

Linear Mixed Models for Longitudinal Data with Nonrandom Dropouts

Ahmed M. Gad and Noha A. Youssif
Cairo University

Abstract: Longitudinal studies represent one of the principal research strategies employed in medical and social research. These studies are the most appropriate for studying individual change over time. The premature withdrawal of some subjects from the study (dropout) is termed nonrandom when the probability of missingness depends on the missing value. Nonrandom dropout is common phenomenon associated with longitudinal data and it complicates statistical inference. Linear mixed effects model is used to fit longitudinal data in the presence of nonrandom dropout. The stochastic EM algorithm is developed to obtain the model parameter estimates. Also, parameter estimates of the dropout model have been obtained. Standard errors of estimates have been calculated using the developed Monte Carlo method. All these methods are applied to two data sets.

Key words: Dropout, longitudinal data, mixed models, stochastic EM.

1. Introduction

Longitudinal studies represent one of the principal research strategies employed in medical and social research. The defining feature of such studies is that subjects are measured repeatedly through time. A pervasive problem associated with longitudinal data studies is the missing values. Missing values arise in longitudinal data whenever one or more of measurement sequences for participating subjects are incomplete, in the sense that the intended measurements are not taken, lost, or otherwise unavailable. The missing data pattern is termed *dropout* when some subjects leave the study prematurely, i.e. any missing value is never followed by an observed value.

Let n be the number of time points and Y_i be $n \times 1$ vector of intended measurements for the i th subject, which would have been obtained if there were no missing values. This vector can be partitioned to $Y_{i,obs}$ and $Y_{i,mis}$, where $Y_{i,obs}$ are the measurements that actually obtained and $Y_{i,mis}$ are the missing measurements. Let Y be the complete set of measurements which can be partitioned as $Y = (Y_{obs}, Y_{mis})$, where Y_{obs} are the observed data and Y_{mis} are the missing

values. Finally, let R denote a set of indicator variables each takes the value of 1 for elements of Y fall into Y_{obs} and the value of 0 for those fall into Y_{mis} .

Missing data analyses usually base on assumptions about the process that creates the missing values. The missing data process can be classified to three different categories (Rubin, 1976). The missing data mechanism is defined as *missing completely at random (MCAR)* if the probability of the missingness does not depend on the measurements. The missing data mechanism is defined as *missing at random (MAR)* if the probability of missingness does not depend on the unobserved measurements. The missing data mechanism is defined as *non-random* if the the probability of missingness depends on unobserved measurements and may be on the observed measurements. Hence, the missing data mechanism is nonrandom if R is dependent on Y_{mis} and may be on Y_{obs} .

In the selection model (Heckman, 1976), the joint distribution of the complete data Y and R can be factorized as $f(Y, R|\theta, \Psi) = f(Y|\theta)P(R|Y, \Psi)$, where Ψ denotes the vector of the parameters that govern the missingness process; θ stands for the other model parameters and $f(R|Y_{obs}, Y_{mis}, \Psi)$ is the distribution that describes missing data mechanism. Diggle and Kenward (1994) propose a selection model for continuous longitudinal data with nonrandom dropout. This model assumes that the probability of dropout for the i th subject at time t_{d_i} depends on the history of the measurements up to and including time t_{d_i} ,

$$P[D_i = d_i|history] = P_{d_i}[H_{id_i}, Y_{id_i}^*; \Psi],$$

where D_i is a random variable such that $2 \leq d_i \leq n$ identifies the dropout time and $d_i = n + 1$ identifies no dropouts, $H_{id_i} = (Y_{i1}, \dots, Y_{id_i-1})$ denotes the observed history up to time t_{d_i-1} and $Y_i^* = (Y_{i1}^*, \dots, Y_{in}^*)$ denotes the complete vector of intended measurements. The joint probability density function for an incomplete sequence with dropout at time d_i is

$$f(Y_i|\theta, \Psi) = f_{id_{i-1}}(Y_i) \prod_{k=2}^{d_i-1} [1 - P_k(H_{ik}, Y_{ik}^*, \Psi)]P(Y_{id_i} = 0|H_{id_i}, Y_{id_i-1} \neq 0),$$

where $f_{id_{i-1}}(Y_i)$ denotes the joint probability density function of the first $(d_i - 1)$ elements of Y_i . Hence, the log-likelihood function for data consist of m subjects is

$$\ell(\theta, \Psi) = \ell_1(\theta) + \ell_2(\Psi) + \ell_3(\theta, \Psi), \quad (1.1)$$

where

$$\ell_1(\theta) = \sum_{i=1}^m \log[f_{id_{i-1}}(Y_i)],$$

$$\begin{aligned}\ell_2(\Psi) &= \sum_{i=1}^m \sum_{k=1}^{d_i-1} \log[1 - P_k(H_{ik}, Y_{ik}^*, \Psi)], \\ \ell_3(\theta, \Psi) &= \sum_{i:d_i \leq n} \log P(D_i = d_i | Y_i, \theta, \Psi),\end{aligned}$$

and

$$P(D_i = d_i | Y_i) = \int P_k(H_{ik}, Y_{ik}, \Psi) f_{ik}(Y_{ik} | H_{ik}; \theta) dY_{ik}.$$

Diggle and Kenward (1994) suggest modelling the probability of dropout as a logistic linear specification, i.e.

$$\text{logit}\{P(H_{ik}, Y_{ik}; \Psi)\} = \Psi_0 + \Psi_1 Y_{ik} + \sum_{j=2}^k \Psi_j Y_{i,k+1-j}. \quad (1.2)$$

Many methods were developed to handle missing data problems in literature. The EM algorithm (Dempester, Laird and Rubin, 1977) is a common approach for parameter estimation in the incomplete data setting. However, calculating the conditional expectation required in the E-step, may be infeasible in many situations. A possible alternative is to perform the expectations via simulation. This is the general idea of the stochastic EM algorithm (Celuex and Diebolt, 1985).

At each iteration of the stochastic EM algorithm the missing data is imputed with a single draw from the conditional distribution of the missing data given the observed data and the current parameter estimates. This imputation of the missing values is based on all our current information about θ , and hence provides us with a plausible pseudo-complete data. Once we have a pseudo-complete data, we can directly maximize its log-likelihood to obtain updated estimates. The whole process is iterated for sufficient number of iterations. If the iterations converge, the final output of the stochastic EM algorithm is a sample from the stationary distribution $\pi(\cdot)$. The mean of this sample, after discarding the first early points as a burn-in period, is considered as an estimate for θ . This mean is called the stochastic EM estimate and denoted by $\tilde{\theta}$.

Due to dropout, the number of measurements varies between subjects. Hence, analysis of longitudinal data requires statistical models which can accommodate this aspect. Also, the intercorrelation between measurements on one subject must be taken into account to draw valid statistical inference. The linear mixed model is a possible candidate for such analysis.

2. Linear Mixed Model

The linear mixed model (Laird and Ware, 1982) assumes that the response vector Y_i for the i th subject satisfy the linear regression model

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i, \quad (2.1)$$

where Y_i is the $n_i \times 1$ vector of observations for the i th subject, X_i is a design matrix, or set of explanatory variables, β is a $p \times 1$ vector of population-specific parameters describing the average trends (fixed effects), Z_i are $n_i \times q$ matrix of known covariates, b_i are a $q \times 1$ vector of subject-specific parameters describe how the i th subject deviate from the population average (random effects). The residuals ϵ_i are assumed to be independent normally distributed with zero means and covariance matrix Σ_i , usually assumed to be $\sigma^2 I_{n_i}$.

Inference for linear mixed effects model can be made using the marginal distribution of the response Y_i , assuming that the b_i are independent random variables with common distribution function $g(b_i)$, called the mixing distribution. The mixing distribution is assumed to be normal with zero means and a variance-covariance matrix D of order $q \times q$. It is common to assume that ϵ_i and b_i are independent. According to the model (2.1), marginally Y_i is normally distributed with mean vector $\mu_i = X_i\beta$ and covariance matrix $V_i = Z_i D Z_i' + \Sigma_i$. The elements of the matrix V_i are known functions of q unknown covariance parameters. Assume that a vector α contains these covariance parameters, and so the model parameters, θ , are β and α . It is common to assume that $\Sigma_i = \sigma^2 I_{n_i}$. A good review of linear mixed models, in the context of longitudinal data, can be found in Cnaan, Laird and Slasor (1997) and Molenberghs and Verbeke (2001).

In the context of longitudinal data, many methods for fitting linear mixed models are available. For nonrandom missingness, in the likelihood context, Ibrahim, Chen and Lipsitz (2001) derive a computationally feasible E-step, of the EM algorithm, by integrating out the random effects and use the Monte Carlo EM (MCEM) algorithm (Wei and Tanner, 1990) to obtain ML estimates. They use Gibbs sampler to implement the imputation of missing values. They don't estimate Ψ (missingness parameters), they consider it as nuisance parameter. Stubbendick and Ibrahim (2003) propose a selection model for estimating parameters in presence of nonrandom missing responses and covariates. Parameters are estimated via maximum likelihood using the Gibbs sampler and the Monte Carlo EM algorithm.

The purpose of this paper is to fit the linear mixed model, in the presence of nonrandom dropouts, for continuous longitudinal data. The Diggle-Kenward selection model is adopted. Hence, we rely on the log-likelihood function defined in Eq. (1.1). The underlying model has many parameters to be estimated, the

θ and Ψ . It would seem then, estimation using the standard techniques, the EM algorithm for example, is difficult to implement. Also, in the E-step of the EM algorithm, the integrations are intractable. To overcome these difficulties, the stochastic EM algorithm is developed to obtain parameter estimates. Also, the Monte Carlo method (Ip, 1994) is developed to find standard errors of parameter estimates. The rest of the paper is organized as follows. In Section 3, the stochastic EM algorithm is developed to carry out inference for linear mixed model. Also, the Monte Carlo method is developed to obtain standard errors. Finally, in Section 4, two examples are presented and inference is conducted using the proposed techniques.

3. The Stochastic EM Algorithm

The stochastic EM algorithm is developed to obtain the parameter estimates. The elegance of the stochastic EM algorithm actually lies in its simplicity. The stochastic EM algorithm has many advantages. The main advantage of this algorithm is that it avoids evaluating the integrations, in the E-step of the EM algorithm. In the S-step these integrations are evaluated using simulation, so no need for numerical integrations. In the current setting, we have to evaluate the integrals included in the log-likelihood function in Eq. (1.1). This complicates the M-step of the EM algorithm, which is not the case with the developed stochastic EM algorithm. Also, the stochastic EM algorithm can recover multimodality of the likelihood surface (Ip, 1994). The deterministic M-step and the stochastic S-step generate a homogeneous, ergodic and irreducible Markov chain which converges to its stationary distribution faster than the usual long chains of conditional stochastic draws used in many Gibbs sampling schemes. The S-step prevents the sequence $\{\theta^{(t)}\}$ from staying near an unstable stationary point of the likelihood function. Thus, the stochastic EM algorithm avoids the cases of slow convergence observed for the EM algorithm. Moreover, the stochastic EM algorithm can avoid the saddle points or the insignificant local maxima. This property of the stochastic EM algorithm may lead to an improvement over the deterministic EM algorithm in terms of convergence to a good stationary point (Ip, 1994). A relatively recent review of the EM algorithm and its extensions is in McLachlan and Krishnan (1997) and references therein.

The complete data for the i th subject are $Y_{i,obs}, Y_{i,mis}, R_i$. The joint probability density function $f(Y_{i,obs}, Y_{i,mis}, R_i | \theta, \Psi)$ can be factorized as

$$f(Y_{i,obs}, Y_{i,mis}, R_i | \theta, \Psi) = f(Y_{i,obs}, Y_{i,mis} | \theta) f(R_i | Y_{i,obs}, Y_{i,mis}; \Psi).$$

Let the early first missing value for the i th subject is denoted as $y_{i,mis}$. We argue that in nonrandom dropout setting, and depending on the dropout model (1.2), if this value is imputed then, given its value, the remaining missing values can be

assumed as missing at random. Using this idea, given that $y_{i,mis}$ for each subject is filled in, any optimization procedure for incomplete data can be used.

In the nonrandom dropout setting, the two steps of the stochastic EM algorithm, at the $(t + 1)$ th iteration, are developed as follows:

- **S-Step:** The first missing value of each subject is simulated from the conditional distribution $f(y_{i,mis}|Y_{i,obs}, R_i, \theta^{(t)}, \Psi^{(t)})$. This distribution does not have a standard form, hence it is not possible to use the direct simulation. To overcome this problem, an accept-reject procedure is proposed to simulate the missing values $y_{i,mis}$. This procedure mimics the dropout process. It is as follows:
 1. A candidate value, y^* , is generated from the conditional distribution function $f(y_{i,mis}|Y_{i,obs}, \theta^{(t)})$ which is normal distribution.
 2. Calculate the probability of dropout for y^* , according to the dropout model (1.2), where the parameters Ψ_j are fixed at the current values $\Psi_j^{(t)}$. Let us denote this probability as P_i^* .
 3. A random variate U is simulated from the uniform distribution on the interval $[0, 1]$, then $y_{i,mis} = y^*$ if $U \leq P_i^*$; otherwise repeat step 1.
- **M-Step:** The M-step consists of two substeps, the logistic step (M1-step) and the normal step (M2-step).

In the M1-step, the maximum likelihood estimates of the dropout parameters in model (1.2) are obtained using any iterative method for likelihood estimation of binary data models (see, for example, McCullagh and Nelder, 1989).

In the M2-step, maximum likelihood estimates of the parameters β and α are obtained using an appropriate optimization approach for incomplete data, Jennrich-Schluchter algorithm (Jennrich and Schluchter, 1986) for example. Note that by using this algorithm, the observed data for each subject are redefined to be the originally observed data $Y_{i,obs}$ augmented by the simulated value of $y_{i,mis}$. The remainder of the sequence need not be simulated because, given the simulated value, the setting become that of MAR, with separable likelihood components.

When implementing the stochastic EM algorithm we need to check the convergence of the resulting chain. Several convergence diagnostics have been proposed in literature, see for example Brooks (1998) and Brooks and Roberts (1998) for details. Some of these techniques depend on the chain output such as the Gelman-Rubin method (Gelman and Rubin, 1992). This method is used to monitor convergence of the stochastic EM iterations in this paper. This method is

based on generating multiple, $s \geq 2$, chains in parallel for $v = 2p$ iterations. For each chain, this method suggests starting from different points for which the starting distribution is over-dispersed compared to the target distribution. This method is separately monitoring the convergence of each scalar parameter of interest by evaluating the Potential Scale Reduction Factor, (PSRF), $\sqrt{\hat{R}}$ as

$$\sqrt{\hat{R}} = \sqrt{\frac{v-1}{v} + \frac{1}{v} \frac{B}{W}},$$

where B/v is the between sequence variance and W is the average of within sequence variances. This calculation depends on the last p iterations of each sequence. The convergence is achieved if the PSRF is close to one which means that the parallel Markov chains are essentially overlapping. If the PSRF is large, then proceeding further simulation may be needed.

Louis (1982) derive an interesting identity relating the observed data log-likelihood and the complete data log-likelihood:

$$-\frac{\partial^2 \ell(\theta, Y_{obs})}{\partial \theta \partial \theta'} = E \left(-\frac{\partial^2 \ell(\theta | Y_{obs}, Y_{mis})}{\partial \theta \partial \theta'} \middle| Y_{obs} \right) - \text{cov} \left(\frac{\partial \ell(\theta | Y_{obs}, Y_{mis})}{\partial \theta} \middle| Y_{obs} \right).$$

Efron (1994) and Ip (1994) propose using simulation to find the two parts in the right hand side of Louis' formula (the Monte Carlo method). The main idea is to simulate M identically independent distributed samples Z_1, Z_2, \dots, Z_M from the conditional distribution of the missing values given the observed values. Then, the two parts can be approximated by their empirical versions, i.e.

$$E \left(\frac{\partial^2 \ell(\theta | Y_{obs}, Y_{mis})}{\partial \theta \partial \theta'} \middle| Y_{obs} \right) \approx \frac{1}{M} \sum_{j=1}^M \frac{\partial^2 \ell(\theta | Y_{obs}, Z_j)}{\partial \theta \partial \theta'},$$

and

$$\text{cov} \left(\frac{\partial \ell(\theta | Y_{obs}, Y_{mis})}{\partial \theta} \middle| Y_{obs} \right) \approx \frac{1}{M} \sum_{j=1}^M \left(\frac{\partial \ell(\theta | Y_{obs}, Z_j)}{\partial \theta} \right)^2 - \left(\frac{1}{M} \sum_{j=1}^M \frac{\partial \ell(\theta | Y_{obs}, Z_j)}{\partial \theta} \right)^2,$$

where the parameter θ is fixed at $\hat{\theta}$.

In the nonrandom dropout setting the conditional distribution of the missing data of the i th subject is $f(Y_{i,mis} | Y_{i,obs}, R_i)$. Again the proposed accept-reject procedure can be used to simulate from this distribution.

4. Applications

4.1 The mastitis data

These data were analyzed by Diggle and Kenward (1994) using fixed effects model and a different technique. Mastitis is an infection of the udder which reduces milk yield of the infected animals. Data were available of milk yields in thousands of liters of 107 dairy cows, from a single herd, in two consecutive years such that, Y_{ij} , $i = 1, \dots, 107$; $j = 1, 2$. In the first year, all animals were supposedly free of mastitis and in the second year, 27 became infected (assumed missing). The question of scientific interest was whether the occurrence of mastitis is related to the yield that would have been observed if mastitis is not occurred. Initial exploration of the data shows two patterns for the first 80 animals (completers). First, as expected there is an increase in the milk yield from the first year to the second year. Second, the variance between milk yield in the first year appears to be smaller than the second year.

Table 1: Parameter estimates of the mixed effects model for the mastitis data

Parameters	Estimates	Standard Errors	p -value
β_0	5.76	0.0000	<0.0001
β_1	0.35	0.0147	<0.0001
d_{11}	0.54	0.0368	<0.0001
d_{12}	0.02	0.0357	0.9715
d_{22}	0.72	0.0685	0.0001
σ^2	0.33	0.0369	<0.0001
Ψ_0	0.19	0.1283	0.1507
Ψ_1	2.07	0.0673	<0.0001
Ψ_2	-2.33	0.0894	<0.0001

The linear mixed effects model (2.1) is fitted for these data. The design matrix X_i is chosen as

$$X_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

The fixed effects parameters vector is of dimension 2 where $\beta = (\beta_0, \beta_1)$. The covariance matrix is $V_i = Z_i D Z_i' + \sigma^2 I$, where

$$D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix},$$

and $Z_i = X_i$. The dropout process is modelled using logistic regression as in Eq. (1.2). The linear mixed model for the yield and the dropout model are

incorporated to give the log-likelihood function in Eq. (1.1). The developed stochastic EM algorithm is used to find the parameter estimates. The number of iterations is fixed at 2000 iterations and the calculations are based on the last 1000 iterations. The results are presented in Table 1.

It is noticed from the results in Table 1 that variability within individuals ($\sigma^2 = 0.33$) is less than the variability between individuals in the first year, d_{11} , or in the second year, d_{22} . In addition the variance between individuals in the second year is greater than the one in the first year. Diggle and Kenward (1994) observed that in nonrandom models, dropout tends to depend on the increment (i.e., the difference between the current and previous measurements, $Y_{d_i} - Y_{d_i-1}$). According to the obtained estimates, the dropout model is as

$$\text{logit}[P(\text{dropout})] = 0.19 + 2.07Y_1 - 2.33Y_2,$$

where Y_1 denotes the first observation and Y_2 denotes the missing one. So, this model can be rewritten as

$$\text{logit}\{P(\text{dropout})\} = 0.19 + 0.26Y_1 - 2.33(Y_2 - Y_1).$$

The model suggests that the larger the decrease in milk yield between the two years, the higher the probability of getting disease. This means that the animals with greater decrease in yield over the two years have a higher probability of getting mastitis. The parameter β_0 represent the population average of the first year response, whereas β_1 represent the population average increment of the second year over the first year.

Standard errors have been obtained using the proposed Monte Carlo method and results are shown in Table 1. Statistical inference about the unknown parameters can be obtained through the standard errors of parameter estimates. Based on the results displayed in Table 1, we can conclude that the null hypothesis $H_0 : \Psi_2 = 0$ cannot be accepted. This may be believed as an evidence for non-random dropout process.

For diagnosing convergence we apply the Gelman-Rubin method using 3 sequences with different starting points. The shrink factor, $\sqrt{\hat{R}}$, has been calculated for each one of the 9 parameters. The results are close to 1 with larger value 1.08 for d_{11} and smaller value 0.99 for β_0 . This means that the generated sequences have converged properly.

4.2 The rats data

These data were analysed by Molenberghs and Verbeke (2001), also in Verbeke and Molenberghs (2004), in which 50 male Wistar rats were randomized to either a control group (15 rats) or one of the two treatment groups. The two treatment

groups are low dose (18 rats) or high dose (17 rats) of the drug Decapeptyl, which is an inhibitor for testosterone production in rats. The primary aim of the study was to investigate the effect of the inhibition of the production of testosterone in male Wistar rats on their craniofacial growth. The research question is: how does craniofacial growth depend on testosterone production? The height of the skull is one of the responses of interest. Height is measured as the distance (in pixels) between two well-defined points on X-ray pictures of the skull. The treatment started at the age of 45 days, and measurements were taken every 10 days, with the first observation taken at the age of 50 days. The measurements were intended to the age of 110 days. Many rats do not survive anaesthesia and therefore drop out before the end of the study, only 22 of them survived to the end of study. This experiment is balanced: the same number of repeated measurements is planned for all participating subjects, at fixed time points.

Verbeke and Molenberghs (2004) suggest the following model,

$$\begin{aligned}
 Y_{ij} &= (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i})t_{ij} + \epsilon_{ij}, \\
 &= \begin{cases} (\beta_0 + b_{1i}) + (\beta_1 + b_{2i})t_{ij} + \epsilon_{ij} & \text{if low dose} \\ (\beta_0 + b_{1i}) + (\beta_2 + b_{2i})t_{ij} + \epsilon_{ij} & \text{if high dose} \\ (\beta_0 + b_{1i}) + (\beta_3 + b_{2i})t_{ij} + \epsilon_{ij} & \text{if control,} \end{cases}
 \end{aligned}$$

where Y_i is a vector of order 7 containing the response values for the i th animal, L_i, H_i and C_i are indicator variables such that

$$L_i = \begin{cases} 1 & \text{if low dose} \\ 0 & \text{otherwise,} \end{cases}$$

$$H_i = \begin{cases} 1 & \text{if high dose} \\ 0 & \text{otherwise,} \end{cases}$$

$$C_i = \begin{cases} 1 & \text{if control} \\ 0 & \text{otherwise,} \end{cases}$$

and $t_{ij} = \ln(1 + (Age - 45)/10)$ which is a transformation of the time scale. The β_0 can be interpreted as the average response at the start of the treatment (independent of treatment) and β_1, β_2 and β_3 as the average time effect for each treatment group.

The design matrix Z_i is then

$$Z_i = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ t_{i1} & t_{i2} & t_{i3} & t_{i4} & t_{i5} & t_{i6} & t_{i7} \end{pmatrix}'.$$

The design matrix X_i is of order 7×4 where all the values of the first column are one. The second, third and fourth columns are t_{ij} for $j = 1, \dots, 7$ if the subject is

in the low, high or control group, respectively. Otherwise, the column elements are zeros.

The stochastic EM algorithm is used to obtain the parameter estimates of this model in the presence of dropout. Also, the estimates standard errors have been obtained using the Monte Carlo method. The results are shown in Table 2. Although the parameter estimates using the stochastic EM algorithm are similar to those obtained by Verbeke and Molenberghs (2004), we have got positive values for the variance component (d_{22} in particular). They cannot obtain positive values for the variance component d_{22} using restricted maximum likelihood method. The parameter estimates of the dropout model (Ψ) suggest that the smaller the height of the skull, the larger the probability of death due to the inhibition of testosterone. It is also noted that there is much variability between rats at the beginning of the experiment and less variability within rats. Due to time effect for each treatment group the variability between rats decreases.

Table 2: Parameter estimates of the mixed effects model for the rats data

Parameter	Estimates	Standard Errors	p -value
β_0	68.65	0.0028	<0.0001
β_1	7.41	0.0053	<0.0001
β_2	6.83	0.0061	<0.0001
β_3	7.22	0.0082	<0.0001
d_{11}	2.92	0.0101	<0.0001
d_{12}	0.24	0.0668	0.0003
d_{22}	0.84	0.3878	0.0308
σ^2	1.52	0.0025	<0.0001
Ψ_0	37.17	0.4039	<0.0001
Ψ_1	0.07	0.0112	<0.0001
Ψ_2	-0.53	0.0159	<0.0001

Convergence has been monitored in the same manner of the previous example. Two chains each of 2000 iterations have been used to implement Gelman-Rubin method. The shrink factor is close to 1 for all parameters with larger value of 1.005. This indicates that convergence has been achieved.

5. Concluding Remarks

We fitted the linear mixed model to continuous longitudinal data with non-random dropout. The estimation process depends on the marginal distribution of the response variable. We assumed that the error terms are normally distributed, hence the marginal distribution of the response is normal. We consider the Diggle-Kenward selection model. In this model, a probability dropout model

is incorporated in the likelihood function. The fixed effects parameters, the variance components and the dropout parameters need to be estimated. The stochastic EM algorithm (Celuex and Diebolt, 1985) is proposed and developed to obtain parameter estimates. Also, the Monte Carlo method Efron (1994) is developed to obtain the standard errors.

We applied the proposed approaches to the mastitis data of diary cattle and the rats data. The linear mixed model has been used to fit these data and parameter estimates are obtained.

The approaches presented in this paper can overcome the drawbacks of other algorithms such as the EM algorithm and the Monte Carlo EM algorithm. Also, the implementation of the stochastic EM algorithm is easier.

Acknowledgments

The Authors are grateful to the Editor for helpful comments and to a referee for important comments and suggestions on the manuscript.

References

- Brooks, S. P., and Roberts, G. O. (1998). Assessing convergence of Markov chain algorithms. *Statistic and Computing* **8**, 319-335.
- Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *The Statistician* **47**, 69-100.
- Celuex, G., and Diebolt, J. (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problems. *Computational Statistics Quarterly* **2**, 73-82.
- Cnaan, A., Laird, N. M. and Slasor, P. (1997). Using the general linear model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine* **16**, 2349-2380.
- Dempester, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistics Society B* **39**, 1-38.
- Diggle, P. J. and Kenward, M. G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Journal of Royal Statistical Society B* **43**, 49-93.
- Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of the American Statistical Association* **89**, 463-475.
- Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457-511.

- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economics and Social Measurement* **5**, 475-492.
- Ibrahim, J. G., Chen, M. and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**, 551-564.
- Ip, E. H. S. (1994). A stochastic EM estimator in the presence of missing data: Theory and applications. PhD thesis, Division of biostatistics, Stanford University, California, USA.
- Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics* **42**, 805-820.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of Royal Statistical Society B* **44**, 226-232.
- McLachlan, G.J. and Krishnan, T. (1997). *The EM algorithm and extensions*. Wiley.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd edititon, Chapman and Hall.
- Molenberghs, G. and Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling* **1**, 235-269.
- Rubin, D. B. (1976). Inference and missing data. *Biometrics* **63**, 581-592.
- Stubbendick, A. L. and Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing response and covariates in random effects models. *Biometrics* **59**, 1140-1150.
- Verbeke, G. and Molenberghs, G. (2004). Longitudinal and incomplete data Analysis. *Proceedings of the international biometric conference*, Cairns, Australia.
- Wei, G. C. G., and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* **85**, 699-704.

Received March 3, 2005; accepted June 16, 2005.

Ahmed M. Gad
Statistics Department
Faculty of Economics and Political Science
Cairo University, Cairo, EGYPT
hagas10@hotmail.com

Noha A. Youssif
Statistics Department
Faculty of Economics and Political Science
Cairo University, Cairo, EGYPT
nohayoussif@hotmail.com