are regarded as fixed constants.

The model can be expressed in matrix notation as

$$\mathbf{y} = \mathbf{X\alpha} + \mathbf{Z\beta} + \mathbf{e},$$

where **y, X, α** and **e** are as defined in the fixed effects model, and

$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \ldots, \beta_n)'$ = random effect/ coefficient parameters.

Z is a second design matrix with dimension *n x q* giving the values of random effects corresponding to each observation. It is specified in exactly the same way as **X** was for the fixed effects, except that an intercept term is not included.

2.2.3 Covariance matrix Σ

We saw in the fixed effects model that all observations have equal variances and the observations are uncorrelated. This leads to the **Σ** matrix being diagonal. Random effects result in correlated observations. The covariance of y, var(y) = **Σ** can be written as

$$\boldsymbol{\Sigma} = \mathrm{var}(\mathbf{X\alpha} + \mathbf{Z\beta} + \mathbf{e})$$

Since we assume that the random effects and the residuals are uncorrelated.

$$\boldsymbol{\Sigma} = \mathrm{var}(\mathbf{X\alpha}) + \mathrm{var}(\mathbf{Z\beta}) + \mathrm{var}(\mathbf{e})$$

Since **α** describes the fixed effects parameters, $\mathrm{var}(\mathbf{X\alpha}) = 0$. Also, **Z** is a matrix of constants. Therefore,

$$\boldsymbol{\Sigma} = \mathbf{Z}\,\mathrm{var}(\boldsymbol{\beta})\mathbf{Z}' + \mathrm{var}(\mathbf{e})$$

We will let **G** denote var(**β**), and since the random effects are assumed to follow normal distributions we may write $\boldsymbol{\beta} \sim N(0, \mathbf{G})$. Similarly, we write $\mathrm{var}(\mathbf{e}) = \mathbf{R}$, the residual covariance matrix and

$\mathbf{e} \sim N(0, \mathbf{R})$. Hence,

$$\boldsymbol{\Sigma} = \mathbf{ZGZ}' + \mathbf{R}$$

2.2.4 Likelihood function

The mixed model can be fitted by maximising the likelihood function for values of the data. The likelihood function, L, measures the likelihood of the model parameters given the data and is defined using the density function of the observations. In models where the observations are assumed independent (e.g. fixed effects models), the likelihood function is simply the product of the density functions for each observation. However, observations in a mixed model are not independent and the likelihood function therefore needs to be based on a multivariate density function for the observations. The likelihood for the variance parameters and the fixed effects can be defined using the multivariate normal distribution for y (the term ‘variance parameters’ encompasses all parameters in the **G** and **R** matrix, i.e. variance components and the covariance parameters). As random effects have expected values of zero and therefore do not affect the mean, this distribution has a mean vector $\mathbf{X\alpha}$ and a covariance matrix $\mathbf{\Sigma}$. The likelihood function based on the multivariate normal density function is then

$$L = \frac{\exp\left[-\frac{1}{2}(Y - X\alpha)' \Sigma^{-1} (Y - X\alpha)\right]}{(2\pi)^{(1/2)n} |\Sigma|^{(1/2)}}$$

In practice the log likelihood function is usually used in place of the likelihood function since it is simpler to work with and its maximum value coincides with that of the likelihood. The log likelihood is given by

$$\log(L) = k - \frac{1}{2}\left[\log|\Sigma| + (Y - Xa)' \Sigma^{-1} (Y - Xa)\right]$$

where

k= $-\frac{1}{2} n \log(2\pi)$ a constant that can be ignored in the maximization process

n= number of observations.

The values of the model parameters which maximize the log likelihood can then be determined.

2.2.5 Hypothesis Testing

To find the mixed effects model that fits better our data we will use the BIC criteria

$$\text{BIC} = -\frac{2}{m}L + p\frac{\log m}{m},$$

where m is the sample size, L is the maximised loglikelihood for the mixed effects model and *p* is the sum of the parameters included in the model. We will use again the BIC criteria so we will be able to compare the results from the several models to find out which fits best.

CHAPTER 3

ANALYSIS

3.1 Outline of the data

We have taken our data from an experiment (Kenward, 1987) where 60 cattle were partitioned randomly into two different groups (A, B) of 30 cattle. Each group received a different treatment on intestinal parasites in order to observe the difference between the weights of the cattle of the two different groups. They were weighted n=11 times over a 133-day period. The first 10 measurements on each animal were made at two-week intervals and the final measurement was made one week later. The measurement times were common across animals and were rescaled to t=1,2,…,10,11.We study each group separately starting with group A.

3.2 Cattle analysis for group A

Our primary intention is to detect a model that will explain in the most effective way the weight through time of the 30 cattle that belong to group A. In the first place we will apply the joint- mean covariance model when the mean is assumed to be a common vector μ and when we describe the mean model with a broken line model. Afterwards we will apply mixed – effects models in our data to discover which one (random intercept-random intercept and slope) suits us the most, also when the mean is assumed to be a common vector μ and when we refer to the mean model as a broken line model.

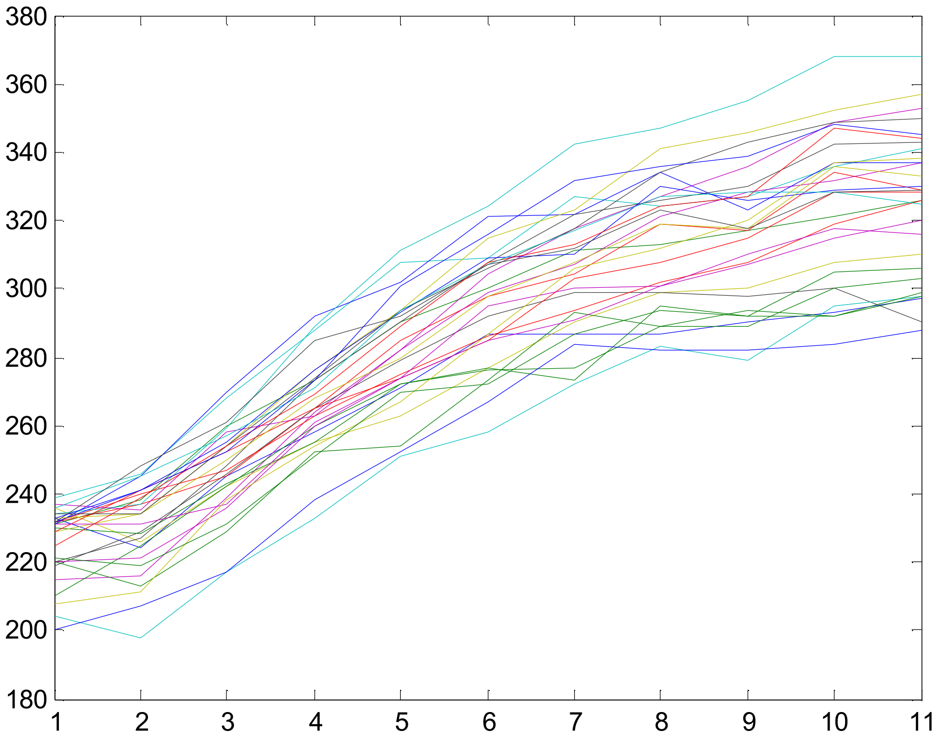

Figure 3.1. Plot of weight through time

This plot is useful for assessing overall aspects of the data. For example, the plot above suggests that there is a general linear increase in the weights across time and that there is considerable individual homogeneity in this.

3.2.1 Joint mean- covariance model for group A

Our purpose is to find, for the treatment group A with m=30 animals, the joint-mean covariance model

$$\mu_t = m(x_t,\beta), \quad \log\sigma_t^2 = u(z_t,\lambda), \quad \varphi_{t,j} = d(z_{t,j},\gamma)$$

(Pourahmadi, 1999) that fits best. Firstly, for the mean model we assume a common mean vector $\mu = (\mu_1, \mu_2,\ldots, \mu_{11})$ where

$\mu_t = \frac{1}{30}\sum_{i=1}^{30} y_{it}$,t=1,2,…,11.We will use this vector μ as initial values to estimate the parameters that will maximize the likelihood. Then, we apply the Cholesky Decomposition in our 11x11 sample covariance matrix

$S=\frac{1}{30}\sum_{i=1}^{30}(y_i-\mu)(y_i-\mu)'$ to find the T,D matrices that give us the $\varphi_{t,j}$'s and σ_t^2's sample parameters.

Since we have found the $\varphi_{t,j}$'s and σ_t^2's sample parameters using Cholesky Decomposition we assume that our data are independent and we are trying to find linear models that will fit well to our parameters. The parameters of the models that fit better to our sample parameters will be used as initial values to the $\log\sigma_t^2=u(z_t,\lambda)$, $\varphi_{t,j}=d(z_{t,j},\gamma)$ submodels of the joint mean-covariance model.

3.2.1.1. Model for the generalized autoregressive parameters

From the plot of $\varphi_{t,j}$ versus time lags j=1,2,...,t-1 we observe that the sample generalized autoregressive parameters could be a linear function and specifically a cubic function.

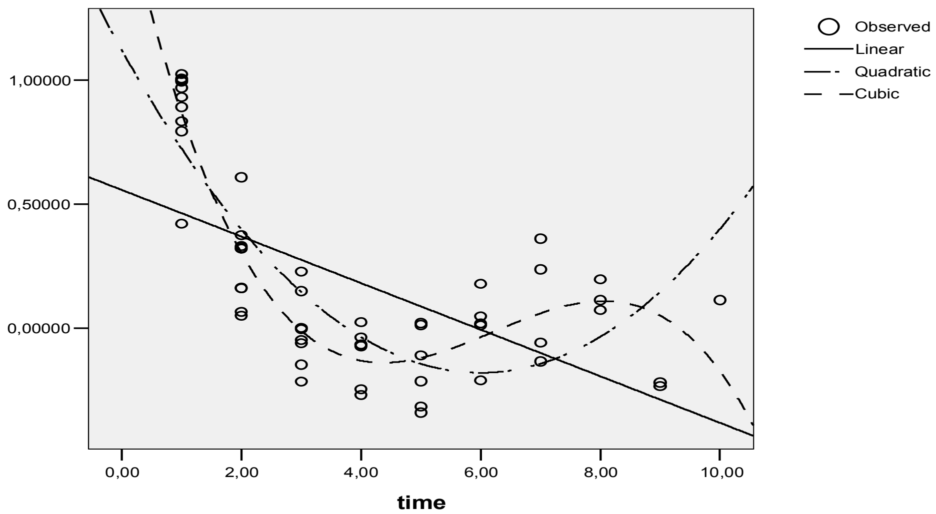

Figure 3.2 . Plot of the sample generalized autoregressive parameters through time with linear, quadratic and cubic function

We applied to our parameters three linear models to find out which is the most suitable according to the F test and the R^2. The models with the corresponding analysis of variance are cited below.

a) The linear polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2 (t-j) + \varepsilon_{t,j,d} \;\Rightarrow \widehat{\phi}_{t,j} = 0{,}5567 - 0{,}0939(t-j) + \varepsilon_{t,j,d}$$

b) The quadratic polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2 (t-j) + \gamma_3 (t-j)^2 + \varepsilon_{t,j,d}$$

$$\Rightarrow \widehat{\phi}_{t,j} = 1{,}121 - 0{,}434(t-j) + 0{,}036(t-j)^2 + \varepsilon_{t,j,d}$$

c) The cubic polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2 (t-j) + \gamma_3 (t-j)^2 + \gamma_4 (t-j)^3 + \varepsilon_{t,j,d} \qquad \Rightarrow$$

$$\widehat{\phi}_{t,j} = 1{,}743 - 1{,}052(t-j) + 0{,}186(t-j)^2 - 0{,}010(t-j)^3 + \varepsilon_{t,j,d}$$

MODEL		**SUM OF SQUARES**	**df**	**MEAN SQUARES**	**F**	**p-value**
LINEAR	REGRESSION	2,910	1	2,910	28,823	0,000
	RESIDUAL	5,352	53	0,101		
	TOTAL	8,262	54			
QUADRATIC	REGRESSION	5,609	2	2,805	54,970	0,000
	RESIDUAL	2,653	52	0,051		
	TOTAL	8,262	54			
CUBIC	REGRESSION	6,837	3	2,279	81,531	0,000
	RESIDUAL	1,425	51	0,028		
	TOTAL	8,262	54			

Table 3.1.Anova for the sample generalized autoregressive parameters

MODEL	**R^2**	**R^2 ADJUSTED**
LINEAR	0,352	0,340
QUADRATIC	0,679	0,667
CUBIC	0,827	0,817

Table 3.2. R^2 values for the sample generalized autoregressive parameters

As far as we can see from the tables above, all models fit well to our generalized autoregressive parameters according to the F-test since p-value<0,001. On the other hand, the F-value appears to take bigger values when we increase the number of parameters γ and respectively the degrees of the polynomial. Similarly the R^2 adjusted seems to augment while we add terms, meaning that the cubic function

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \varepsilon_{t,j,d} \Rightarrow$$

$$\widehat{\phi}_{t,j} = 1{,}743 - 1{,}052(t-j) + 0{,}186(t-j)^2 - 0{,}010(t-j)^3 + \varepsilon_{t,j,d}$$

is the most appropriate for the parameters.

3.2.1.2. Model for the innovation parameters

From the plot of σ_t^2 versus time t=1,2,...,11 we observe that the logarithms of the sample innovation parameters could also be a linear function and specifically a cubic function.

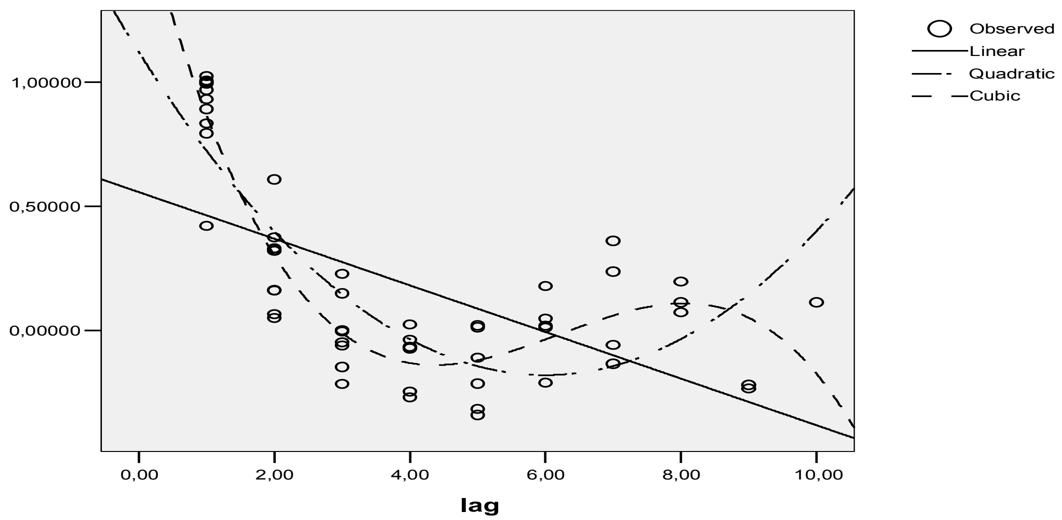

Figure 3.3 . Plot of the sample innovation parameters through time with linear, quadratic and cubic function

We applied to our parameters four linear models to find out which is the most suitable according to the F test and the R^2. The models with the corresponding analysis of variance are cited below.

a) The linear polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t \varepsilon_{t,v} \Rightarrow \log \hat{\sigma}_t^2 = 4{,}229 - 0{,}142t + \varepsilon_{t,v}$$

b) The quadratic polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 \varepsilon_{t,v} \Rightarrow \log \hat{\sigma}_t^2 = 4{,}393 - 2{,}18t + 0{,}006t^2 + \varepsilon_{t,v}$$

c) The cubic polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \varepsilon_{t,v} \Rightarrow$$
$$\log \hat{\sigma}_t^2 = 5{,}735 - 1{,}326t + 0{,}227t^2 - 0{,}012t^3 + \varepsilon_{t,v}$$

d) The 4th degree polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \lambda_5 t^4 + \varepsilon_{t,v} \Rightarrow$$
$$\log \hat{\sigma}_t^2 = 6{,}0611 - 1{,}7196t + 0{,}3607t^2 - 0{,}029t^3 + 0{,}0007t^4 + \varepsilon_{t,v}$$

MODEL		**SUM OF SQUARES**	**df**	**MEAN SQUARES**	**F**	**p-value**
LINEAR	REGRESSION	2,229	1	2,229	13,839	0,005
	RESIDUAL	1,450	9	0,161		
	TOTAL	3,679	10			
QUADRATIC	REGRESSION	2,263	2	1,131	6,393	0,022
	RESIDUAL	1,416	8	0,177		
	TOTAL	3,679	10			
CUBIC	REGRESSION	3,196	3	1,065	15,438	0,002
	RESIDUAL	0,483	7	0,069		
	TOTAL	3,679	10			
4th DEGREE	REGRESSION	3,210	4	0,8025	10,275	0,004
	RESIDUAL	0,469	6	0,0781		
	TOTAL	3,679	10			

Table 3.3.Anova for the sample innovation parameters

MODEL	R^2	R^2 ADJUSTED
LINEAR	0,606	0,562
QUADRATIC	0,615	0,519
CUBIC	0,869	0,812
4^{th} DEGREE	0,874	0,790

Table 3.4. R^2 values for the sample innovation parameters

As far as we can see from the tables above all models fit adequately according to the F-test since p-value<0,01 for all models except the quadratic polynomial. Particularly, we observe that the models that explain better the variance of the innovation parameters are the cubic and the 4^{th} degree polynomials. We will prefer the cubic polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \varepsilon_{t,v} \Rightarrow$$
$$\log \hat{\sigma}_t^2 = 5{,}735 - 1{,}326t + 0{,}227t^2 - 0{,}012t^3 + \varepsilon_{t,v}$$

because its R^2_{adjusted} value is bigger (0,812> 0,790).

3.2.1.3 Broken line model

In the linear mean- covariance model that we structured above we let a full model for the mean. In this session we will replace the common mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_{11})$ where $\mu_t = \frac{1}{30}\sum_{i=1}^{30} y_{it}$,t=1,2,…,11 that we assumed before with a broken line model in order to reduce the number of the mean parameters. As we can observe from the plot of the weights of the cattle through time, we can represent the data easily with three broken lines. We see in the plot two turning points that distinguish on time points 2 and 5. We construct a new design matrix for the mean that consist of three new variables x, y, z .

Let $x = \begin{cases} 1 & t=1 \\ 2 & t>1 \end{cases}$, $y = \begin{cases} 0 & t\le 2 \\ t-2 & t>2 \end{cases}$, $z = \begin{cases} 0 & t\le 7 \\ t-7 & t>7 \end{cases}$

Afterwards we apply this design matrix for the mean and we study how this model fits to our longitudinal data.

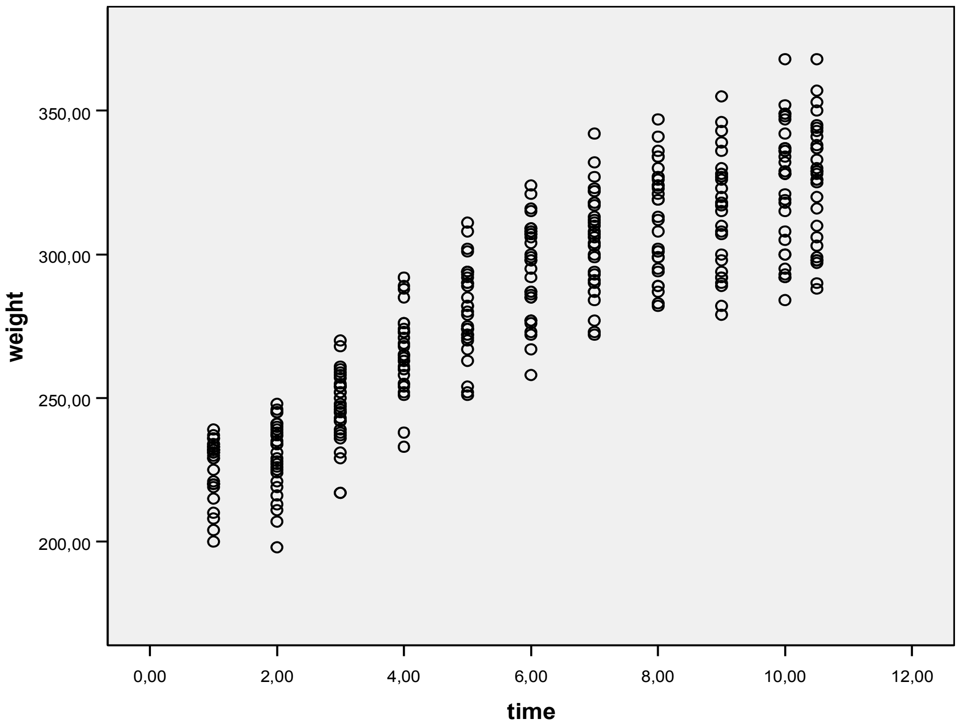

Figure 3.4. Plot of weight through time

We suppose the linear model

$$y_{ij}= \beta_0 + \beta_1 x_{ij} +\beta_2 y_{ij} +\beta_3 zij + \varepsilon_{ij}$$

for the mean where i=1, 2 ,.., 30 , j=1, 2, …, 11 and $\varepsilon_{ij} \sim N(0, \sigma^2)$

From the tables below we find out that according to regression linear analysis our broken line model has the form

$y_{ij}= \beta_0 + \beta_1 x_{ij} +\beta_2 y_{ij} +\beta_3 z_{ij}+ \varepsilon_{ij} \Rightarrow y_i= 219{,}685 + 6{,}515x_{ij} +15{,}125y_{ij} +4{,}459z_{ij} + \varepsilon_{ij}$

The analysis of variance indicates that this model fits adequately to our data and explain a significant percentage of variance. The p-value of the F-test is 0,000 meaning that the F-value= 485,857 belongs to the $F_{3,326}$ distribution. We are also able to see that the R^2 value of the model is very satisfactory (R^2 = 0,817). On the other hand we observe from the scatterplot of the studentized residuals and the unstandardised predicted values that the residuals of the model do not have a random form that indicates a bad fit for our model.

Model	Sum of Squares	df	Mean Square	F	p-value
Regression	407101,496	3	135700,499	485,857	,000
Residual	91052,201	326	279,301		
Total	498153,697	329			

Table 3.5. Anova for broken line model

Model	Unstandardized Coefficients		Standardized Coefficients	t	p-value
	B	Std. Error	Beta		
Constant	219,685	6,477		33,918	,000
x	6,515	3,744	,048	1,740	,083
y	15,125	,662	,775	22,851	,000
z	4,459	,844	,158	5,285	,000

Table 3.6. Estimated coefficients and t-test for broken line model

R	R Square	Adjusted R Square	Std. Error of the Estimate
,904	,817	,816	16,71231

Table 3.7. Model summary (R^2 values) for broken line model

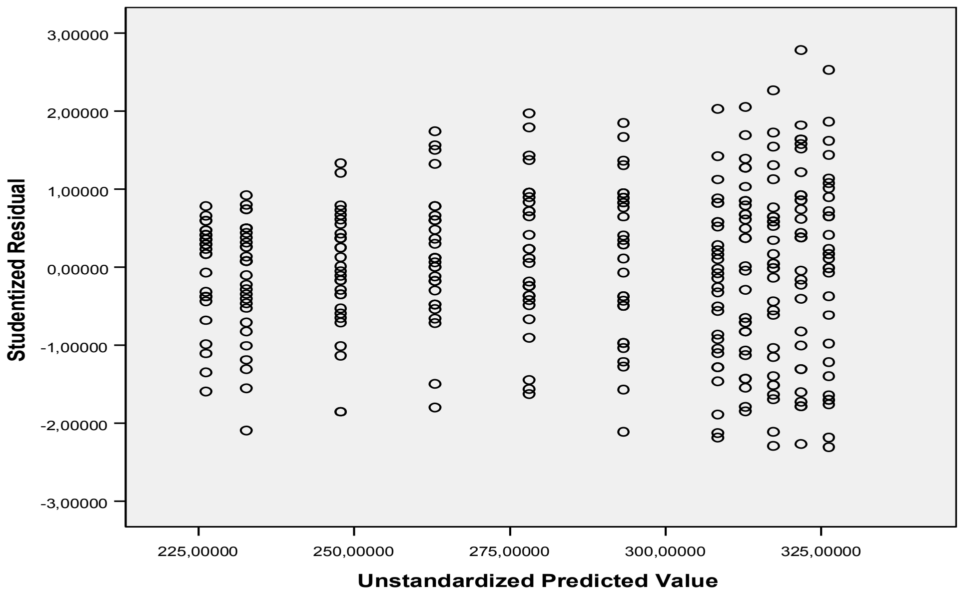

Figure 3.5. Studentized residuals of broken line model

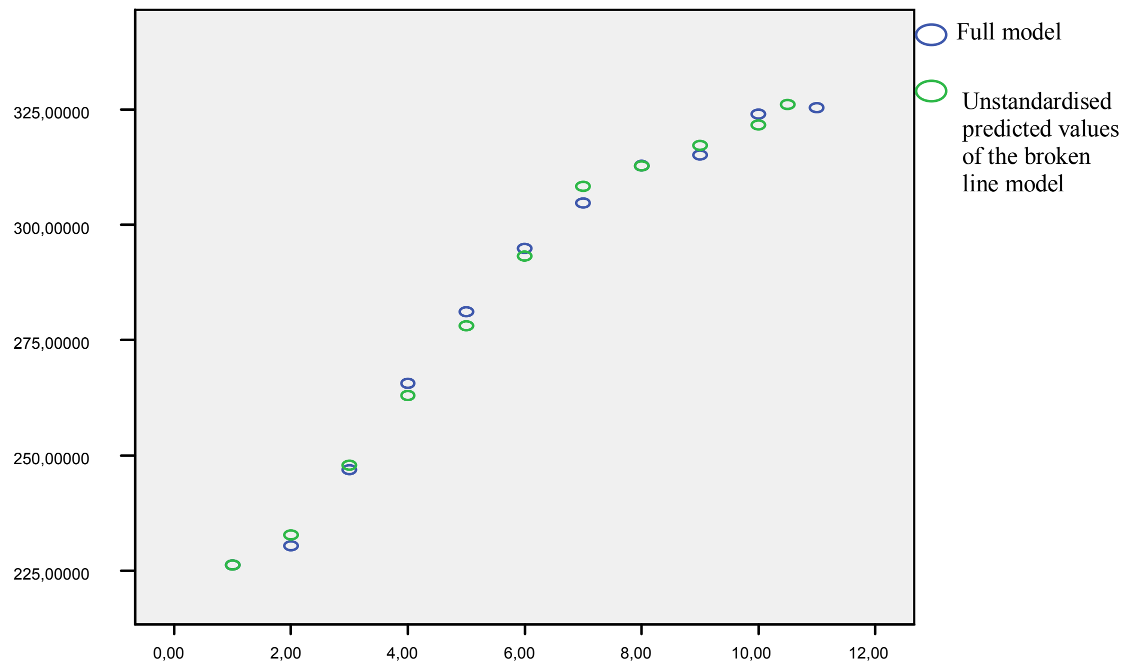

Figure 3.6. Plot of estimated values (mean model)

The figure above presents graphically how the predicted values for the mean of the broken line model abstain from the initial model for the mean.

<u>3.2.2 Fit of the joint- mean covariance model</u>

As we have already mentioned we will use the BIC criteria to distinguish the models with the best fit among those presented in the former sessions. We have decided after specific statistical analysis that when we assume a common vector μ for the mean in the joint- mean covariance model the generalised autoregressive parameters and the innovation variances are cubic functions. On the other hand, when we assume a broken line model this may change. We have applied several combinations of linear functions for the generalised autoregressive parameters and the innovation variances to the broken line model with the intention to find a model that will give satisfactory results. The following table shows the -2loglikelihood and the BIC criteria of several models that we have applied to our data. The criteria in the table are displayed in smaller-is-better forms. The first line shows the result from the first model where then mean is a common vector, the generalised autoregressive parameters and the innovation variances are cubic functions. The sum of parameters in this case is 19 since we have 11 parameters for the mean, 4 for the autoregressive coefficients and 4 for the innovation variances. After the application of many combinations of potential functions we detect no model that fits better than the former ($3.2.1). All the broken line models that we have applied in our data have bad fit since they result in larger values of -2loglikelihood and bigger values of BIC, even though we have reduced the number of parameters. We conclude that the full model for the mean is though without a great difference a better one.

Mean	**Autoregressive coef.**	**Innovation variance**	**-2loglikelihood**	**BIC**
11	*4*	*4*	*1.570,1*	*54,4908*
4	4	4	1.628,6	55,6471
4	3	3	1.635,4	55,6471
4	3	5	1.630,8	55,7205
4	5	5	1.602,1	54,9906
4	5	3	1.610,1	55,0305
4	4	5	1.623,5	55,5905
4	5	4	1.602,7	54,8972
4	6	4	1.580,0	54,2539

Table 3.8 Number of parameters and fit of the joint mean-covariance model

3.2.3 Mixed effects model

Our purpose in this session is to find the essential mixed effects model for our data where depending variable is the weight of the cattle (y_{ij}).

3.2.3.1 Random intercept model

Firstly we assume a mixed effects model with a random intercept u_{0i}

$$y_{ij} = \mu + t_j + u_{0i} + e_{ij}$$

This is the random-intercepts model with only time as a regressor (factor), where time t_j is treated using incremental values from j=1 to 11 and i=1, 2 ,.., 30. We proceed under the assumption that the structure of the covariance matrix is an identity matrix of the form $\sigma^2 \begin{bmatrix} 1 & 0 & . & 0 \\ 0 & 1 & 0 & . \\ . & 0 & . & 0 \\ 0 & . & 0 & 1 \end{bmatrix}$. This structure has constant variance and there is assumed to be no correlation between any elements. The characteristics of the model is thoroughly presented below

-2 Log Likelihood	2412,628
Schwarz's Bayesian Criterion (BIC)	81,8948

Table 3.9. Information Criteria

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	time	11		10
Random Effects	Intercept	1	Identity	1
Residual				1
Total		13		13

Table 3.10. Model Dimension

Source	Numerator df	Denominator df	F	p-value
Intercept	1	30	11.330E4	,000
time	10	300,000	647,989	,000

Table 3.11.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	325,4667	3,0078	47,87	108,205	,000	319,41	331,514
[time=1]	-99,266	2,0502	300	-48,416	,000	-103,3	-95,231
[time=2]	-95,133	2,0502	300	-46,400	,000	-99,16	-91,098
[time=3]	-78,600	2,0502	300	-38,336	,000	-82,63	-74,565
[time=4]	-59,833	2,0502	300	-29,183	,000	-63,86	-55,798
[time=5]	-44,3	2,0502	300	-21,607	,000	-48,33	-40,265
[time=6]	-30,6	2,0502	300	-14,925	,000	-34,63	-26,565
[time=7]	-20,7333	2,0502	300	-10,112	,000	-24,76	-16,698
[time=8]	-12,6	2,0502	300	-6,146	,000	-16,63	-8,565
[time=9]	-10,33	2,0502	300	-5,040	,000	-14,36	-6,29
[time=10]	-1,40	2,0502	300	-,683	,495	-5,43	2,63
[time=11]	0[a]	0	.	.	.	.	.
a. This parameter is set to zero because it is redundant.							

Table 3.12. Estimates of Fixed Effects

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		63,054	5,148	12,247	,000	53,729	73,997
Intercept	Variance	208,36	55,28	3,769	,000	123,875	350,472

Table 3.13. Estimates of Covariance Parameters

We proceed with another mixed effects model with a random intercept and an identity covariance matrix with the difference that time is a covariate, meaning that the model is of the form

$$y_{ij} = \beta_0 + \beta_{1ij} + u_{0i} + e_{ij}$$

where i=1, 2 ,.., 30 and j=1, 2, …, 11.The results of this analysis are shown in the following tables.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	time	1		1
Random Effects	Intercept	1	Identity	1
Residual				1
Total		3		4

Table 3.14. Model Dimension

-2 Log Likelihood	2596
Schwarz's Bayesian Criterion (BIC)	86,9868

Table 3.15. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		116,2740	9,493	12,247	,000	99,079	136,453
Intercept	Variance	203,5246	55,28	3,681	,000	119,507	346,608

Table 3.16. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	41,844	5.624	,000
time	1	300	3.377	,000

Table 3.17.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	217,896	2,9055	41,84	74,992	,000	212,032235	223,760968
time	11,152	,1919	300	58,109	,000	10,774548	11,529901

Table 3.18. Estimates of Fixed Effects

3.2.4 Broken lines in mixed effects model

In this session we present mixed effects models as before but not directly to our data. We use the broken line design matrix that we have already applied to the joint mean-covariance model. Given that we use the design matrix of the broken line model we are seeking for the mixed effects model that will have the best fit.

3.2.4.1 Random intercept model

First of all we apply a random intercept model of the form

$$y_{ij} = \beta_0 + \beta_1 x_j + \beta_2 y_j + \beta_3 z_j + u_{0i} + e_{ij}$$

for i=1, 2 ,.., 30 and j=1, 2, …, 11 where x, y, z are covariates and u_{0i} is a random intercept. The covariance matrix is assumed to have an identity structure. The results of this analysis are shown in the following tables.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	x	1		1
	y	1		1
	z	1		1
Random Effects	Intercept	1	Identity	1
Residual				1

Table 3.19. Model Dimension

-2 Log Likelihood	2435
Schwarz's Bayesian Criterion (BIC)	81,8469

Table 3.20. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		68,00289	5,552	12,247	,000	57,946	79,804
Intercept	Variance	207,9129	55,28	3,761	,000	123,468	350,110

Table 3.21. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	144,683	2.815	,000
x	1	300,000	12,434	,000
y	1	300,000	2.145	,000
z	1	300,000	114,714	,000

Table 3.22.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	219,6851	4,140555	144,68	53,057	,000	211,501	227,868
x	6,514902	1,847560	300	3,526	,000	2,879	10,150
y	15,12549	,326606	300	46,311	,000	14,482	15,768
z	4,459216	,416342	300	10,710	,000	3,639	5,278

Table 3.23. Estimates of Fixed Effects

3.2.4.2 Random intercept and slope model

The following model is a random intercept and slope model.

$$y_{ij} = \beta_0 + \beta_1 x_j + \beta_2 y_j + \beta_3 z_j + u_{0i} + u_{1i} x_j + u_{2i} y_j + u_{3i} z_j + e_{ij}$$

for i=1, 2 ,.., 30 and j=1, 2, …, 11 where x, y, z are covariates, u_{0i} is a random intercept and u_{1i}, u_{2i}, u_{3i}, are the random slope coefficients of the x, y, z respectively. The covariance matrix is also assumed to have an identity structure.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	x	1		1
	y	1		1
	z	1		1
Random Effects	Intercept + x + y + z	4	Identity	1
Residual				1
Total		8		6

Table 3.24. Model Dimension

-2 Log Likelihood	2220
Schwarz's Bayesian Criterion (BIC)	74,6802

Table 3.25. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		21,05	1,963	10,724	,000	17,539	25,278
Intercept + x + y + z	Variance	13,9	2,321	5,988	,000	10,020	19,283

Table 3.26. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	297,138	1,331E4	,000
x	1	325,311	27,918	,000
y	1	82,807	460,898	,000
z	1	90,043	38,460	,000

Table 3.27.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	219,6	1,904	297,138	115,370	,000	215,937	223,432
x	6,514	1,233	325,311	5,284	,000	4,089	8,940
y	15,12	,704	82,807	21,469	,000	13,724	16,526
z	4,459	,719	90,043	6,202	,000	3,030	5,887

Table 3.28. Estimates of Fixed Effects

The model presented below is also a random intercept and slope model.

$$y_{ij} = \beta_0 + \beta_1 x_j + \beta_2 y_j + \beta_3 z_j + u_{0i} + u_{1i} x_j + u_{2i} y_j + u_{3i} z_j + e_{ij}$$

for i=1, 2 ,.., 30 and j=1, 2, …, 11 where x, y, z are covariates, u_{0i} is a random intercept and u_{1i}, u_{2i}, u_{3i}, are the random slope coefficients of the x, y, z respectively. In this case the covariance structure is assumed to be diagonal.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	x	1		1
	y	1		1
	z	1		1
Random Effects	Intercept + x + y + z	4	Diagonal	4
Residual				1
Total		8		9

Table 3.29. Model Dimension

-2 Log Likelihood	2196
Schwarz's Bayesian Criterion (BIC)	74,2204

Table 3.30. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		20,1136	1,893	10,620	,000	16,724	24,190
Intercept + x + y + z	Var: Intercept	60,5942	2,734	2,216	,027	25,023	146,725
	Var: x	21,6505	9,261	2,338	,019	9,361	50,070
	Var: y	6,37083	1,844	3,455	,001	3,612	11,235
	Var: z	6,54617	2,071	3,160	,002	3,520	12,172

Table 3.31. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	58,279	9.574	,000
x	1	59,224	24,515	,000
y	1	31,472	937,963	,000
z	1	30,522	73,790	,000

Table 3.32.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	219,685	2,245	58,279	97,848	,000	215,191	224,178
x	6,5149	1,315	59,224	4,951	,000	3,882	9,147
y	15,1254	,493	31,472	30,626	,000	14,118	16,132
z	4,4592	,519	30,522	8,590	,000	3,399	5,518

Table 3.33. Estimates of Fixed Effects

Finally the last model is a random intercept and slope model with the difference that even if the random part of the model is the same as before the fixed part has changed. This model assumes only a fixed integer and not fixed slopes for the three covariates. On the contrary there are only random slopes assumed

$$y_{ij} = \beta_0 + u_{0i} + u_{1i}x_j + u_{2i}y_j + u_{3i}z_j + e_{ij}$$

for i=1, 2 ,.., 30 and j=1, 2, …, 11 where x, y, z are the covariates, u_{0i} is a random intercept and u_{1i}, u_{2i}, u_{3i}, are the random slope coefficients of the x, y, z respectively. The covariance matrix is assumed to have a diagonal structure.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
Random Effects	Intercept + x + y + z	4	Diagonal	4
Residual				1
Total		5		6

Table 3.34. Model Dimension

-2 Log Likelihood	2360
Schwarz's Bayesian Criterion (BIC)	79,3469

Table 3.35. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		20,75	2,0433	10,156	,000	17,109	25,168
Intercept + x + y + z	Var: Intercept	51,72	29,010	1,783	,075	17,229	155,278
	Var: x	46,38	17,219	2,694	,007	22,406	96,023
	Var: y	245,18	63,561	3,857	,000	147,513	407,527
	Var: z	25,737	7,0558	3,648	,000	15,038	44,047

Table 3.36. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	28,319	13.640	,000

Table 3.37.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval Lower Bound	Upper Bound
Intercept	224,04	1,918	28,319	116,79	,000	220,118	227,972

Table 3.38. Estimates of Fixed Effects

3.2.5 Fit of the mixed-effects models

As we can figure out from the results of the former sessions the mixed-effects models is comparatively to the joint mean covariance model a very bad fit. The table below is a summary of the mixed –effects models we have applied. The BIC values of these models vary from 74,220 to 86,986 while the BIC value of the joint – mean covariance model with the worst fit was 55,7205. The same behavior we observe with the loglikelihood. In this case the smallest value of the -2loglikelihood is 2.196 while the biggest value observed in the joint – mean covariance models is 1.635,4. This indicates that the joint- mean covariance model excels significantly the mixed effect model. Furthermore, we can notice from the table that the broken line models seem to fit better in our data as they also seemed to fit better in the previous case of the joint mean covariance model. That indicates that the broken line model seems to give better results in both models we have applied.

Covariance structure	Fixed effects	Random effects	Sum of param.	-2 loglikel.	BIC
Identity	Intercept & time effect	Intercept	13	2.412	81,894
Identity	Intercept & time effect	Intercept	4	2.596	86,986
Identity	Broken lines	Intercept	6	2.435	81,847
Identity	Broken lines	Broken lines	6	2.220	74,680
Diagonal	Broken lines	Broken lines	9	2.196	74,220
Diagonal	Intercept	Broken lines	6	2.360	79,347

Table 3.39. Fit of the mixed-effects models

3.2.6 Analysis of the best model

According to the results of former section we have concluded that the model that fits better to the data of group A is the model that assumes a common vector μ for the mean and a cubic function for the generalized autoregressive parameters and the innovation variances as it is shown below.

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \varepsilon_{t,j,d} \Rightarrow$$

$$\widehat{\phi}_{t,j} = -1{,}1014 + 0{,}5455(t-j) - 0{,}0845(t-j)^2 + 0{,}0043(t-j)^3 + \varepsilon_{t,j,d}$$

$$\log \widehat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \varepsilon_{t,v} \Rightarrow$$

$$\log \widehat{\sigma}_t^2 = 5{,}1911 - 0{,}7143t + 0{,}0817t^2 - 0{,}0023t^3 + \varepsilon_{t,v}$$

This model concentrates the largest value of the loglikelihood and the smallest value of BIC criteria. We observe from the following scatterplots that the distribution of the standardized residuals does not seem to follow a specific pattern which is an indicator of the good fit of our model. Furthermore, the graph below and the Kolmogorov –Smirnov test (Z= 0.631), with p-value=0.821>0.05, show that the residuals follow the normal distribution.

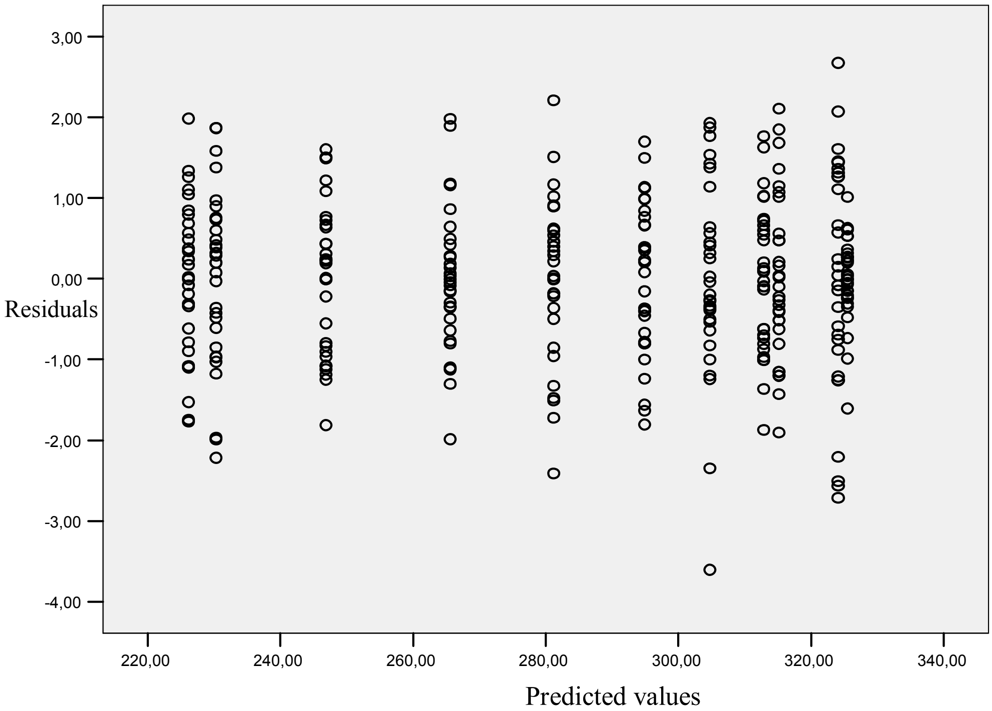

Figure 3.7. Plot of the standardized residuals

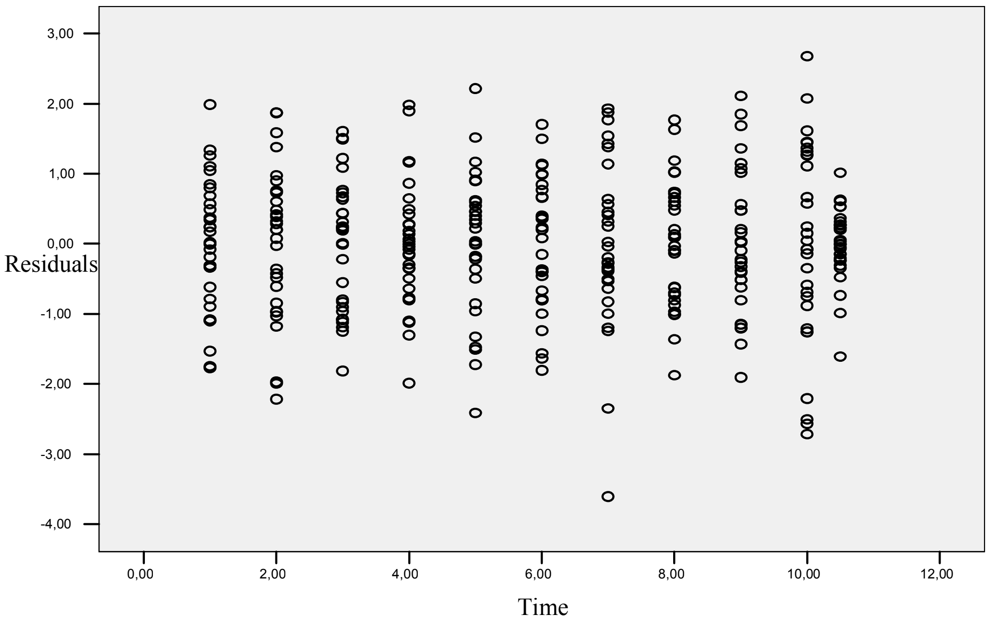

Figure 3.8. Plot of residuals through time

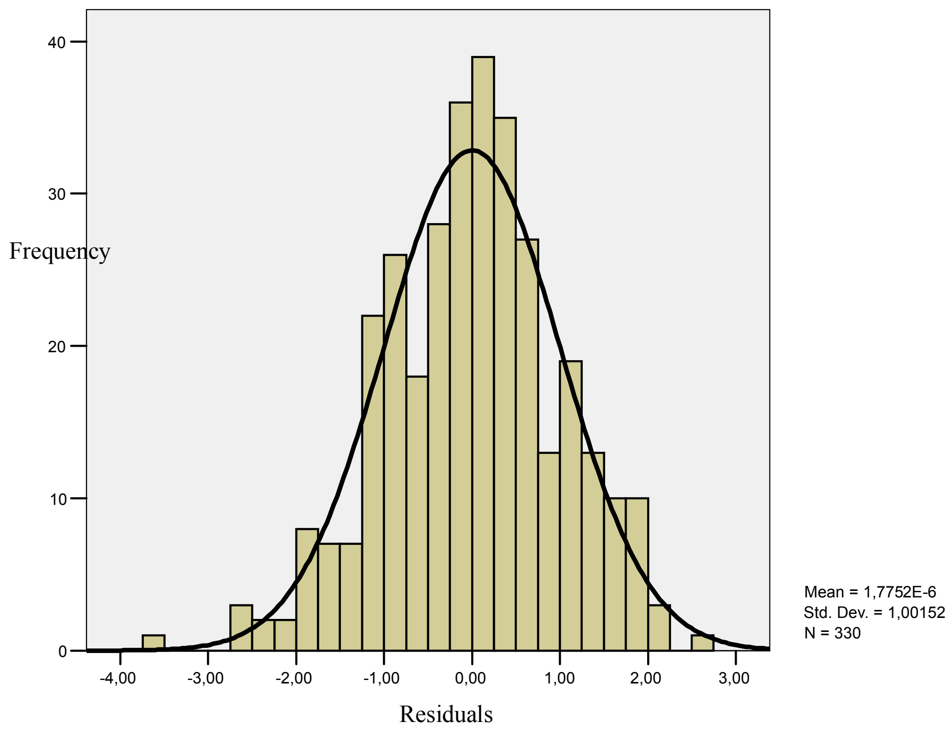

Figure 3.9. Distribution of the standardized residuals

3.3 Cattle analysis of group B

In group B we will proceed in the same way as we have worked with group A. We start again with the joint mean covariance model. The structure of group B is more complicated in modelling than that of group A. Firstly we will find potential models for the innovation variances and for the autoregressive parameters. Afterwards we will apply the joint mean covariance model under the assumption of these models and a common vector μ for the mean. We will continue with the application of the broken lines model for the mean with several models for the innovation variances and for the autoregressive parameters.

We will also study the fit of mixed effects models to our data with weight as a depending variable and time as an independent. Finally, we will try to find a satisfactory mixed effects model under the assumption of a new design matrix of broken lines.

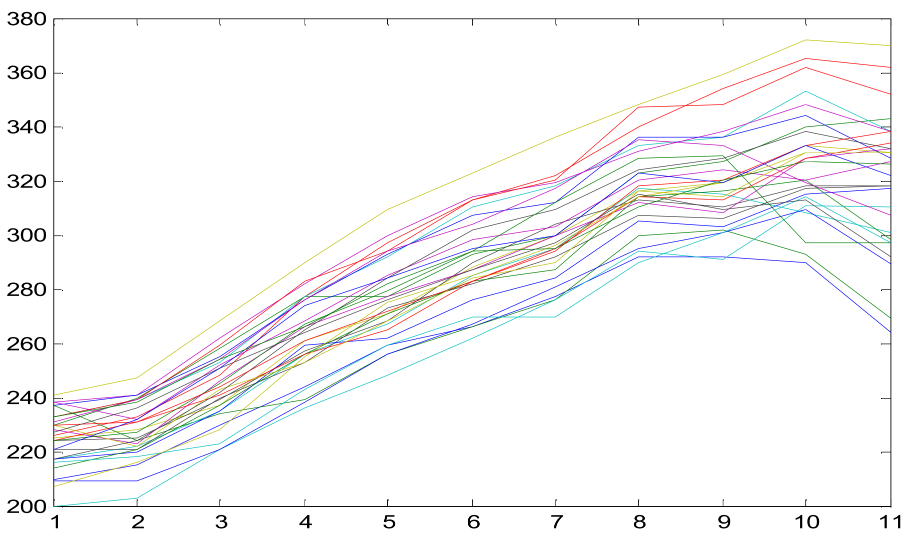

Figure 3.10. Plot of weight through time

In the figure above we assess that the weights tend to increase over time with exception the last time point when we observe a massive decline. The subjects also seem to have overall the same behaviour.

3.3.1 Joint mean- covariance model for group B

We are seeking for the treatment group B with m=30 animals, the joint-mean covariance model

$$\mu_t = m(x_t,\beta), \quad \log\sigma_t^2=u(z_t,\lambda), \quad \varphi_{t,j}=d(z_{t,j},\gamma)$$

(Pourahmadi, 1999) that fits best. Firstly, for the mean model we assume a common mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_{11})$ where $\mu_t = \frac{1}{30}\sum_{i=1}^{30} y_{it}$,t=1,2,…,11. Then, we apply the Cholesky Decomposition in the 11x11 sample covariance matrix $S=\frac{1}{30}\sum_{i=1}^{30}(y_i-\mu)(y_i-\mu)'$ to find the T,D matrices that give us the $\varphi_{t,j}$'s generalized autoregressive sample parameters and σ_t^2's sample innovation variances.

3.3.1.1. Model for the generalized autoregressive parameters

The plot $\varphi_{t,j}$ versus time lags j=1,2,…,t-1 does not really enable us to figure out the degree of the linear function that will suit to the autoregressive parameters. This plot also shows us the linear, quadratic and cubic function that indicate the need of a high degree polynomial.

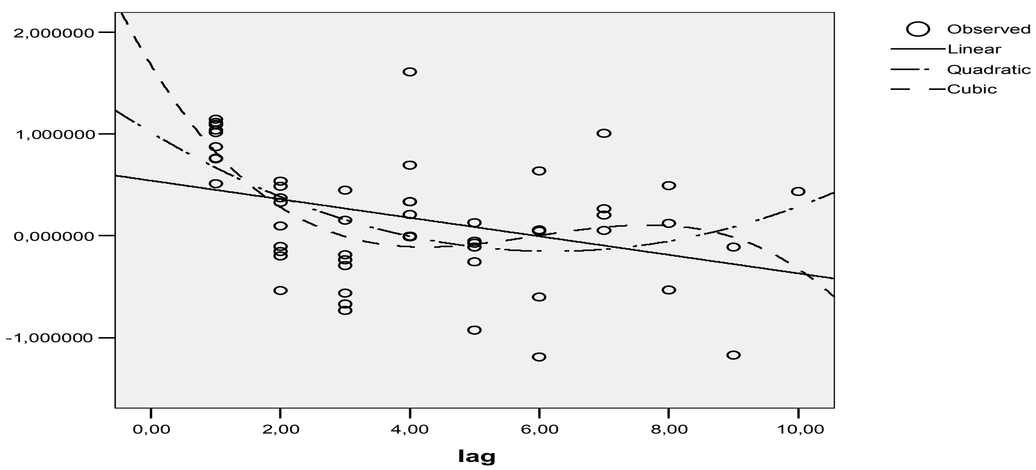

Figure 3.11 . Plot of the sample generalized autoregressive parameters through time with linear, quadratic and cubic function

We applied to our parameters eight linear models to find out which is the most suitable according to the R^2. The models are cited below.

a) The linear polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \varepsilon_{t,j,d} \Rightarrow \widehat{\phi}_{t,j} = 0.542 - 0.091(t-j) + \varepsilon_{t,j,d}$$

b) The quadratic polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \varepsilon_{t,j,d}$$

$$\Rightarrow \widehat{\phi}_{t,j} = 1{,}017 - 0{,}377(t-j) + 0{,}03(t-j)^2 + \varepsilon_{t,j,d}$$

c) The cubic polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \varepsilon_{t,j,d} \Rightarrow$$

$$\widehat{\phi}_{t,j} = 1{,}680 - 1{,}036(t-j) + 0{,}19(t-j)^2 - 0{,}011(t-j)^3 + \varepsilon_{t,j,d}$$

d) The 4th degree polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \gamma_5(t-j)^4 + \varepsilon_{t,j,d} \Rightarrow$$

$$\widehat{\phi}_{t,j} = 2{,}774 - 2{,}54(t-j) + 0{,}7832(t-j)^2 - 0{,}0976(t-j)^3 + 0{,}004208(t-j)^4 + \varepsilon_{t,j,d}$$

e) The 5th degree polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \gamma_5(t-j)^4 + \gamma_6(t-j)^5 + \varepsilon_{t,j,d} \Rightarrow$$

$$\widehat{\phi}_{t,j} = 2{,}971 - 2{,}877(t-j) + 0{,}9697(t-j)^2 - 0{,}1414(t-j)^3 + 0{,}008774(t-j)^4 - 0{,}0001732(t-j)^5 + \varepsilon_{t,j,d}$$

f) The 6th degree polynomial

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \gamma_5(t-j)^4 + \gamma_6(t-j)^5 + \gamma_7(t-j)^6 + \varepsilon_{t,j,d}$$

$$\Rightarrow$$

$$\widehat{\phi}_{t,j} = 7{,}612 - 12{,}27(t-j) + 7{,}59(t-j)^2 - 2{,}306(t-j)^3 + 0{,}3653(t-j)^4 - 0{,}02884(t-j)^5 +$$
$$+ 0{,}0008938(t-j)^6 + \varepsilon_{t,j,d}$$

MODEL	**R^2**	**R^2 ADJUSTED**
LINEAR	0,136	0,120
QUADRATIC	0,231	0,202
CUBIC	0,301	0,259
4^{th} DEGREE	0,361	0,310
5^{th} DEGREE	0,361	0,296
6^{th} DEGREE	0,418	0,346

Table 3.40. R^2 values for the sample generalized autoregressive parameters

From the table above we observe that the R^2 and the R^2 adjusted seem to enlarge while we add terms but none of these values indicate that there is a model that fits adequately to the parameters. Because of this we will not be restricted to one model as we have done with group A but we will try several combinations of models to find out which maximizes the loglikelihood.

3.3.1.2. Model for the innovation parameters

We assume that our data is independent and we are trying to find a linear model that will be the best fit in order to use its parameters as initial values From the plot of σ_t^2 versus time t=1,2,…,11 we observe that the logarithms of the sample innovation parameters could be a linear function and specifically a quadratic function.

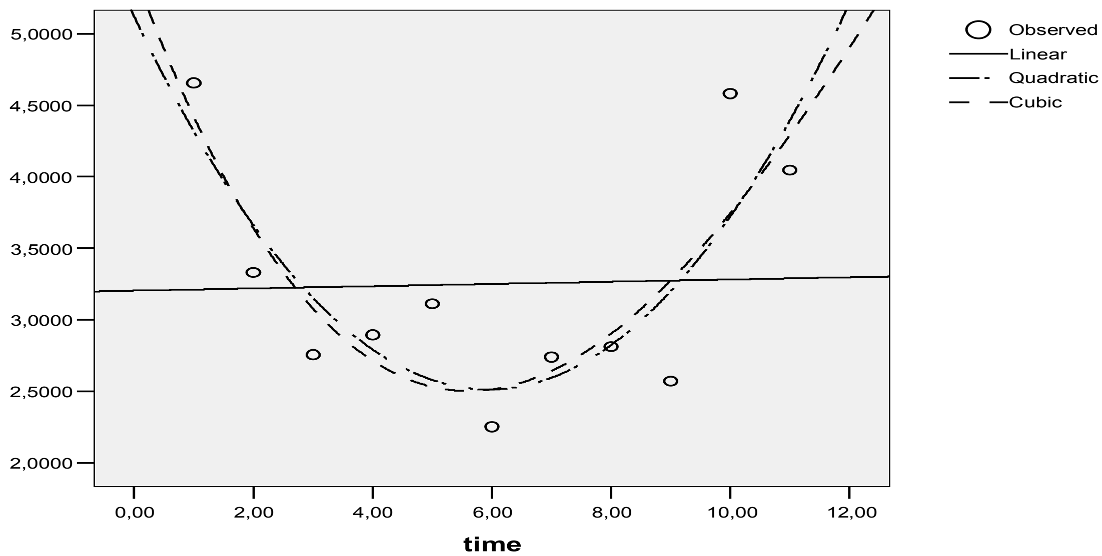

Figure 3.12. Plot of the sample innovation parameters through time with linear, quadratic and cubic function

We applied to our parameters four linear models to find out which is the most suitable according to the F test and which has the biggest value of R^2. The models with the corresponding analysis of variance are cited below.

a) The linear polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t \varepsilon_{t,v} \Rightarrow \log \hat{\sigma}_t^2 = 3{,}204 + 0{,}008t + \varepsilon_{t,v}$$

b) The quadratic polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 \varepsilon_{t,v} \Rightarrow \log \hat{\sigma}_t^2 = 5{,}121 - 0{,}877t + 0{,}074t^2 + \varepsilon_{t,v}$$

c) The cubic polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \varepsilon_{t,v} \Rightarrow$$
$$\log \hat{\sigma}_t^2 = 5{,}431 - 1{,}133t + 0{,}125t^2 - 0{,}003t^3 + \varepsilon_{t,v}$$

d) The 4th degree polynomial

$$\log \hat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \lambda_4 t^3 + \lambda_5 t^4 + \varepsilon_{t,v} \Rightarrow$$
$$\log \hat{\sigma}_t^2 = 5{,}974 - 1{,}788t + 0{,}3465t^2 - 0{,}03069t^3 + 0{,}00116t^4 + \varepsilon_{t,v}$$

MODEL		**SUM OF SQUARES**	**df**	**MEAN SQUARES**	**F**	**p-value**
LINEAR	REGRESSION	0,007	1	0,007	0,009	0,926
	RESIDUAL	6,685	9	0,743		
	TOTAL	6,692	10			
QUADRATIC	REGRESSION	4,674	2	2,337	9,263	0,008
	RESIDUAL	2,018	8	0,252		
	TOTAL	6,692	10			
CUBIC	REGRESSION	4,723	3	1,574	5,599	0,028
	RESIDUAL	1,968	7	0,281		
	TOTAL	6,692	10			
4th DEGREE	REGRESSION	4,779	4	1,194	3,754	0,000
	RESIDUAL	1,913	6	0,318		
	TOTAL	6,692	10			

Table 3.41. Anova for the sample innovation parameters

MODEL	**R^2**	**R^2 ADJUSTED**
LINEAR	0,001	-0,110
QUADRATIC	0,698	0,623
CUBIC	0,706	0,580
4^{th} DEGREE	0,714	0,523

Table 3.42. R^2 values for the sample innovation parameters

From the statistical analysis presented in the tables above we conclude that the quadratic and the linear functions have the better fit to the innovation variances and especially the quadratic polynomial that has the biggest value of $R^2_{adjusted} = 0,623$. Furthermore according to the linear regression analysis it has the largest F- value = 9,263 from all the other models with p-value = 0,008 < 0,01

3.3.1.3 Broken line model

In this session we will replace the common mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_{11})$ where $\mu_t = \frac{1}{30}\sum_{i=1}^{30} y_{it}$,t=1,2,…,11 , that we assumed before, with a broken line model in order to minimise the number of the mean parameters. In this case we observe from the plot of the weights of the cattle through time that we can represent our data with four broken lines. We see in the plot three turning points that distinguish on time points 2 and 7 and 10.

Let $y = \begin{cases} 1 & t=1 \\ 2 & t>2, \end{cases}$ $z = \begin{cases} 0 & t\leq 2 \\ t-2 & t>2 \end{cases}$, $d = \begin{cases} 0 & t\leq 7 \\ t-7 & 11>t>7 \\ 3 & t=11 \end{cases}$ and $q = \begin{cases} 0 & t<11 \\ 1 & t=11 \end{cases}$

We construct a new design matrix for the mean that consist of the above 4 new variables y, z, d, q and we study how this model fits to our longitudinal data.

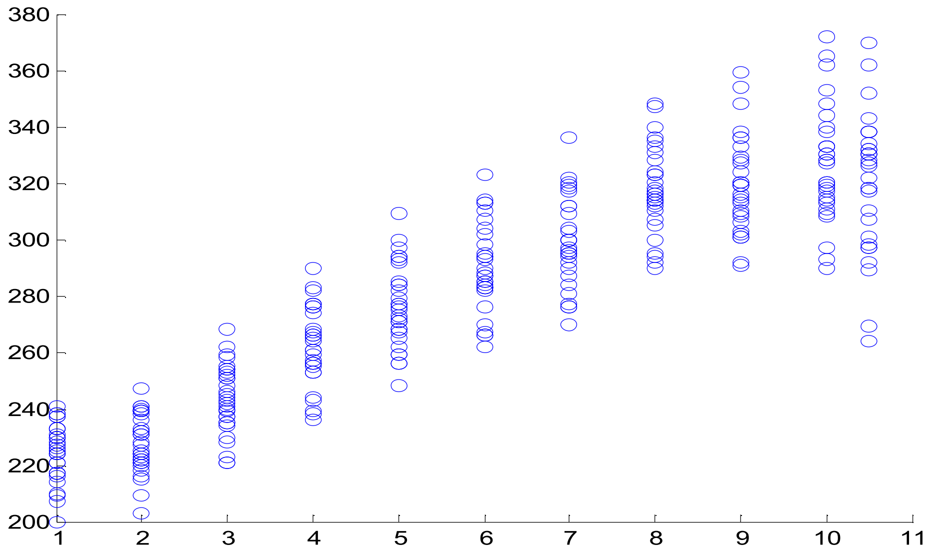

Figure 3.13. Plot of weight through time

We suppose the linear model

$$y_{ij} = \beta_0 + \beta_1 y_{ij} + \beta_2 z_{ij} + \beta_3 d_{ij} + \beta_4 q_{ij} + \varepsilon_{ij}$$

for the mean where i=1, 2 ,.., 30 and j=1, 2, …, 11and $\varepsilon_{ij} \sim N(0, \sigma^2)$

From the tables below we find out that according to regression linear analysis our broken line model has the form

$y_{ij} = \beta_0 + \beta_1 y_{ij} + \beta_2 z_{ij} + \beta_3 d_{ij} + \beta_4 q_{ij} + \varepsilon_{ij} \Rightarrow y_{ij} = 219{,}640 + 4{,}960 y_{ij} + 15{,}037 z_{ij} + 7{,}81 d_{ij} - 7{,}873 q_{ij} + \varepsilon_{ij}$

The analysis of variance indicates that this new design matrix gives satisfying results and the new model for the mean explain a significant percentage of the variance since R^2 = 0,840 and $R^2_{adjusted}$=0,838 values are close to 1. The p-value of the F-test is 0,000 meaning that the F-value= 425,697 belongs to the $F_{4,325}$ distribution. On the other hand we observe from the scatterplot of the studentized residuals and the unstandardised predicted values that the residuals seem to follow a specific pattern that indicates a bad fit for our model.

.

Model		Unstandardized Coefficients		Standardized Coefficients	t	p-value
		B	Std. Error	Beta		
1	(Constant)	219,640	6,214		35,347	,000
	y	4,960	3,596	,036	1,379	,169
	z	15,037	,654	,754	22,986	,000
	d	7,810	1,133	,234	6,893	,000
	q	-7,873	3,789	-,057	-2,078	,039

Table 3.43. Estimated coefficients and t-test for broken line model

R	R Square	Adjusted R Square	Std. Error of the Estimate
,916	,840	,838	16,02405

Table 3.44. Model summary for broken line model

Model		Sum of Squares	df	Mean Square	F	p-value
	Regression	437224,848	4	109306,212	425,697	,000
	Residual	83450,307	325	256,770		
	Total	520675,155	329			

Table 3.45. Anova for broken line model

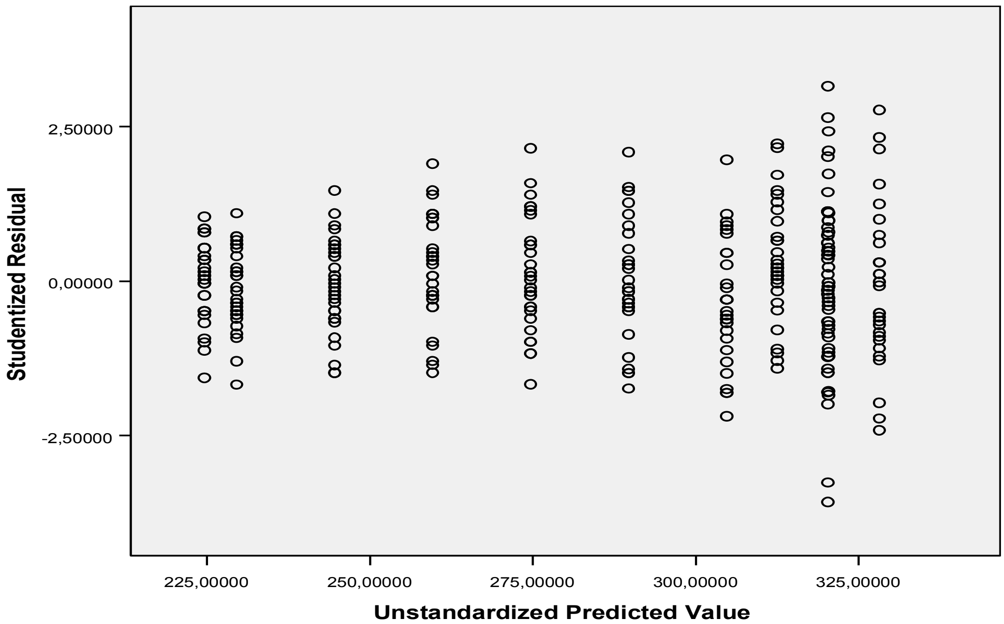

Figure 3.14. Studentized residuals of broken line

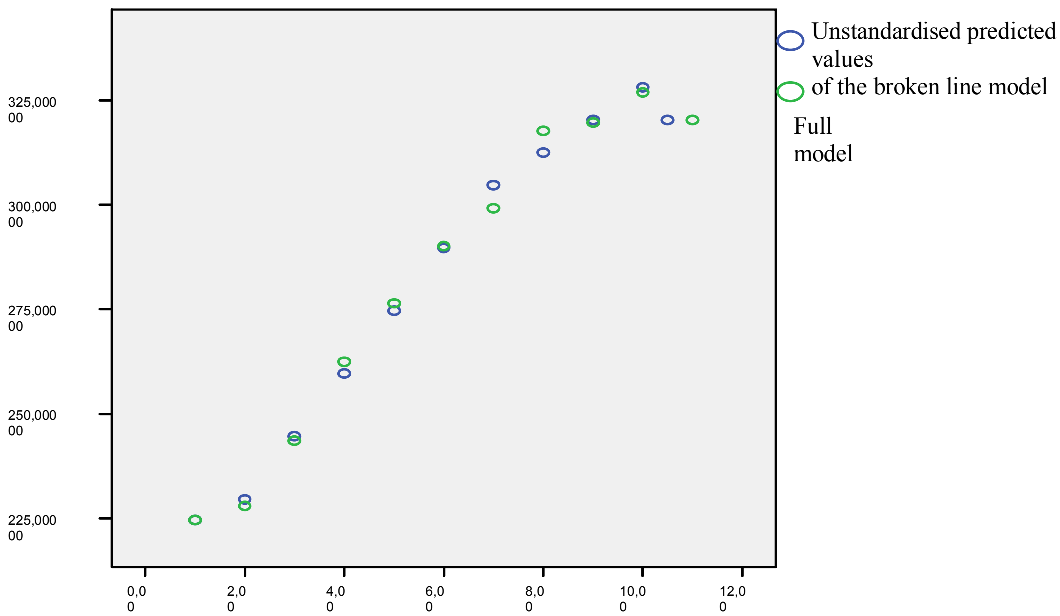

Figure 3.15. Plot of estimated values (mean model)

The figure above presents the predicted values of the broken lines model and the means of the cattle weights at each time point meaning the model for the mean initially assumed. We assess graphically that there is no great difference between the two models since the observations are close to each other.

3.3.2 Fit of the joint- mean covariance model

We judge the fit of our models according to the BIC criteria that depends on the number of parameters included in the model and the value of the loglikelihood. The joint- mean covariance model as we have already explained is partitioned of three submodels. The parameters included are p for the mean model, q_1 for the autoregressive parameters and q_2 for the innovation variances. Our goal is to find the combination of the three submodels, with the least possible parameters, that maximises the loglikelihood. The table below presents the number of parameters that were included in every submodel and the values of the -2loglikelihood and the BIC criteria that resulted from the analysis.

The models where we assume a common vector μ for the mean are those with 11 mean parameters and the models with 5 mean parameters are those where the broken lines model is assumed for the mean. As far as we can see from the table the models with 11 parameters for the mean have a better fit, meaning that the full model must be preferred. Specifically, the model with the smallest

-2loglikelihood value is that where a 5^{th} degree polynomial was assumed for the autoregressive coefficients and a 7^{th} degree polynomial was assumed for the innovation variances. However, the smallest BIC value was presented in the model where a 5^{th} degree polynomial was also assumed for the autoregressive coefficients but a cubic polynomial was assumed for the innovation variances, which has fewer parameters and that may have resulted in a smaller BIC value.

Mean Parameters	**Autoregressive coef.**	**Innovation variance**	**-2loglikelihood**	**BIC**
11	4	3	1.504,6	52,1941
11	4	4	1.503,0	52,2541
11	4	5	1.497,2	52,1741
11	4	6	1.530,5	53,3975
11	4	7	1.494,1	52,2975
11	5	3	1.496,7	52,0441
11	5	4	1.496,2	52,1408
11	5	5	1.492,4	52,1275
11	5	6	1.518,2	53,1009
11	5	7	1.489,9	52,2709
5	5	7	1.624,6	56,0807
5	5	6	1.626,4	56,0273
5	5	5	1.629,4	56,0139
5	5	4	1.631,0	55,9539
5	5	3	1.636,7	56,0305
5	4	7	1.625,6	56,0006
5	4	6	1.628,0	55,9673
5	4	5	1.631,2	55,9606
5	4	4	1.632,6	55,8939
5	4	3	1.637,8	55,9538

Table 3.46. Number of parameters and fit of the joint mean-covariance model

3.3.3 Mixed effects model

Our purpose in this session is to find the essential mixed effects model for our data where depending variable is the weight of the cattle (y_{ij}).

3.3.3.1 Random intercept model

We start with a random intercept model with time as a depending variable

$$y_{ij} = \mu + t_j + u_{0i} + e_{ij}$$

where u_{0i} is the random intercept, j=1 to 11 and i=1, 2 ,.., 30. We proceed under the assumption that the structure of the covariance matrix is an identity matrix of the form $\sigma^2 \begin{bmatrix} 1 & 0 & . & 0 \\ 0 & 1 & 0 & . \\ . & 0 & . & 0 \\ 0 & . & 0 & 1 \end{bmatrix}$. This structure has constant variance. There is assumed to be no correlation between any elements. The characteristics of the model is thoroughly presented below.

-2 Log Likelihood	2.406
Schwarz's Bayesian Criterion (BIC)	81,67

Table 3.47. Information Criteria

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	time	11		10
Random Effects	Intercept	1	Identity	1
Residual				1
Total		13		13

Table 3.48. Model Dimension

Source	Numerator df	Denominator df	F	p-value
Intercept	1	30,000	12.660	,000
time	10	300	702,660	,000

Table 3.49.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	320,3	2,864	50,27	111,817	,000	314,547	326,052
[time=1,00]	-95,7	2,041	300	-46,868	,000	-99,718	-91,6817
[time=2,00]	-92,4	2,041	300	-45,252	,000	-96,418	-88,3817
[time=3,00]	-76,76	2,041	300	-37,596	,000	-80,7849	-72,7484
[time=4,00]	-57,80	2,041	300	-28,307	,000	-61,8182	-53,7817
[time=5,00]	-43,86	2,041	300	-21,483	,000	-47,8849	-39,8484
[time=6,00]	-30,16	2,041	300	-14,774	,000	-34,1849	-26,1484
[time=7,00]	-21,06	2,041	300	-10,317	,000	-25,0849	-17,0484
[time=8,00]	-2,633	2,041	300	-1,290	,198	-6,6515	1,3849
[time=9,00]	-,633333	2,041	300	-,310	,757	-4,6515	3,3849
[time=10,0]	6,633	2,041	300	3,249	,001	2,615	10,6515
[time=11,0]	0[a]	0	.	.	.	.	.

Table 3.50. Estimates of Fixed Effects

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		62,53986	5,106358	12,247	,000	53,291	73,393
Intercept	Variance	183,6194	48,88051	3,756	,000	108,974	309,39

Table 3.51. Estimates of Covariance Parameters

We continue with another mixed effects model with a random intercept and an identity covariance matrix with the difference that time is a covariate not a factor, meaning that the model is of the form

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{0i} + e_{ij}$$

where i=1, 2 ,.., 30 and j=1, 2, …, 11.The results of this analysis are shown in the following tables

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	time	1		1
Random Effects	Intercept	1	Identity	1
Residual				1
Total		3		4

Table 3.52. Model Dimension

-2 Log Likelihood	2.600
Schwarz's Bayesian Criterion (BIC)	87,00

Table 3.53. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		119,07	9,722	12,247	,000	101,461	139,734
Intercept	Variance	178,48	48,88	3,651	,000	104,338	305,307

Table 3.54. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	43,867	5.974	,000
time	1	300,000	3.548	,000

Table 3.55.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	213,74	2,765	43,86	77,29	,000	208,167	219,315
time	11,568	,1942	300	59,56	,000	11,186	11,95

Table 3.56. Estimates of Fixed Effects

3.3.4 Broken lines in mixed effects model

In this session we present mixed effects models as before but not directly to our data. We use the broken line design matrix that we have already applied to the joint mean-covariance model. Given that we have assumed the design matrix of the broken line model we are seeking for the mixed effects model that will have the best fit.

3.3.4.1 Random intercept model

First of all we apply a random intercept model of the form

$$y_{ij} = \beta_0 + \beta_1 y_{ij} + \beta_2 z_{ij} + \beta_3 d_{ij} + \beta_4 q_{ij} + u_{0i} + e_{ij}$$

for i=1, 2 ,.., 30 and j=1, 2, …, 11 where y, z, d, q are covariates and u_{0i} is a random intercept. The covariance matrix is assumed to have an identity structure.The results of this analysis are shown in the following tables.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	y	1		1
	z	1		1
	d	1		1
	q	1		1
Random Effects	Intercept	1	Identity	1
Residual				1
Total		6		7

Table 3.57. Model Dimension

-2 Log Likelihood	2.440
Schwarz's Bayesian Criterion (BIC)	82,12

Table 3.58. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		69,932	5,7099	12,247	,000	59,59	82,06
Intercept	Variance	182,94	48,881	3,743	,000	108,36	308,85

Table 3.59. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value.
Intercept	1	164,186	2.904	,000
y	1	300,000	6,984	,009
z	1	300,000	1.940	,000
d	1	300,000	174,443	,000
q	1	300,000	15,850	,000

Table 3.60.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	219,64	4,076	164,18	53,886	,000	211,591	227,688
y	4,960	1,8768	300	2,643	,009	1,266	8,653
z	15,036	,3414	300	44,044	,000	14,364	15,708
d	7,8100	,5913	300	13,208	,000	6,646	8,973
q	-7,873	1,977	300	-3,981	,000	-11,765	-3,981

Table 3.61. Estimates of Fixed Effects

3.3.4.2 Random intercept and slope model

The following model is a random intercept and slope model.

$$y_{ij} = \beta_0 + \beta_1 y_{ij} + \beta_2 z_{ij} + \beta_3 d_{ij} + \beta_4 q_{ij} + u_{0i} + u_{1i} y_j + u_{2i} z_j + u_{3i} d_j + u_{4i} q_j + e_{ij}$$

for i=1, 2 ,.., 30 and j=1, 2, …, 11 where y, z, d, q are covariates, u_{0i} is a random intercept and u_{1i}, u_{2i}, u_{3i}, u_{4i} are the random slope coefficients of the y, z, d, q respectively. The covariance matrix is also assumed to have an identity structure.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	y	1		1
	z	1		1
	d	1		1
	q	1		1
Random Effects	Intercept + y + z + d + q	5	Identity	1
Residual				1
Total		10		7

Table 3.62. Model Dimension

-2 Log Likelihood	2.316
Schwarz's Bayesian Criterion (BIC)	77,99

Table 3.63. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		30,331	2,9786	10,183	,000	25,021	36,769
Intercept + y + z + d + q	Variance	15,008	2,7439	5,470	,000	10,488	21,475

Table 3.64. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	267,606	9.532	,000
y	1	329,002	12,130	,001
z	1	74,040	410,474	,000
d	1	106,196	93,561	,000
q	1	326,066	28,220	,000

Table 3.65.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	219,64	2,2497	267,6	97,629	,000	215,210	224,069
y	4,96	1,4241	329,002	3,483	,001	2,158	7,761
z	15,036	,7421	74,04	20,260	,000	13,557	16,515
d	7,81	,8074	106,19	9,673	,000	6,209	9,410
q	-7,873	1,482	326,06	-5,312	,000	-10,789	-4,957

Table 3.66. Estimates of Fixed Effects

The model presented below is also a random intercept and slope model.

$$y_{ij} = \beta_0 + \beta_1 y_{ij} + \beta_2 z_{ij} + \beta_3 d_{ij} + \beta_4 q_{ij} + u_{0i} + u_{1i} y_j + u_{2i} z_j + u_{3i} d_j + u_{4i} q_j + e_{ij}$$

for i=1, 2 ,.., 30 and j=1, 2, …, 11 where y, z, d, q are covariates, u_{0i} is a random intercept and u_{1i}, u_{2i}, u_{3i}, u_{4i} are the random slope coefficients of the y, z, d, q respectively. In this case the covariance structure is assumed to be diagonal.

		Number of Levels	Covariance Structure	Number of Parameters
Fixed Effects	Intercept	1		1
	y	1		1
	z	1		1
	d	1		1
	q	1		1
Random Effects	Intercept + y + z + d + q	5	Diagonal	5
Residual				1
Total		10		11

Table 3.67. Model Dimension

-2 Log Likelihood	2.276
Schwarz's Bayesian Criterion (BIC)	77,11

Table 3.68. Information Criteria

Parameter		Estimate	Std. Error	Wald Z	p-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		26,26	2,537	10,349	,000	21,731	31,738
Intercept + y + z + d + q	Var: Intercept	65,118	30,37	2,144	,032	26,101	162,460
	Var: y	13,17	8,533	1,544	,123	3,700	46,892
	Var: z	3,194	1,096	2,914	,004	1,630	6,259
	Var: d	9690	3,382	2,865	,004	4,888	19,207
	Var: q	99,09	36,56	2,710	,007	48,081	204,249

Table 3.69. Estimates of Covariance Parameters

Source	Numerator df	Denominator df	F	p-value
Intercept	1	65,486	7.883	,000
y	1	63,813	13,963	,000
z	1	33,772	1.505	,000
d	1	32,366	134,256	,000
q	1	30,640	12,990	,001

Table 3.70.Tests of Fixed Effects

Parameter	Estimate	Std. Error	df	t	p-value	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	219,64	2,473	65,486	88,786	,000	214,700	224,579
y	4,96	1,327	63,813	3,737	,000	2,308	7,611
z	15,03	,387	33,772	38,791	,000	14,248	15,824
d	7,81	,674	32,366	11,587	,000	6,437	9,182
q	-7,873	2,184	30,640	-3,604	,001	-12,330	-3,415

Table 3.71. Estimates of Fixed Effects

3.3.5 Fit of the mixed-effects models

In order to conclude the results of the former sessions we present the following table. As we can see from the table the BIC values are much higher from those of the joint mean-covariance models as we have already observed for group A. This means that the joint mean-covariance models fit much more adequately to all our data from the mixed-effects models. However among the mixed -effects models the smallest BIC values are concentrated in the broken lines models. Finally, the best model is the last one where a diagonal covariance structure is assumed.

Covariance structure	Fixed effects	Random effects	Sum of param.	-2loglikel.	BIC
Identity	Intercept & time effect	Intercept	13	2.406	81,6739
Identity	Intercept & time effect	Intercept	3	2.600	87,0068
Identity	Broken lines	Intercept	7	2.440	82,1269
Identity	Broken lines	Broken lines	7	2.316	77,9936
Diagonal	Broken lines	Broken lines	11	2.276	77,1138

Table 3.72. Fit of the mixed-effects models

3.3.6 Analysis of the best model

The analysis fulfilled in former sections proves that the best model for group B is, likewise group A, the one that assumes a common vector μ for the mean and in this case a 4th degree polynomial for the generalized autoregressive parameters and a quadratic function for the innovation variances.

The estimated parameters for these functions are the following.

$$\widehat{\phi}_{t,j} = \gamma_1 + \gamma_2(t-j) + \gamma_3(t-j)^2 + \gamma_4(t-j)^3 + \gamma_5(t-j)^4 + \varepsilon_{t,j,d} \Rightarrow$$

$$\widehat{\phi}_{t,j} = -1{,}7165 + 1{,}0883(t-j) - 0{,}2094(t-j)^2 + 0{,}0124(t-j)^3 - 0{,}000(t-j)^4 + \varepsilon_{t,j,d}$$

$$\log\widehat{\sigma}_t^2 = \lambda_1 + \lambda_2 t + \lambda_3 t^2 + \varepsilon_{t,v} \Rightarrow \log\widehat{\sigma}_t^2 = 5{,}2879 - 0{,}9402t + 0{,}0854t^2 + \varepsilon_{t,v}$$

This model concentrates the largest value of the loglikelihood and the smallest BIC= 52,04. We observe from the following scatterplots that the standardized residuals seem to be distributed randomly which indicates a good fit for our model. Furthermore, the graph below and the Kolmogorov – Smirnov test (Z= 0,714) with p-value=0,687>0.05 show that the residuals follow the normal distribution.

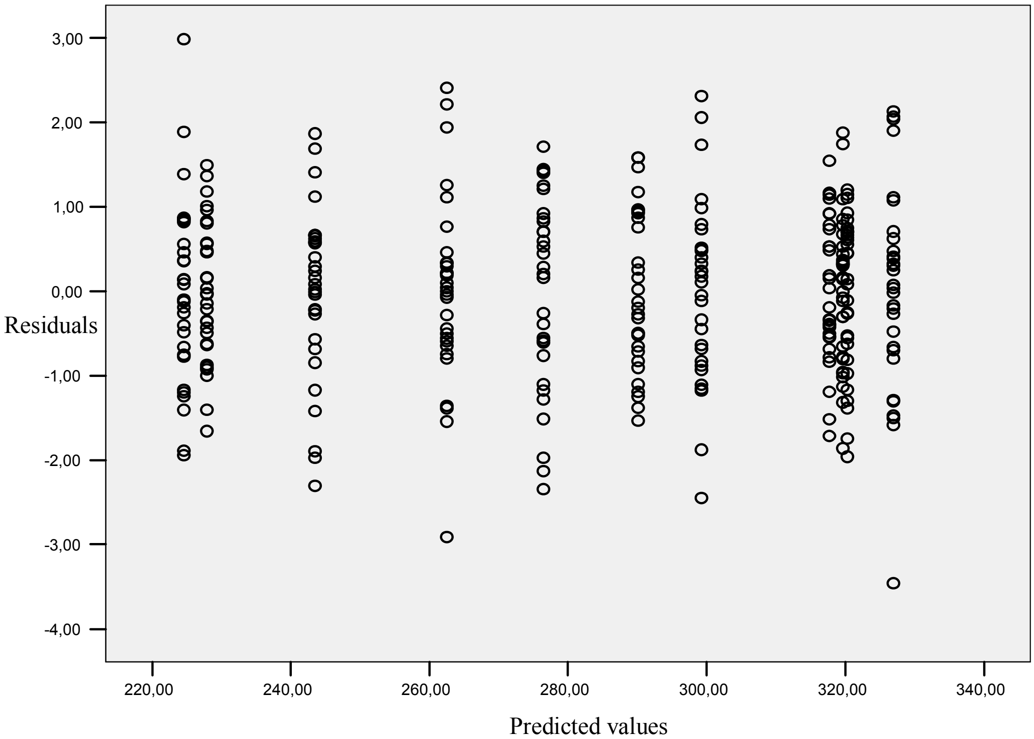

Figure 3.16. Plot of standardized residuals

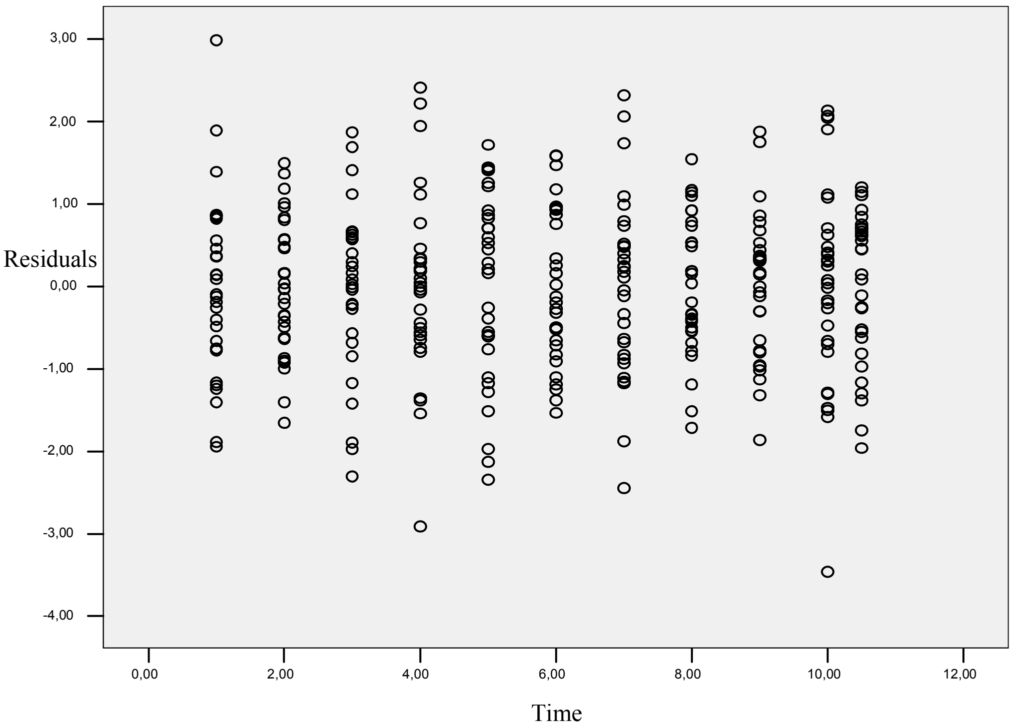

Figure 3.17. Plot of residuals through time

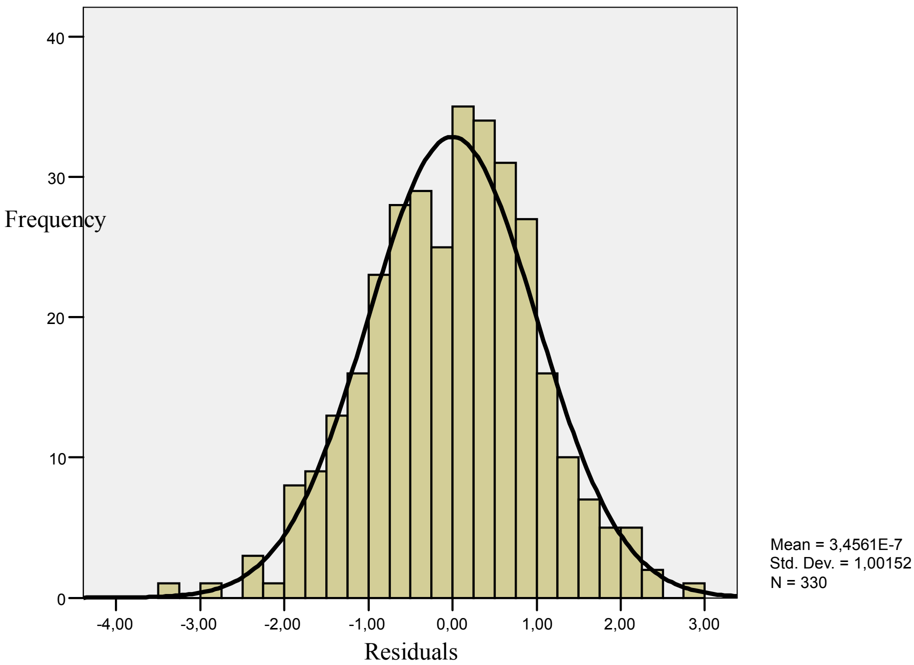

Figure 3.18. Distribution of the standardized residuals

APPENDIX A

Table A1- Kenward (1987) cattle data

SUBJECT	TIME	WEIGHT	GROUP	SUBJECT	TIME	WEIGHT	GROUP
1	1	210	2	31	1	233	1
1	2	215	2	31	2	224	1
1	3	230	2	31	3	245	1
1	4	244	2	31	4	258	1
1	5	259	2	31	5	271	1
1	6	266	2	31	6	287	1
1	7	277	2	31	7	287	1
1	8	292	2	31	8	287	1
1	9	292	2	31	9	290	1
1	10	290	2	31	10	293	1
1	11	264	2	31	11	297	1
2	1	230	2	32	1	231	1
2	2	240	2	32	2	238	1
2	3	258	2	32	3	260	1
2	4	277	2	32	4	273	1
2	5	277	2	32	5	290	1
2	6	293	2	32	6	300	1
2	7	300	2	32	7	311	1
2	8	323	2	32	8	313	1
2	9	327	2	32	9	317	1
2	10	340	2	32	10	321	1
2	11	343	2	32	11	326	1
3	1	226	2	33	1	232	1
3	2	233	2	33	2	237	1
3	3	248	2	33	3	245	1
3	4	277	2	33	4	265	1
3	5	297	2	33	5	285	1
3	6	313	2	33	6	298	1
3	7	322	2	33	7	304	1
3	8	340	2	33	8	319	1
3	9	354	2	33	9	317	1
3	10	365	2	33	10	334	1
3	11	362	2	33	11	329	1
4	1	233	2	34	1	239	1
4	2	239	2	34	2	246	1
4	3	253	2	34	3	268	1
4	4	277	2	34	4	288	1
4	5	292	2	34	5	308	1
4	6	310	2	34	6	309	1
4	7	318	2	34	7	327	1
4	8	333	2	34	8	324	1
4	9	336	2	34	9	327	1
4	10	353	2	34	10	336	1
4	11	338	2	34	11	341	1
5	1	238	2	35	1	215	1
5	2	241	2	35	2	216	1
5	3	262	2	35	3	239	1
5	4	282	2	35	4	264	1
5	5	300	2	35	5	282	1
5	6	314	2	35	6	299	1
5	7	319	2	35	7	307	1

Table A1 (continued)

SUBJECT	TIME	WEIGHT	GROUP	SUBJECT	TIME	WEIGHT	GROUP
5	8	331	2	35	8	321	1
5	9	338	2	35	9	328	1
5	10	348	2	35	10	332	1
5	11	338	2	35	11	337	1
6	1	225	2	36	1	236	1
6	2	228	2	36	2	226	1
6	3	237	2	36	3	242	1
6	4	261	2	36	4	255	1
6	5	271	2	36	5	263	1
6	6	288	2	36	6	277	1
6	7	300	2	36	7	290	1
6	8	316	2	36	8	299	1
6	9	319	2	36	9	300	1
6	10	333	2	36	10	308	1
6	11	330	2	36	11	310	1
7	1	224	2	37	1	219	1
7	2	225	2	37	2	229	1
7	3	239	2	37	3	246	1
7	4	257	2	37	4	265	1
7	5	268	2	37	5	279	1
7	6	290	2	37	6	292	1
7	7	304	2	37	7	299	1
7	8	313	2	37	8	299	1
7	9	310	2	37	9	298	1
7	10	318	2	37	10	300	1
7	11	318	2	37	11	290	1
8	1	237	2	38	1	231	1
8	2	241	2	38	2	245	1
8	3	255	2	38	3	270	1
8	4	276	2	38	4	292	1
8	5	293	2	38	5	302	1
8	6	307	2	38	6	321	1
8	7	312	2	38	7	322	1
8	8	336	2	38	8	334	1
8	9	336	2	38	9	323	1
8	10	344	2	38	10	337	1
8	11	328	2	38	11	337	1
9	1	237	2	39	1	230	1
9	2	224	2	39	2	228	1
9	3	234	2	39	3	243	1
9	4	239	2	39	4	255	1
9	5	256	2	39	5	272	1
9	6	266	2	39	6	276	1
9	7	276	2	39	7	277	1
9	8	300	2	39	8	289	1
9	9	302	2	39	9	289	1
9	10	293	2	39	10	300	1
9	11	269	2	39	11	303	1
10	1	233	2	40	1	232	1
10	2	239	2	40	2	240	1
10	3	259	2	40	3	247	1

Table A1 (continued)

SUBJECT	TIME	WEIGHT	GROUP	SUBJECT	TIME	WEIGHT	GROUP
10	4	283	2	40	4	263	1
10	5	294	2	40	5	275	1
10	6	313	2	40	6	286	1
10	7	320	2	40	7	294	1
10	8	347	2	40	8	302	1
10	9	348	2	40	9	308	1
10	10	362	2	40	10	319	1
10	11	352	2	40	11	326	1
11	1	217	2	41	1	234	1
11	2	222	2	41	2	237	1
11	3	235	2	41	3	259	1
11	4	256	2	41	4	289	1
11	5	267	2	41	5	311	1
11	6	285	2	41	6	324	1
11	7	295	2	41	7	342	1
11	8	317	2	41	8	347	1
11	9	315	2	41	9	355	1
11	10	308	2	41	10	368	1
11	11	301	2	41	11	368	1
12	1	228	2	42	1	237	1
12	2	223	2	42	2	235	1
12	3	246	2	42	3	258	1
12	4	266	2	42	4	263	1
12	5	277	2	42	5	282	1
12	6	287	2	42	6	304	1
12	7	300	2	42	7	318	1
12	8	312	2	42	8	327	1
12	9	308	2	42	9	336	1
12	10	328	2	42	10	349	1
12	11	332	2	42	11	353	1
13	1	241	2	43	1	229	1
13	2	247	2	43	2	234	1
13	3	268	2	43	3	254	1
13	4	290	2	43	4	276	1
13	5	309	2	43	5	294	1
13	6	323	2	43	6	315	1
13	7	336	2	43	7	323	1
13	8	348	2	43	8	341	1
13	9	359	2	43	9	346	1
13	10	372	2	43	10	352	1
13	11	370	2	43	11	357	1
14	1	221	2	44	1	220	1
14	2	221	2	44	2	227	1
14	3	240	2	44	3	248	1
14	4	253	2	44	4	273	1
14	5	273	2	44	5	290	1
14	6	282	2	44	6	308	1
14	7	292	2	44	7	322	1
14	8	307	2	44	8	326	1
14	9	306	2	44	9	330	1
14	10	317	2	44	10	342	1

Table A1 (continued)

SUBJECT	TIME	WEIGHT	GROUP	SUBJECT	TIME	WEIGHT	GROUP
14	11	318	2	44	11	343	1
15	1	217	2	45	1	232	1
15	2	220	2	45	2	241	1
15	3	235	2	45	3	255	1
15	4	259	2	45	4	276	1
15	5	262	2	45	5	293	1
15	6	276	2	45	6	309	1
15	7	284	2	45	7	310	1
15	8	305	2	45	8	330	1
15	9	303	2	45	9	326	1
15	10	315	2	45	10	329	1
15	11	317	2	45	11	330	1
16	1	214	2	46	1	210	1
16	2	221	2	46	2	225	1
16	3	237	2	46	3	242	1
16	4	256	2	46	4	260	1
16	5	271	2	46	5	272	1
16	6	283	2	46	6	277	1
16	7	287	2	46	7	273	1
16	8	314	2	46	8	295	1
16	9	316	2	46	9	292	1
16	10	320	2	46	10	305	1
16	11	298	2	46	11	306	1
17	1	224	2	47	1	229	1
17	2	231	2	47	2	241	1
17	3	241	2	47	3	252	1
17	4	256	2	47	4	265	1
17	5	265	2	47	5	274	1
17	6	283	2	47	6	285	1
17	7	295	2	47	7	303	1
17	8	314	2	47	8	308	1
17	9	313	2	47	9	315	1
17	10	328	2	47	10	328	1
17	11	334	2	47	11	328	1
18	1	200	2	48	1	204	1
18	2	203	2	48	2	198	1
18	3	221	2	48	3	217	1
18	4	236	2	48	4	233	1
18	5	248	2	48	5	251	1
18	6	262	2	48	6	258	1
18	7	276	2	48	7	272	1
18	8	294	2	48	8	283	1
18	9	291	2	48	9	279	1
18	10	311	2	48	10	295	1
18	11	310	2	48	11	298	1
19	1	238	2	49	1	220	1
19	2	232	2	49	2	221	1
19	3	252	2	49	3	236	1

Table A1 (continued)

SUBJECT	TIME	WEIGHT	GROUP	SUBJECT	TIME	WEIGHT	GROUP
19	4	268	2	49	4	260	1
19	5	285	2	49	5	274	1
19	6	298	2	49	6	295	1
19	7	303	2	49	7	300	1
19	8	320	2	49	8	301	1
19	9	324	2	49	9	310	1
19	10	320	2	49	10	318	1
19	11	327	2	49	11	316	1
20	1	230	2	50	1	233	1
20	2	222	2	50	2	234	1
20	3	243	2	50	3	250	1
20	4	253	2	50	4	268	1
20	5	268	2	50	5	280	1
20	6	284	2	50	6	298	1
20	7	290	2	50	7	308	1
20	8	316	2	50	8	319	1
20	9	314	2	50	9	318	1
20	10	330	2	50	10	336	1
20	11	330	2	50	11	333	1
21	1	217	2	51	1	234	1
21	2	224	2	51	2	234	1
21	3	242	2	51	3	254	1
21	4	265	2	51	4	274	1
21	5	284	2	51	5	294	1
21	6	302	2	51	6	306	1
21	7	309	2	51	7	318	1
21	8	324	2	51	8	334	1
21	9	328	2	51	9	343	1
21	10	338	2	51	10	349	1
21	11	332	2	51	11	350	1
22	1	209	2	52	1	200	1
22	2	209	2	52	2	207	1
22	3	221	2	52	3	217	1
22	4	238	2	52	4	238	1
22	5	256	2	52	5	252	1
22	6	267	2	52	6	267	1
22	7	281	2	52	7	284	1
22	8	295	2	52	8	282	1
22	9	301	2	52	9	282	1
22	10	309	2	52	10	284	1
22	11	289	2	52	11	288	1
23	1	224	2	53	1	220	1
23	2	227	2	53	2	213	1
23	3	245	2	53	3	229	1
23	4	267	2	53	4	252	1
23	5	279	2	53	5	254	1
23	6	294	2	53	6	273	1
23	7	312	2	53	7	293	1
23	8	328	2	53	8	289	1
23	9	329	2	53	9	294	1

Table A1 (continued)

SUBJECT	*TIME*	*WEIGHT*	*GROUP*	*SUBJECT*	*TIME*	*WEIGHT*	*GROUP*
23	10	297	2	53	10	292	1
23	11	297	2	53	11	298	1
24	1	230	2	54	1	225	1
24	2	231	2	54	2	239	1
24	3	244	2	54	3	254	1
24	4	261	2	54	4	269	1
24	5	272	2	54	5	289	1
24	6	283	2	54	6	308	1
24	7	294	2	54	7	313	1
24	8	318	2	54	8	324	1
24	9	320	2	54	9	327	1
24	10	333	2	54	10	347	1
24	11	338	2	54	11	344	1
25	1	216	2	55	1	236	1
25	2	218	2	55	2	245	1
25	3	223	2	55	3	257	1
25	4	243	2	55	4	271	1
25	5	259	2	55	5	294	1
25	6	270	2	55	6	307	1
25	7	270	2	55	7	317	1
25	8	290	2	55	8	327	1
25	9	301	2	55	9	328	1
25	10	314	2	55	10	328	1
25	11	297	2	55	11	325	1
26	1	231	2	56	1	231	1
26	2	239	2	56	2	231	1
26	3	254	2	56	3	237	1
26	4	276	2	56	4	261	1
26	5	294	2	56	5	274	1
26	6	304	2	56	6	285	1
26	7	317	2	56	7	291	1
26	8	335	2	56	8	301	1
26	9	333	2	56	9	307	1
26	10	319	2	56	10	315	1
26	11	307	2	56	11	320	1
27	1	207	2	57	1	208	1
27	2	216	2	57	2	211	1
27	3	228	2	57	3	238	1
27	4	255	2	57	4	254	1
27	5	275	2	57	5	267	1
27	6	285	2	57	6	287	1
27	7	296	2	57	7	306	1
27	8	314	2	57	8	312	1
27	9	319	2	57	9	320	1
27	10	330	2	57	10	337	1
27	11	330	2	57	11	338	1
28	1	227	2	58	1	232	1

Table A1 (continued)

SUBJECT	*TIME*	*WEIGHT*	*GROUP*	*SUBJECT*	*TIME*	*WEIGHT*	*GROUP*
28	2	236	2	58	2	248	1
28	3	251	2	58	3	261	1
28	4	264	2	58	4	285	1
28	5	276	2	58	5	292	1
28	6	287	2	58	6	307	1
28	7	297	2	58	7	312	1
28	8	315	2	58	8	323	1
28	9	309	2	58	9	318	1
28	10	313	2	58	10	328	1
28	11	292	2	58	11	329	1
29	1	221	2	59	1	233	1
29	2	232	2	59	2	241	1
29	3	251	2	59	3	252	1
29	4	274	2	59	4	273	1
29	5	284	2	59	5	301	1
29	6	295	2	59	6	316	1
29	7	300	2	59	7	332	1
29	8	323	2	59	8	336	1
29	9	319	2	59	9	339	1
29	10	333	2	59	10	348	1
29	11	322	2	59	11	345	1
30	1	233	2	60	1	221	1
30	2	238	2	60	2	219	1
30	3	254	2	60	3	231	1
30	4	266	2	60	4	251	1
30	5	282	2	60	5	270	1
30	6	294	2	60	6	272	1
30	7	295	2	60	7	287	1
30	8	310	2	60	8	294	1
30	9	320	2	60	9	292	1
30	10	327	2	60	10	292	1
30	11	326	2	60	11	299	1

APPENDIX B

Program 1

Cattle data are entered in a format which uses 4 variables. The first denotes the subject, the second denotes the group, the third denotes the time and the fourth is the observation (weight).

The next lines in Matlab transpose observed data (4th column) into 11 columns as a

typical longitudinal data set and compute the empirical covariance matrix

```
load cattle.dat
k=0;
y0=[];
for i=1:60
   y1=[];
   for j=1:11
      y1=[y1 cattle(k+j,4)];
   end
   y0=[y0;cattle(k+1,2) y1];
   k=k+11;
end
% The data used are those of group 1
y=y0((y0(:,1)==1),2:12);
[n,q]=size(y);
% empirical covariance matrix and modified cholesky decomposition
sig1=((y-ones(n,1)*mean(y))'*(y-ones(n,1)*mean(y)))/(n-1);
[t1,d1]=modchol(sig1);
```

Program 2

Calculates in Matlab the modified Cholesky decomposition of the empirical covariance matrix of observation (sigma)

```
function [t,d]=modchol(sigma)
r=chol(sigma);
d=diag(diag(r))*diag(diag(r));
t=inv((inv(diag(diag(r)))*r)');
```

Program 3

Cattle data analysis in Matlab (for one group) through Pourahmadi's method and Pan, McKenzie. This program finds the -2loglikelihood value, the estimated generalized autoregressive parameters and the estimated innovation parameters given the initial values (gamma, delta).

```
y=cattle(:,4);
[n0 p0]=size(y);
n=n0/11;
x=[(cattle(:,3)==1) (cattle(:,3)==2) (cattle(:,3)==3) (cattle(:,3)==4)
(cattle(:,3)==5) (cattle(:,3)==6)...
    (cattle(:,3)==7) (cattle(:,3)==8) (cattle(:,3)==9) (cattle(:,3)==10)
(cattle(:,3)==10.5)];
 [n0,p]=size(x);
beta=[226.2;230.3;246.9;265.6;281.2;294.9;304.7;312.9;315.1;324.1;325.5];
betaml=beta+0.1;
gamma=[6.7346;-1.4561;7.2275;-0.0123];
delta=[1.7434;-1.0522;0.1857;-0.01];
zeta=[];
for j=1:11
    for i=j+1:11
        zeta=[zeta;1 cattle(i,3)-cattle(j,3) (cattle(i,3)-cattle(j,3))^2 (cattle(i,3)-
cattle(j,3))^3];
    end
end
eta=[ones(11,1) cattle(1:11,3) cattle(1:11,3).^2 cattle(1:11,3).^3];
nparam=[4 4];
prm=[gamma; delta;beta];
param=zeros(sum(nparam)+p,1);
q=11;
options=optimset('MaxFunEvals',200000,'MaxIter',200000,'TolX',1e-
8,'TolFun',1e-8,'LargeScale','off');
m=0;
```

Program 3 (continued)

```
while (norm(beta-betaml)/norm(beta)>0.00001)
   m=m+1
beta=betaml;
prminit=prm(1:sum(nparam));
gamma=prm(1:nparam(1));
delta=prm(nparam(1)+1:sum(nparam));
[invsigmaml,tml,dml]=makesig(q,gamma,delta,zeta,eta);
k=0;
y0=[];
for i=1:30
      y0=[y0;y(k+(1:11))'];
      k=k+11;
end
sig=((y0-(ones(n,1)*(beta')))'*(y0-(ones(n,1)*(beta'))))/(n);
[t,d]=modchol(sig);
[prm,mxlik,exitflag,output]=fminsearch(@lik_pour,prminit,options,n0,q,bet
a,y,x,nparam,zeta,eta);
gamma=prm(1:nparam(1));
delta=prm(nparam(1)+1:sum(nparam));
[invsigmaml,tml,dml]=makesig(q,gamma,delta,zeta,eta);
tmp1=zeros(p,p);
tmp2=zeros(p,1);
betaml=zeros(p,1);
m0=0;
for i=1:n0/11
   tmp1=tmp1+x(m0+(1:11),:)'*invsigmaml*x(m0+(1:11),:);
   tmp2=tmp2+x(m0+(1:11),:)'*invsigmaml*y(m0+(1:11));
   m0=m0+11;
end
betaml=inv(tmp1)*tmp2;
end
```

Program 4

Cattle data (for one group) analysis in Matlab through Pourahmadi's method and Pan, McKenzie when we assume a broken line model for the mean. This program finds the -2loglikelihood value, the estimated generalized autoregressive parameters and the estimated innovation parameters given the initial values (gamma, delta).

```
y=cattle(:,4);
[n0 p0]=size(y);
n=n0/11;
x=matrix1
[n0,p]=size(x);
means=[226.2;230.3;246.9;265.6;281.2;294.9;304.7;312.9;315.1;324.1;325.
5];
beta=[219.685;6.515;15.125;4.459];
betaml=beta+0.1;
gamma=[1.746;-0.6785;-0.4973;0.432;-0.1365;0.02175;-
0.001722;0.00005342];
delta=[1.7434;-1.0522;0.1857;-0.01];
zeta=[];
for j=1:11
    for i=j+1:11
        zeta=[zeta;1 cattle(i,3)-cattle(j,3) (cattle(i,3)-cattle(j,3))^2 (cattle(i,3)-
cattle(j,3))^3 cattle(i,3)-cattle(j,3)^4 (cattle(i,3)-cattle(j,3))^5 (cattle(i,3)-
cattle(j,3))^6 (cattle(i,3)-cattle(j,3))^7];
    end
end
eta=[ones(11,1) cattle(1:11,3) cattle(1:11,3).^2 cattle(1:11,3).^3];
nparam=[8 4];
prm=[gamma; delta;beta];
param=zeros(sum(nparam)+p,1);
q=11;
options=optimset('MaxFunEvals',200000,'MaxIter',200000,'TolX',1e-
8,'TolFun',1e-8,'LargeScale','off');
```

Program 4 (continued)

```
m=0;
while (norm(beta-betaml)/norm(beta)>0.00001)
   m=m+1
beta=betaml;
prminit=prm(1:sum(nparam));
gamma=prm(1:nparam(1));
delta=prm(nparam(1)+1:sum(nparam));
[invsigmaml,tml,dml]=makesig(q,gamma,delta,zeta,eta);
k=0;
y0=[];
for i=1:30
      y0=[y0;y(k+(1:11))'];
      k=k+11;
end
sig=((y0-(ones(n,1)*(means')))'*(y0-(ones(n,1)*(means'))))/(n);
[t,d]=modchol(sig);
[prm,mxlik,exitflag,output]=fminsearch(@lik_pour,prminit,options,n0,q,bet
a,y,x,nparam,zeta,eta);
gamma=prm(1:nparam(1));
delta=prm(nparam(1)+1:sum(nparam));
[invsigmaml,tml,dml]=makesig(q,gamma,delta,zeta,eta);
tmp1=zeros(p,p);
tmp2=zeros(p,1);
betaml=zeros(p,1);
m0=0;
for i=1:n0/11
   tmp1=tmp1+x(m0+(1:11),:)'*invsigmaml*x(m0+(1:11),:);
   tmp2=tmp2+x(m0+(1:11),:)'*invsigmaml*y(m0+(1:11));
   m0=m0+11;
end
betaml=inv(tmp1)*tmp2;
end
```

Program 5

The following program in Matlab is a function that is used in programs 3 and 4 to calculate the value of the likelihood:

```
function lik_pour=lik_pour(param,n0,q,beta,y,x,nparam,zeta,eta)
gamma=param(1:nparam(1))
delta=param(nparam(1)+1:sum(nparam))
[n0,p]=size(x);
lik_pour=0;
[invsigma,t,d]=makesig(q,gamma,delta,zeta,eta);
k=0;
for i=1:n0/q
    r=y(k+(1:q))-x(k+(1:q),:)*beta;
    lik_pour=lik_pour+r'*invsigma*r+log(prod(1./diag(d)));
    k=k+q;
end
lik_pour
```

Program 6

The following program in Matlab is a function that is used in programs 3 and 4 to compute the estimated inverted covariance matrix given the estimated parameters:

```
function [invsigma,t,d]=makesig(q,gamma,delta,zeta,eta)
phihat=zeta*gamma;
logshat=eta*delta;
t=zeros(q);
k=0;
for j=1:q
    t(j,j)=1;
    for i=j+1:q
        k=k+1;
        t(i,j)=phihat(k);
    end
end
d=diag(1./exp(logshat));
```

BIBLIOGRAPHY

Anderson,T.W.(1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann.Statist*.1,135-41.

Anderson,T.W.(1984). An introduction to Multivariate Analysis, *2nd ed. New York: John Wiley.*

Barndorff-Nielsen, O. (1991). Likelihood theory. In *Statistical Theory and Modeling, Ed.D.V.Hinkley, N. Reid. E.J. Snell*, pp.232-64. London: Chapman and Hall

Berger,O.J.Liseo, B. & Wolpert, R. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statist.Sci*. 14, 1-28.

Brown,P.J. LE, N.D. & Zidek, J.V. (1994). Inference for a covariance matrix. *In Aspects of Uncertainty, Ed. P.R. Freeman and A.F.M. Smith,* pp.77-90. Chichester: John Wiley.

Brown, H. & Prescott, R. (1999). Applied mixed models in medicine. *Medical statistics Unit. University of Edinburgh.* Chichester : Wiley

Chiu,T.Y.M.,Leonard,T.& Tsui,K.W.(1996). The matrix-logarithm covariance model. *J. Am. Statist. Assoc*.91,198-210.

Cook, R.D.&Weisberg,S.(1983). Diagnostics for heteroscedasticity in regression. *Biometrika* 70,1-10.

Cox, D.R.& Wermuth, N. (1996). Multivariate Dependencies: Models, Analysis and Interpretation. *London: Chapman and Hall.*

Crowder, M.(1995). On the use of a working correlation matrix in using generalized linear models for repeated measures.*Biometrika* 82,407-10.

Dempster,A.P.(1972). Covariance selection. *Biometrics* 28. 157-75.

Diggle,P.J.(1988). An approach to the analysis of repeated measurements. *Biometrics* 44,959-71.

Diggle,P.J.,Liang, K.Y. & Zeger,S.L.(1994). Analysis of Longitudinal Data. *Oxford: Oxford University Press.*

Gabriel,K.R.(1962). Ante-dependence analysis of an ordered set of variables. *Ann. Math.Statist.*33,201-12.

Heagerty,P.J.& Zeger,S.L (1998). Lorelogram: A regression approach to exploring dependence in longitudinal categorical response. *J. Am.Statist. Assoc.*93,150-62.

Kenward, M.G. (1987). A method for comparing profiles of repeated measurements. *Appl.Statist.*36,296-308.

Laird, N.M. & Ware, J.J. (1982). Random effects model for longitudinal data. *Biometrics* 38,963-74.

Leonard, T.& Hsu, J.S.J.(1992). Bayesian inference for a covariance matrix. *Ann. Statist.* 20, 1669-96.

Liang, K.Y.& Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73,13-22.

Macchiavelli, R.E. & Arnold, S.F. (1994). Variable order antedependence models. *Commun. Statist. A* 23, 2683-99.

Martinez, Wendy, L., Angel, R. (2002) Computational statistics handbook with MATLAB. *Boca Raton: Chapman & Hall.*

McCullagh, P. & Nelder, J.A. (1989). Generalized linear models, *2nd ed. London: Chapman & Hall.*

Muñoz, A.,Carey, V.,Schouten, J.P., Segal, M. & Rosner, B. (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 48, 733-42.

Newton, H.J. (1988). TIMESLAB:A Time Series Analysis Laboratory. *Pacific Grove, CA: Wadsworth & Brooks/Cole.*

Pan, J. & Mackenzie,G. (2003). On modeling mean-covariance structures in longitudinal studies. *Biometrika* 90,1,pp.239-244.

Pinheiro, J.D. & Bates, D.M. (1996). Uncostrained parameterizations for variance covariance matrices. *Statist. Comp.* 6, 289-96.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Uncostrained parameterization.*Biometrika* 86,677-90

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika* 87, 425-35

Rao C.R.(1987). Prediction of future observations in growth curve models. Statist.Sci. 2,434-71.

Smyth, G.K. (1989). Generalized Linear Models with varying dispersion. *J.R. Statist.Soc. B* 51,47-60.

Tukey, J.W. (1961). Curves as parameters, and touch estimation. In *Proc.* 4^{th} *Berkeley Symp. Math. Statist. Prob.,1,Ed. J. Neyman* ,pp.681-94. Berkeley: University of California Press.

Verbyla,A.P.(1993). Modelling variance heterogeneity: Residual maximum likelihood and diagnostics. *J.R. Statist. Soc*. B 55, 493-508.

Zimmerman, D.L. & Núñez-Antón, V. (1997). Structured antedependence models for longitudinal data. In Modeling Longitudinal and Spatially Correlated Data: Methods, applications, and Future Directions, *Springer Lecture Notes in Statistics, No. 122, Ed.T.G. Gregoire et al., pp.63-76. New York: Springer- Verlag.*

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ