American Journal of Applied Mathematics and Statistics, 2013, Vol. 1, No. 2, 30-35
Available online at http://pubs.sciepub.com/ajams/1/2/3
© Science and Education Publishing
DOI:10.12691/ajams-1-2-3

A Shared Parameter Model for Longitudinal Data with Missing Values

Ahmed M. Gad*, Nesma M. M. Darwish

Department of Statistics, Faculty of Economics and Political Science, Cairo University, Cairo, Egypt
*Corresponding author: ahmed.gad@feps.edu.eg

Received March 15, 2013; Revised April 24, 2013; Accepted April 26, 2013

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 prematurely 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. The shared parameter 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. The proposed approach performance is evaluated through a simulation study. Also, the proposed approach is applied to a real data set.

Keywords: *longitudinal data, missing data, Monte Carlo, nonrandom missing, repeated measures, shared parameters, standard errors, stochastic EM*

1. Introduction

In longitudinal studies each subject is measured repeatedly for the same response variable at different times or different condition or both. For example, if the weights of a sample of individuals are measured once a week for twenty consecutive weeks, the collection of these weights is longitudinal data. The main advantage of longitudinal studies is that it can distinguish changes over time within individuals and enabling direct study of that change.

Longitudinal data are very common in biomedical research and clinical trials where some of measurement of a person develops over time, for example the status of a disease of one person or the value of a car, evolves or develops over time. In these cases one variable is the underlying characteristic or measurement. Longitudinal studies are in contrast to cross–sectional studies in which single outcome are measured for each individual (taken at only one fixed point in time).

Missing data are very common with longitudinal studies. The missing data occur whenever, one or more of, measurement sequences are incomplete. The missing values could be for many reasons. Missing data can be categorized into two different patterns; intermittent missing pattern and dropout pattern. In intermittent pattern a missing value could be followed by an observed value. Dropout means a missing value never followed by an observed value.

A distinguishing feature of incomplete longitudinal data analysis is the need to address the underlying causes of missing values. Missing data mechanism is classified to three different types due to [1,2]. These types are missing completely at random, missing at random, and nonrandom missingness. A nonresponse process is *missing completely at random* (MCAR) if the missingness is independent of both unobserved and observed data, and *missing at random* (MAR) if, conditional on the observed data, the missingness is independent of the unobserved measurements. A process that is neither MCAR nor MAR is *nonrandom* (MNAR). For likelihood inference, and when the parameters describing the measurement process are functionally independent of the parameters describing the missingness process, MCAR and MAR are *ignorable*, in which case the missingness process can be ignored when interest is in inference for the longitudinal process only. The missing data mechanism is referred to as *informative* if the probability of missingness is related to the underlying response process [3]. Follmann and Wu [4] have shown that informative missing mechanism is a special case of the nonignorable missing mechanism.

Ignoring the missing values with longitudinal data analysis would lead to biased inference. Many authors have tried to model jointly the response process and the missing data process. This modeling framework includes the selection models, pattern mixture models and shared parameter models. In shared parameter models a random effect is shared between the repeated measures model and the missing data mechanism model.

Many authors have proposed a shared parameter model for longitudinal data subject to informative missingness. Wu and Carroll [3,5] proposed a model for continuous normally distributed longitudinal data. Follmann and Wu [4,6,7,8,9] proposed models for binary longitudinal responses. Albert and Follmann [10] proposed methodology for longitudinal count data.

The EM algorithm [11] can be used to obtain the maximum-likelihood estimates for incomplete data. However, in the nonrandom case, the simplicity of the EM algorithm is lost. The expectation step is problematic and does not admit a closed form solution. Also, in some situations, the M-step is computationally unattractive. Many authors have tried to introduce new variants of the EM algorithm that can overcome the complexity of the E-step. A possible solution for the intractable E-step is to use the Monte Carlo EM algorithm [12,13] and a stochastic version of the EM algorithm [13-18]. A relatively recent review of the EM algorithm and its extensions is in [19] and references therein. The stochastic EM (SEM) algorithm is a stochastic version of the EM algorithm, which has been introduced by [14], and subsequently in [16], as a way for executing the E-step using simulation.

The EM algorithm does not provide directly the standard errors of the estimates. Hence, methods for evaluating these standard errors need to be considered. Several methods have been introduced to solve this problem, see for example, Louis [20,21,22]. Efron [23] and [24] have introduced a stochastic version of the Louis' method (the Monte Carlo method).

In the current paper, we propose a model in which a random effect is shared between the response process and the missing data mechanism. We develop the stochastic EM algorithm (SEM algorithm) to estimate the model parameters. Also, the Monte Carlo method is developed to obtain the standard errors. In Section 2 we discuss the motivating opiate clinical trial example. In Section 3, we develop the random effects transition model and discuss parameter estimation in Section 4. We illustrate this methodology with the opiate clinical trial data in Section 5. A discussion follows in Section 6.

2. The Models

Assume that the number of subjects is m and the intended measurements for the ith subject are n_i . Assume that due to missing data only n_o measurements are available of the n_i, whereas n_{im} measurements are missing, $n_i=n_{io}+n_{im}$. Let y_{ij} represents the jth measurement on the ith subject, $i=1, \ldots, m$, $j=1, \ldots, n_i$. Let Y_i be an $n_i \times 1$ vector containing the responses that would be obtained, for the ith subject, if there were no missing values. Assume that the observed and missing components of Yi are denoted as $Y_{i,\text{obs}}$ and $Y_{i,\text{mis}}$, respectively. Let R_i be a vector of missingness indicators. For a particular realization of (Y_i,R_i), each element of R_i takes a value of one if the corresponding value of Y_i is observed and the value of zero if the corresponding value of Y_i is missing. In notation;

$$R_{ij} = \begin{cases} 0 & \text{if } y_{ij} \text{ is observed} \\ 1 & \text{if } y_{ij} \text{ is intermittent missing} \\ 2 & \text{if } y_{ij} \text{ is dropout.} \end{cases}$$

It is common to jointly model the response process and the missingness process. The complete data of the i*th* subject can be viewed as $(Y_{i,obs},Y_{i,mis},R_i)$, and the full density function is $f(Y_{i,obs},Y_{i,mis},R_i|\ \theta,\ \psi)$, where the parameters vectors θ and ψ describe the measurement and missingness processes, respectively.

The selection model and pattern mixture model are different factorization of the full density function $f(Y_{i,obs},Y_{i,mis},R_i|\ \theta,\ \psi)$. The *selection model* framework is based on the factorization;

$$f\left(Y_{i,obs},Y_{i,mis},R_i|\ \theta,\ \psi\right)= f(Y_{i,obs},Y_{i,mis}|\theta)\, f(R_i|Y_{i,obs},Y_{i,mis},\ \psi)\ .$$

The first factor is the marginal density of the measurement process and the second one is the density of the missingness process, conditional on the response. The *pattern-mixture model* (Little, 1994) is based on the factorization;

$$f\left(Y_{i,obs},Y_{i,mis},R_i|\ \theta,\ \psi\right)= f(Y_{i,obs},Y_{i,mis}|R_i,\theta)\, f(R_i|\psi).$$

This can be seen as a mixture of different populations, characterized by the observed pattern of missingness. Instead of using the selection model or pattern-mixture model, the measurement and the missingness process can be jointly modelled by using a *shared-parameter model* [3,5,6]. These models assume that there is a vector of random effects b_i, that is shared between the response and missingness process.

Different missing data mechanisms defined by [1] can be defined according to the conditional distribution $f(R_i \mid Y_{i,obs},Y_{i,mis},\ \psi)$. The missing data mechanism is MCAR if $f(R_i \mid Y_{i,obs},Y_{i,mis},\ \psi)= f(R_i \mid \psi)$, the missing data mechanism is MAR if $f(R_i \mid Y_{i,obs},Y_{i,mis},\ \psi)= f(R_i \mid Y_{i,obs},\ \psi)$, otherwise the missing data mechanism is MNAR.

The shared parameter model assume that the response process Y_i and the missing data mechanism indicator R_i are conditionally independent of each other, given a group of parameters, b_i. Hence the density function of the complete data $f(Y_{i,obs},Y_{i,mis},R_i|\ \theta,\ \psi)$ can be written as

$$\begin{aligned} & f(Y_{iobs},Y_{imis},R_i|\theta,\varphi) \\ &= \int f(Y_{iobs},Y_{imis}|b_i,\theta)\, f(R_i|b_i,\varphi)\, f(b_i)db_i \end{aligned}$$

Shared parameters b_i affect both the response Y_i and the missing data indicator R_i, thus can be either observable variables (e.g., gender) or latent variables.

Assuming that the response variable Y_i is continuous so, the mixed effects model assumes that the response vectors Y_i satisfies the linear regression model;

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i \ ,$$

where X_i is a set of explanatory variables (design matrix), β is a $p\times 1$ vector of fixed effect parameter, Z_i is the random effects covariates and b_i is a shared parameter. The shared parameters b_i are assumed to be normally distributed with a zero mean and a variance equal to σ^2. The errors ε_i are assumed to be independent normally distributed with zero means and V_i covariance matrix. The matrix V_i may be unstructured and hence it contains $n_i(n_i+1)/2$ parameters. Also, the covariance matrix may have a parametric structure, i.e. its elements are functions of a smaller number (vector) of parameters α. In this case it can be written as $V_i(\alpha)$. The main reason for modelling the covariance matrix, Vi, as a function of parameters α is to examine different covariance structures, and for parsimony.

Tsonaka et al [25] have shown that the shared parameters model, by construction, implies a missing not at random (MNAR) mechanism. The conditional distribution $f(R_i \mid Y_{i,obs}, Y_{i,mis}, \psi)$ can be viewed as

$$f(R_i \mid Y_{i,obs}, Y_{i,mis}, \psi) = \int f(R_i \mid b_i, \psi) f(b_i \mid Y_{i,obs}, Y_{i,mis}, \psi) d\, b_i$$

which shows that the probability of nonresponse depends on $Y_{i,mis}$ through the posterior $f(b_i \mid Y_{i,obs}, Y_{i,mis}, \psi)$, corresponding therefore to a nonignorable mechanism.

The missing data process, conditional on the random effects b_i, can be modeled as [9],

$$P(R_{ij} = k \mid b_i, R_{ij} \neq 2) = \begin{cases} \dfrac{1}{1+\sum_{k=1}^{k=2} 1+exp(w'_{kij}\,\eta_k + r_k b_i)} & if\ k=0 \\ \dfrac{exp(w'_{kij}\,\eta_k + \gamma_k b_i)}{1+\sum_{k=1}^{k=2} 1+exp(w'_{kij}\,\eta_k + \gamma_k b_i)} & if\ k=1,2 \end{cases} \quad (1)$$

where w_{kij} are vectors of covariates and η_k are their corresponding regression coefficients. Also, the parameters γ_k relate the missingness process (intermittent or dropout) to the response process.

The likelihood function for the parameters (θ, ψ),

$$L(\theta, \varphi) = \prod_{i=1}^{m} \int_{b_i} \prod_{j=1}^{n_{oi}} L_{ij} \prod_{j=1}^{t_i} P_{1ij}^{\,I(R_{ij}=1)} P_{2ij}^{\,I(R_{ij}=2)} (1 - P_{1ij} - P_{2ij})^{I(R_{ij}=0)} g(b_i) db_i \quad (2)$$

where n_{oi} is the number of observed measurements for subject i, t_i is the last measurement time, $P_{lij}=P(R_{ij}=l|b_i, R_{ij-1}\neq 2)$, $l=1,2$ are as given in Eq. (1) and I(.) are indicator function which equal 1 if the condition is met and zero otherwise. Note that,

$$\prod_{j=1}^{n_{oi}} L_{ij} = L_i = f(Y_{iobs} \mid b_i, \theta) = \frac{1}{\sqrt{2\pi |v_{iobs}|}} e^{-1/2 (y_i - X_{iobs}\beta)' v_{iobs}^{-1} (y_i - X_{iobs}\beta)}$$

where v_{iobs} and X_i is a suitable partition of the V_i and X_i respectively and $\theta=(\beta,\alpha)$.

Maximizing this likelihood function we can obtain the parameter estimates. However, this maximization is not easy to implement and computationally intractable. We suggest and develop the stochastic EM algorithm to obtain the parameters estimates.

3. Estimation

We propose fitting the shared parameters model using the stochastic EM algorithm. Gad and Ahmed [26] proposed and developed this algorithm in selection models context. In the shared parameters model context the complete data are Y_{iobs}, Y_{imis}, R_i and b_i. In the S-step we need to simulate from the missing data distribution given the observed data, i.e. the conditional distribution

$$f(Y_{imis}, , b_i \mid Y_{iobs,} R_i).$$

This distribution can be partitioned as;

$$f(Y_{imis}, b_i \mid Y_{iobs}, R_i) = f(b_i) f(Y_{imis} \mid b_i, Y_{iobs}, R_i).$$

Hence, to simulate from this conditional distribution of the missing data given the observed data, we need to simulate from two distributions. First, we simulate from the marginal distribution of b_i, $f(b_i)$. This is a normal distribution. Second, we simulate from the conditional distribution of the missing data, $f(Y_{imis}|b_i, Y_{iobs}, R_i)$. We argue that this simulation can be performed from the conditional distribution $f(Y_{imis}|b_i, Y_{iobs},)$, since the missing data depend on the missing data indicator through the random effect parameter. This distribution now is a normal distribution. The developed EM algorithm iterates two main steps; the stochastic step (S-step) and the maximization step (M-step). Henc at the (t+1)th iteration iterates the following steps.

S-step:

This step consists of two sub-steps; sub-step I and sub-step II.

Sub-step I: For each subject i a single draw is obtained from the marginal distribution of b_i; $f(b_i|\sigma^{2(t)})$. This distribution is the normal distribution with mean zero and variance σ^2.

Sub-step II: The missing values of each subject is simulated from the conditional distribution $f(Y_{imis}|Y_{iobs}, b_i, \theta^{(t)}, \psi^{(t)})$. Note that $\sigma^{2(t)}$, $\theta^{(t)}$ and $\psi^{(t)}$ are the current parameters estimates. In case of the dropout pattern we can simulate the first missing value only for each subject. The remaining missing values in this case can be assumed missing completely at random.

M-Step:

The M-step consists of two sub-steps, 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 are obtained using any iterative method for likelihood estimation of binary data models (see, for example [27]). In the M2-step, the maximum likelihood estimates of the parameters β and α are obtained using an appropriate optimization approach for incomplete data. We recommend using the Jennrich-Schluchter algorithm [28].

4. Standard Errors

Louis [20] suggest that the information matrix can be approximated by

$$I(\theta) = E(-\frac{\partial^2 l(\theta|Y_{obs}, Y_{mis})}{\partial\theta\partial\theta}|Y_{obs}) - cov(\frac{\partial l(\theta|Y_{obs}, Y_{mis})}{\partial}|Y_{obs}) \quad (3)$$
$$= -E - C$$

where θ is fixed at the stochastic EM estimates and $l(\theta|Y_{obs}, Y_{mis})$ is the log-likelihood function.

Evaluating the integrals in the formula in Eq. (3), in the current setting, may not be easy. Efron [23], also in Ip [24], suggest using simulation (the Monte Carlo method) to approximate the integrations in Eq. (3). The missing values are simulated from their conditional distribution and then integrations are evaluated by their empirical versions.

The main idea is to simulate M identically distributed samples, $q_1, q_2, \ldots, q_M$ from the conditional distribution of the missing values given the observed values and the

parameters estimates, $f(Y_{mis}|Y_{obs}, \hat{\theta})$. Hence the Louis formula (3) can be approximated by its empirical version, i.e.

$$E \approx \frac{1}{M}\sum_{j=1}^{M}\frac{\partial^2 l(\theta|Y_{obs}, q_j)}{\partial\theta\partial\theta}$$

And

$$C \approx \mathrm{cov}\left(\frac{\partial l(\theta|Y_{obs}, q_j)}{\partial\theta}\right)$$

The Monte Carlo method is developed to find standard errors of the stochastic EM estimates of parameters. The main idea is to simulate M independent identically distributed samples from the conditional distribution of the missing data given the observed data. Hence, we simulate $q_1, q_2, \ldots, q_M$ samples from the conditional distribution $f(Y_{imis}|b_i, Y_{iobs}, R_i)$ and $h_1, h_2, \ldots, h_M$ from the conditional distribution $f(b_i)$. Then the two parts in the right hand side of the formula (3) can be approximated by their empirical versions. In notation,

$$E \sim \frac{1}{M}\sum_{j=1}^{M}\frac{\partial^2 l(\theta|Y_{obs}, R, q_j, h_j)}{\partial\theta\partial\theta}$$

and

$$C \sim \mathrm{cov}\left(\frac{\partial l(\theta|Y_{obs}, R, q_j, h_j)}{\partial\theta}\right)$$

where the parameters $\theta = (\beta, \alpha, \varphi)$ is fixed at the SEM estimates, $\hat{\theta} = (\hat{\beta}, \hat{\alpha}, \hat{\varphi})$.

Having the M pseudo-complete data, the first and second order derivatives of the log-likelihood function are evaluated for each sample, and then it is possible to calculate the quantities E and C and hence the information matrix. The inverse of the information matrix is the covariance matrix of the stochastic EM estimates. The standard error estimates are the square root of the main diagonal elements of this matrix.

5. Application (Anti-Depressant Trial)

This data set is taken from a multicenter clinical trial on the treatment of depression. In each of six centers subjects were randomized to one of three treatments, approximately 20 subjects receiving each treatment in each center. The total number of subjects was 367. Each subject was rated on the Hamilton depression score (HAMD); a sum of 16 test items producing a response on a 0-50 scale. Measurements were made on each of five weekly visits. The first measurement made before the treatment and the remaining four measurements made during treatment. Dropout occurs from the third measurement onwards. At the end of the trail 123 (33%) subjects had left.

A subset of these data have been analyzed by [29], who considered several analyses, including a maximum likelihood analysis. They have shown that an ante-dependence covariance structure of order 2, AD(2), is appropriate for these data.

Diggle and Kenward [30] use maximum likelihood analysis for these data using the same covariance structure, AD(2), as in [29], with a less restricted model for the mean response. For the mean profile, they consider a model in which each center is allowed to have a different intercept and quadratic regression relationships for each treatment group.

Figure 1 shows the set of simple mean profiles, based on the observed data for each center. The figure suggests that there is a nonlinear relationship between mean profiles and time. For this reason [30] suggest modeling the mean profile using quadratic regression for each treatment group.

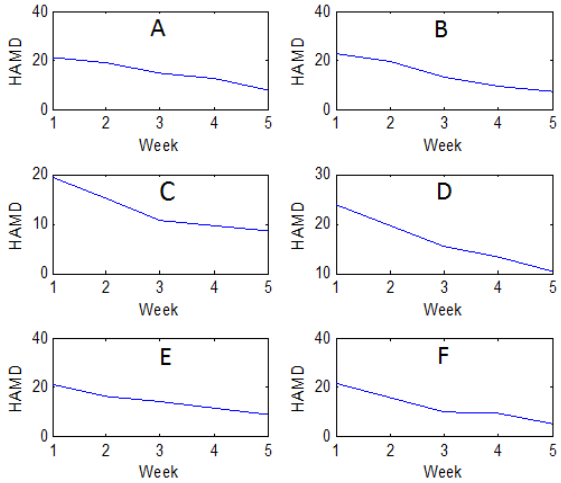

Figure 1. Observed mean response profiles from antidepressant data trail for each center: (A) center 1; (B) center 2; (c) center 3; (D) center4; (E) center 5; (F) center 6

The set of profiles for the completers at each center are plotted in Figure 2. Several aspects can be concluded from this figure. First, there is typical decrease of HAMD score over time in all the centers. Second, the dispersion of measurements between subjects at week 5 higher than week 1. Third, in center 2, there is a subject stars and remains at a high value.

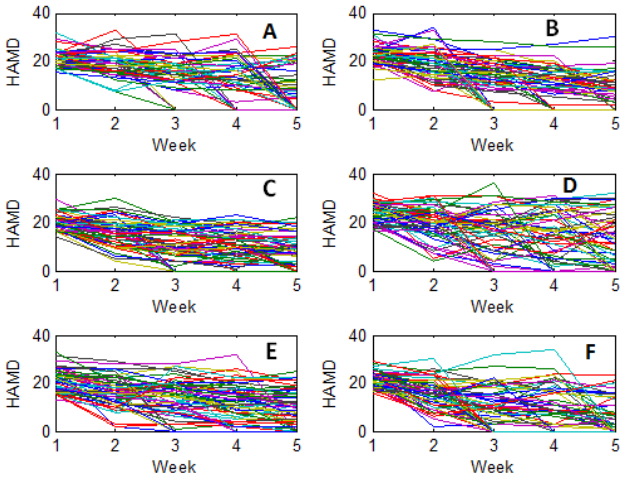

Figure 2. Observed measurements of antidepressant data and centers: : (A) center 1; (B) center 2; (c) center 3; (D) center4; (E) center 5; (F) center 6

Table 1 shows number of subjects who dropout at each treatment in each center. The positive and negative columns shows number of subjects with positive and

negative score increment, respectively, who dropout over the trial period. Center 6 has the highest number of dropouts.

Generally subjects with negative score increments tend to dropout except for few subjects. Dropout occurs consistently from week 3 to week 5, except in the second treatment in center 6. In general week 3 has the highest number of dropouts.

Table 1. Number of dropouts at each center treatment combination of antidepressant data

Center	Treatment		Treatment 2		Treatment 3		Total
	-ve	+ve	-ve	+ve	-ve	+ve	
Center 1	7	1	6	2	2	6	24
Center 2	6	4	3	2	4	1	20
Center 3	3	2	6	4	2	0	17
Center 4	1	4	4	3	4	3	19
Center 5	3	2	4	2	4	1	16
Center 6	10	1	7	1	4	1	27

Table 2. The stochastic EM estimates and their standard errors for antidepressant data

Standard Errors	Estimates	Parameter
β_0	12.03	0.45
β_1	9.70	0.47
β_2	5.57	0.42
β_3	2.83	0.44
β_4	1.97	0.46
β_5	1.07	0.41
γ_1	0.27	0.03
σ_{11}	14.72	1.08
σ_{12}	12.40	1.32
σ_{13}	8.70	1.40
σ_{14}	7.80	1.35
σ_{15}	5.60	1.45
σ_{22}	36.13	2.64
σ_{23}	21.93	2.41
σ_{24}	17.50	2.31
σ_{25}	12.50	2.39
σ_{33}	37.66	2.87
σ_{34}	24.02	2.65
σ_{35}	22.90	2.71
σ_{44}	41.38	2.90
σ_{45}	22.90	2.78
σ_{55}	40.51	3.69

The proposed model is used for these data. The responses are modeled as:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

where Y_i is the antidepressant measures, X_i is a design matrix, β is a vector of unknown parameters represent the centers effect in addition to a constant parameter; $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$, $\boldsymbol{b_i}$ is the random effects and $\boldsymbol{Z_i}$ is a design matrix associated with the shared parameters.

Table 3. The stochastic EM estimates vs. true parameters values for the simulated data

Parameter	Actual	Estimate	Relative Bias %
β_0	8.00	8.02	0.3
β_1	6.00	5.45	9.2
β_2	5.00	5.08	1.6
β_3	7.00	6.35	9.3
β_4	3.00	2.58	14.0
β_5	2.00	1.85	7.5
γ_1	0.30	0.28	7.3
σ_{11}	0.80	0.71	11.2
σ_{12}	0.60	0.62	3.3
σ_{13}	0.90	1.02	13.3
σ_{14}	1.25	1.03	17.6
σ_{15}	1.50	1.62	8.0
σ_{22}	0.50	0.58	16.0
σ_{23}	0.37	0.40	8.1
σ_{24}	0.34	0.31	8.8
σ_{25}	0.16	0.18	12.5
σ_{33}	2.44	2.56	4.9
σ_{34}	0.89	1.03	15.7
σ_{35}	0.92	0.89	3.3
σ_{44}	1.63	1.78	9.2
σ_{45}	2.38	2.50	5.0
σ_{55}	0.40	0.51	27.5

The dropout process is modeled according to the model in Eq. (1). Note that the dropout can be considered as a special case of Eq. (3). The model is

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = w_{1i}\eta_1 + \gamma_1 b_i$$

The developed stochastic EM algorithm has been applied to the models. The parameters estimates of the response model and the dropout model are presented in Table 2. These parameter estimates have been obtained for 3000 iteration and stop when the difference between the

last two is less than .0001 for Jennrich-Schluchter algorithm.

From the results in Table 2 we conclude that the centers effects are statistically significant. Also, the covariance parameters are highly significant. The shared parameter γ_1 is significant which support the nonrandom dropout.

6. Simulation Study

The aim of this simulation study is to validate the obtained stochastic EM estimate by comparing them with the true parameters. The simulation setup is as follows. A random effects linear model is used for the response as

$$Y_i = X_i \beta + b_i + \varepsilon_i,$$

where β is a vector consists of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$. The residuals ε_i are assumed to be independent normally distributed with zero means and covariance matrix V_i. The dropout process is modeled using the logistic model. Hence, we have 6 parameters of the response and the dropout model there is only one parameter (γ).

The developed stochastic EM algorithm is used to find the parameter estimates. The number of iterations is fixed at 3000. The results are presented in Table 3.

Depending on this simulation we can see that the absolute relative bias is small to moderate. The maximum relative bias is around 27%. This means that the proposed approach produce parameters estimates close to the true parameters values. Hence, we can conclude that the proposed approach is reliable and gives reasonable results.

References

[1] Rubin, D. B. "Inference and missing data". Biometrika, 63, 581-592. 1976.

[2] Little, R.J.A., Rubin, D.B., "Statistical Analysis with Missing Data",.Wiley, NewYork.1987.

[3] Wu, M. C. and Carroll, R. J. "Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process", Biometrics, 44, 175-188. 1988.

[4] Follmann, D. and Wu. M. "An approximate generalized linear model with random effects for informative missing Data", Biometrics, 51, 151- 168. 1995.

[5] Wu, M. C. and Bailey, K. R. "Estimation and comparison of changes in the presence of informative right censoring: conditional linear model", Biometrics, 45, 939-955. 1989.

[6] Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., and Landis, J. R. "Mixed effects logistic regression models for longitudinal binary response data with informative drop-out". Biometrics 54, 367-383. 1998.

[7] Pulkstenis, E.P., Ten Have, T. R. and Landis, J. R. "Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication", Journal of the American Statistical Association, 93, 438-450. 1998.

[8] Wu, M.C. and Follmann, D. A. "Use of summary measures to adjust for informative missingness in repeated measures data with random effects", Biometrics, 55, 75-84. 1999.

[9] Albert, P. S and Follumann, D.A., "A random effects transition model for longitudinal binary data with informative missigness", Statistica Neerlandica, 57, 100-111. 2003.

[10] Albert, P.S. and Follmann, D. A. "Modeling repeated count data subject to informative dropout", Biometrics, 56, 667-677. 2000.

[11] Dempster, A.P., Larid, N.M. and Rubin, D.B., "Maximum likelihood from incomplete data via the EM algorithm (with discussion)", Journal of Royal Statistical Society B, 39, 1-38. 1997.

[12] Tanner, M.A.,Wong,W.H., "The calculation of posterior distributions by data augmentation (with discussion)", Journal of American Statistical Association, 82, 528-550. 1987.

[13] Wei, G.C.G., Tanner, M.A., "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm," Journal Royal Statistical Society B, 55, 425-437. 1990.

[14] Celuex, G., Diebolt, J., "The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problems," Computational Statistics Quarterly, 2, 73-82. 1985.

[15] Delyon, B., Lavielle, M., Moulines, E., "Convergence of a stochastic approximation version of the EM Algorithm," Annals of Statistics, 27, 94-128. 1999.

[16] Diebolt, J., Ip, E.H.S., "Stochastic EM: method and application". In: Gilks,W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London. (Chapter 15). 1996.

[17] Gu, M.G., Kong, F.H., "A stochastic approximation algorithm with Markov chain Monte Carlo method for incomplete data estimation problems," Proc. Natl. Acad. Sci. USA 98, 7270-7274. 1998.

[18] Zhu, H.T., Lee, S.Y., "Analysis of generalized linear mixed models via a stochastic approximation algorithm with Markov chain Monte Carlo method," Statist. Comput., 12, 175-183. 2002.

[19] McLachlan, G.J., Krishnan, T., "The EM Algorithm and Extensions", Wiley, New York. 1997.

[20] Louis, T.A., "Finding the observed information matrix when using the EM algorithm". Journal of Royal Statistical Society, B 44, 226-232. 1982.

[21] Meilijson, I., "A fast improvement to the EM algorithm on its own terms", Journal of Royal Statistical Society, B 51, 127-138. 1989.

[22] Meng, X.L., Rubin, D.B., "Maximum likelihood estimation via the ECM algorithm: a general framework", Biometrika, 80, 267-278. 1993.

[23] Efron, B., "Missing data, imputation, and the bootstrap". Journal of American Statistical Association, 89, 463-475. 1994.

[24] Ip, E.H.S., "A stochastic EM estimator in the presence of missing data: theory and applications". Technical Report, Division of Biostatistics, Stanford University, Stanford, California, US. 1994.

[25] Tsonaka, R., Verbeke, G. and Lesaffre, E. "A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness", Biometrics 65, 81-87. 2009.

[26] Gad, A.M and Ahmed, A. S. "Analysis of longitudinal data with intermittent missing values using the stochastic EM algorithm". Computational Statistics & Data Analysis, 50, 2702-2714. 2006.

[27] McCullagh, P. and Nelder, J. A. "Generalized Linear Models". 2nd edititon, Chapman and Hall, England. 1989.

[28] Jennrich, R.I., Schluchter,M.D., "Unbalanced repeated measures models with structured covariance matrices". Biometrika 42, 805-820. 1986.

[29] Heyting, A. and Tolboom, J. T. B. M. and Essers, J. G. A. "Statistical handling of dropouts in longitudinal clinical trials", Statistics in Medicine, 11, 2043-2062. 1992.

[30] Diggle, P.J. and Kenward, M.G. "Informative dropout in longitudinal data analysis", Journal of Royal Statistical Society B, 43, 49-93. 1994.