



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



Storck, Kastner, Toutenburg:

Longitudinal data with dropouts: a comparison of  
pattern mixture models with complete case analysis

Sonderforschungsbereich 386, Paper 188 (2000)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



# Longitudinal data with dropouts: a comparison of pattern mixture models with complete case analysis

S. Storck<sup>1 0</sup>   C. Kastner<sup>1</sup>   H. Toutenburg<sup>2</sup>

March 16, 2000

## Abstract

Pattern mixture models constitute an alternative to selection models (Rubin, 1987). Little and Wang (1996) introduced pattern mixture models for analyzing multivariate normal longitudinal data with missing values. This paper was the theoretical foundation and the induce to investigate the small sample properties of pattern mixture models compared with complete case analysis. The main point of interest, of the simulations, was the mean square error of the estimated model parameters. Parameters estimated by the pattern mixture model are very satisfying under ignorable mechanism but they have to be scanned carefully under nonignorable mechanism.

## 1 Introduction

In longitudinal data often wave nonresponse occurs. In this case Maximum Likelihood (ML) is one possibility to deal with nonresponse (Little and Rubin, 1987). To factorize the likelihood two common approaches exist: selection models (SM) and pattern mixture models (PMM). Where the main interest is not the marginal mean averaged over the pattern, but the conditional mean of the completers together with the probability to complete or to drop out of the study, then PMM is the suitable approach to the problem. The typical feature of PMM is that the distribution of the missing mechanism only depends on the covariates and not on the outcome variable. This is caused by the way the likelihood is factorized.

A new class of pattern mixture models for multivariate incomplete data were first discussed in Little (1993) where missingness is assumed to depend on an

---

<sup>1</sup> Institute of Statistics, Ludwig-Maximilians University of Munich, Germany

<sup>0</sup> email: storck@stat.uni-muenchen.de

<sup>1</sup> Institute of Statistics, Ludwig-Maximilians University of Munich, Germany, email: kchris@stat.uni-muenchen.de

<sup>2</sup> Institute of Statistics, Ludwig-Maximilians University of Munich, Germany, email: toutenb@stat.uni-muenchen.de

arbitrary unspecified function of a linear combination of two variables. Little (1994) has a look on PMM in the bivariate normal case with no covariates. Little and Wang (1996) extended this approach to longitudinal data with wave nonresponse for multivariate normal data with fixed covariate matrix. The main focus is on the theoretical framework for ignorable missing mechanism (IM) and non-ignorable missing mechanism (NIM) and how to calculate the parameters of the model in each case. Introduced are ML methods, which were also integrated into the EM algorithm for over-identified models. In the following years further papers were published. Molenberghs, Michiels, Kenward and Diggle (1998) show that the taxonomy of missing data models can also applied to PMM. Fairclough, Peterson, Cella and Bonomi (1998) illustrates the use of several different model-based methods, including PMM, for different kind of missing mechanism based on real data. Ekholm and Skinner (1998) discusses the advantage and disadvantage of the PMM and SM by reanalyzing the set of binary, partially incomplete data from Woolson and Clarke (1984). Michiels, Molenberghs and Lipsitz (1999) have a look on the similarities and the differences of modeling incomplete data with SM and PMM. The main stress is on MAR. A special view on random-effect models accommodated to a pattern-mixture model is shown in the papers of Hedeker and Gibbons (1997). A further approach is to combined SM and PMM. This is discussed in Molenberghs, Michiels and Kenward (1998). A sensitivity analyze based on location and scale changes is discussed in Daniels and Hogan (1999)

The aim of this paper is to show how influence factors, which can not be influenced by the analyst, will influence the estimate of the parameters in the PMM. The subject of our investigation is the influence of correlation among the measure points, number of subjects and share of missing values. The survey is made for the ignorable and nonignorable nonresponse.

In Section 2 we will give an introduction to the theory of PMM. The structure of the simulation is described in Section 3. In Section 4 we present the results. A final discussion is given in Section 5.

## 2 Theory of Pattern Mixture Models

We want to analyze data from a random sample of  $n$  independent observations of  $T$  continuous variables  $Y = (y_1, \dots, y_T)'$  and  $p$  categorical covariates  $X = (x_1, \dots, x_p)$ , where  $X$  and a subset  $T_1$  of  $Y$ ,  $Y_1 = (y_1, \dots, y_{T_1})'$ , is completely observed for all  $n$  subjects and the remaining  $T_2 = T - T_1$  measure points,  $Y_2 = (y_{T_1+1}, \dots, y_T)'$  are observed for  $n_0$  and missing for  $n_1 = n - n_0$  subjects. Further we want to introduce the indicator variable  $m$ , with  $m = 0$  if  $y_{it}$  is observed for all  $t = 1, \dots, T$  and  $m = 1$  otherwise. This structure could appear in longitudinal data with one dropout point, but could also be generalized to a monotone missing pattern. So we divide our dataset twice with the indicator

vector  $m$ . Therefore we have four parts of datasets

$$\begin{pmatrix} Y_1^0 & Y_2^0 \\ Y_1^1 & Y_2^1 \end{pmatrix},$$

which could be characterized by their moments. The matrices indicated by 0 include all completers and the matrices indicated by 1 include subjects dropping out during the study. To define the different patterns indicated by  $m$  we introduce the scalar  $g \in 0, \dots, G - 1$  where  $G$  is the number of different missing patterns.

We specify the distribution of the missing mechanism and for every strata of missing pattern a distribution for  $Y$  given  $X$ :

$$\begin{aligned} (m | X) &\sim_{ind} \text{Bernoulli}(\pi(X)) \\ (y_i | X, m = g) &\sim_{ind} N_T(B^g X, \Sigma^g) \end{aligned}$$

with  $i = 1, \dots, n$ .  $\pi(X)$  is the probability for  $m = 1$  conditioned on the covariate strata.  $\pi$  is received by a standard logistic regression from  $m$  on  $X$ .  $(y_i | X, m = g)$  is a multivariate regression of  $Y$  on  $X$  for pattern  $g$  with regression coefficient matrix  $B^g$  and covariance matrix  $\Sigma^g$ .

Using ML methods we have to factorize the joint distribution of  $Y$  and  $m$  given  $X$ . There are two common approaches: the selection model

$$f(Y, m | X) = f(Y | X)f(m | Y, X) \quad (1)$$

and the pattern mixture model

$$f(Y, m | X) = f(m | X)f(Y | m, X). \quad (2)$$

In both the joint distribution of  $Y$  and the missing mechanism  $m$  has to be specified, but the way of factorizing this joint distribution is different. Pattern mixture models assume that  $f(Y | m, X)$  is a mixture over the missing patterns. This means that the parameters for the model have to be calculated separately for each pattern. An additional characteristic feature of this approach is that the distribution of the missing mechanism depends only on the covariates and not on  $Y$ . In contrast to this the selection model specifies the complete data distribution  $f(Y | X)$  and the missing mechanism  $f(m | Y, X)$  depends on  $Y$ .

To treat nonresponse in the correct way, we have to distinguish between ignorable nonresponse and nonignorable nonresponse (Rubin, 1987). Ignorable nonresponse exists, when missingness depends only on the covariates and the complete observed measure points:

$$P(m = 1 | Y_1, Y_2, X) = P(m = 1 | Y_1, X). \quad (3)$$

In this case ML estimate are straightforward and we are allowed to ignore the missing mechanism. Nonignorable nonresponse exists, when the missingness depends on the unobserved outcome variable:

$$P(m = 1 | Y_1, Y_2, X) = P(m = 1 | Y_2, X). \quad (4)$$

To get consistent estimates we have to specify the missing mechanism in a correct way otherwise the ML estimate will be biased.

The aim of PMM is to calculate the moments of the distribution averaged over the pattern. First the moments from the strata and time points, which weren't observed, have to be estimated, namely all those related to  $Y_2^1$ . After factorizing the distribution like (2) we are able to estimate  $f(m | X)$  but not  $f(Y | m, X)$ . If missing depends on  $X$  and  $Y_1$  (IM), we factorize

$$f(Y_1, Y_2 | X, m = g, \phi^g) = f(Y_2 | Y_1, X, m = g, \phi_{2.1}^g) \times f(Y_1 | X, m = g, \phi_1^g),$$

where  $\phi_1^g = (B_{1x.x}^g, \Sigma_{11.x}^g)$  contains the regression coefficient and the residual covariance matrix from the multivariate regression:  $Y_1$  on  $X$  and  $\phi_{2.1}^g = (B_{21.1x}^g, B_{2x.1x}^g, \Sigma_{22.1x}^g)$  contains the two regression coefficient matrices and the residual covariance matrix from multivariate regression  $Y_2$  on  $Y_1$  and  $X$ . In this case (3) is converted to  $\phi_{2.1}^0 = \phi_{2.1}^1$ . Now we are able to calculate the missing moments:

$$\begin{aligned} \hat{B}_{1x.x} &= (1 - \hat{p}_x) \hat{B}_{1x.x}^0 + \hat{p}_x \hat{B}_{1x.x}^1, \\ \hat{\Sigma}_{11.x} &= (1 - \hat{p}_x) \hat{\Sigma}_{11.x}^0 + \hat{p}_x \hat{\Sigma}_{11.x}^1 \\ &\quad + \hat{p}_x (1 - \hat{p}_x) (\hat{B}_{1x.x}^0 - \hat{B}_{1x.x}^1) X X' (\hat{B}_{1x.x}^0 - \hat{B}_{1x.x}^1)', \\ \hat{B}_{2x.x} &= \hat{B}_{2x.x}^0 + \hat{B}_{21.1x}^0 (\hat{B}_{1x.x}^1 - \hat{B}_{1x.x}^0), \\ \hat{\Sigma}_{21.x} &= \hat{\Sigma}_{21.x}^0 + \hat{B}_{21.1x}^0 (\hat{\Sigma}_{11.x}^1 - \hat{\Sigma}_{11.x}^0), \\ \hat{\Sigma}_{22.x} &= \hat{\Sigma}_{22.x}^0 + \hat{B}_{21.1x}^0 (\hat{\Sigma}_{11.x}^1 - \hat{\Sigma}_{11.x}^0) \hat{B}_{21.1x}^{0'}. \end{aligned} \quad (5)$$

where  $p_x = P(m = 1 | X, \hat{\pi})$  and  $\hat{\pi}$  is the ML estimate of  $\pi$ .

If missing depends on  $X$  and  $Y_2$  (NIM) we have to factorize  $f(Y | m, X)$  in the other way:

$$f(Y_1, Y_2 | X, m = g, \phi^g) = f(Y_1 | Y_2, X, m = g, \phi_{1.2}^g) \times f(Y_2 | X, m = g, \phi_2^g)$$

where  $\phi_2^g = (B_{2x.x}^g, \Sigma_{22.x}^g)$  contains the regression coefficient and the residual covariance matrix from the multivariate regression of  $Y_2$  on  $X$  and  $\phi_{1.2}^g = (B_{12.2x}^g, B_{1x.2x}^g, \Sigma_{11.2x}^g)$  contains the two regression coefficient matrices and the residual covariance matrix from multivariate regression of  $Y_1$  on  $Y_2$  and  $X$ . We convert restriction (4) to  $\phi_{1.2}^0 = \phi_{1.2}^1$ .

The structure of ML estimate now depends on the dimension of  $T_1$  and  $T_2$ . For  $T_1 = T_2$  ML estimate of  $B_{1x.x}$  and  $\Sigma_{11.x}$  are equal to (5) and ML estimates of the remaining parameters are

$$\begin{aligned} \hat{B}_{2x.x} &= \hat{B}_{2x.x}^0 + (\hat{B}_{12.2x}^0)^{-1} (\hat{B}_{1x.x}^1 - \hat{B}_{1x.x}^0), \\ \hat{\Sigma}_{21.x} &= \hat{\Sigma}_{21.x}^0 + (\hat{B}_{12.2x}^0)^{-1} (\hat{\Sigma}_{11.x}^1 - \hat{\Sigma}_{11.x}^0), \\ \hat{\Sigma}_{22.x} &= \hat{\Sigma}_{22.x}^0 + (\hat{B}_{12.2x}^0)^{-1} (\hat{\Sigma}_{11.x}^1 - \hat{\Sigma}_{11.x}^0) \left( \hat{B}_{12.2x}^{0'} \right)^{-1}. \end{aligned}$$

This type we want to call NIMED (nonignorable mechanism with equal dimensions). For  $T_1 > T_2$  we will have an overidentified model, which requires iterative methods. This type we call NIMUD (nonignorable mechanism with unequal dimensions). A convenient approach is the EM algorithm.  $C_1$  includes all subjects with pattern  $m = 1$ . The E-step at each iteration estimates the statistics in  $C_1$ , namely,

$$S_{12} = \sum_{i \in C_1} y_{1i} y'_{2i}, \quad S_{x2} = \sum_{i \in C_1} x_i y'_{2i}, \quad S_{22} = \sum_{i \in C_1} y_{2i} y'_{2i},$$

by their conditional expected values given the observed data and current estimates of the parameters. These quantities are output from the regression of  $Y_2$  on  $X$  and  $Y_1$  from cases in  $C_1$ , with parameters set at their current estimates for  $C_1$ . The M-step computes new parameters estimates by a complete-data maximization subject to the constraints induced by the missing-data assumption. In particular, for the restrictions of (4), parameters of the regression of  $Y_2$  on  $X$  (namely,  $\phi_2^0$  and  $\phi_2^1$ ) are computed separately for each pattern, and parameters  $\phi_{1.2}$  of the regression of  $Y_1$  on  $Y_2$  and  $X$  are computed pooled over pattern.

If  $T_1 < T_2$  then the model is not identified. We would need more restrictions to identify the model. One approach is  $P(m = 1 | Y_1, Y_2, X) = f(Y_{2*}, X)$ , where  $Y_{2*}$  is a subset of  $Y_2$  with dimension  $T_{2*} \leq T_1$ . In this case we can get the solution by extending the  $T_1 = T_2$  approach, if  $T_1 = T_{2*}$  and the  $T_1 > T_2$  approach, if  $T_1 > T_{2*}$ . Therefore we don't want to have a closer look to this case, because we can transform the structure of the data and treat it like explained above.

### 3 A simulation study

Aim of the survey was to investigate small sample properties of PMM compared with those of the complete case (CC) and the available case (AC) estimator. Therefore, the conduct of the estimators was investigated by changing the frame conditions of the data namely correlation structure between time points, number of subjects, missing mechanism and share of missing values. The second investigation unit became necessary, due to appearing problems in the nonignorable case. Further we investigate the small sample properties under misspecification with regard to the missing mechanism.

A useful tool for multivariate regression is the sweep-operator from Dempster (1969). Here we followed the suggestion of Little and Wang (1996) to implement the PMM. The algorithm is implemented using C++ (Stroustrup, 1997). Basic data structures were supplied by the library of Heumann, Fieger and Kastner (1998). Random numbers are generated using the DRAND48 generator, which is supplied by SunOS 5.5 as a C-library function (SunOS, 1995, man Pages(3C)).

Following the example in Little and Wang (1996) we create datasets with three treatment groups when we have 250 subjects and two treatment groups when we have 50 subjects. The regression parameters are dependent on the

time points and the treatment group. The distribution of the error term was multivariate normal. For details see Storck (1999).

After data generation missing values were produced using the function  $P(m = 1 | Y_1, Y_2, X) = aY_1 + bY_2 + cX$ . With the right specification of  $a$ ,  $b$  and  $c$  the missing mechanism mentioned above can be simulated. With every simulation adjustment we investigate the three missing mechanisms IM, NIMED and NIMUD. Every single simulation contains 1000 runs under the same adjustment. We want to give a short view on the different adjustments for the three investigation units.

Unit one has to compare CC, AC and PMM in regard to the MSE criterion. For that we combine the adjustments: number of subjects ( $n \in \{50, 250\}$ ), share of missing value ( $mis \in \{30\%, 60\%\}$ ), correlation ( $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ ). Hence we have 60 Simulations in our first unit.

One problem occurring during the simulation was that the NIMED type delivered very bad results when correlation was low. So we decided to calculate NIMED with the EM algorithm of NIMUD. This is subject of our second investigation with ten simulations. Possibilities of adjustments were  $n = 250$  with  $mis = 30\%$  or  $n = 50$  with  $mis = 60\%$  combine with  $\rho \in \{0.1, 0.2, 0.5, 0.7, 0.9\}$

In the third unit, containing 15 simulations, we were interested in what will happen if the a priori assumption is wrong. For this case we have chosen fixed assumptions with  $n = 250$ ,  $mis = 30\%$  and  $\rho = 0.5$  and combine those with the extent of misspecification. The misspecification will range from 0 (which means right assumption) to 1 (total wrong assumption). Between this intervall (namely 0.2, 0.4, 0.6, 0.8), missingness will depend on  $Y_1$  and  $Y_2$  in different weights.

We decided to choose the mean square error to compare the three estimators. This choice is theoretical based on the matrix-valued MSE criterion as discussed in full detail e. g. Rao and Toutenburg (1999). The MSE will cover as well bias as variance of the estimates. So we will not assess an estimator higher than another when he has little bias but too strong variance and the other way round. To calculate the MSE we used the equation:

$$E \left\{ \left( \hat{B} - B \right) \left( \hat{B} - B \right)' \right\},$$

where  $\hat{B}$  is the estimated regression parameter and  $B$  the known regression parameter.

## 4 Results

For AV and PMM there was no reason to complain for the estimate of the parameters for the completely observed time points, right through all missing mechanisms. For CC we could see, that only under low correlation and non-ignorable mechanism the estimate had a MSE equal zero. For the ignorable mechanism we notified that even at a very low correlation the CC has a high MSE, which will increase with a stronger correlation (see Figure 1). For the

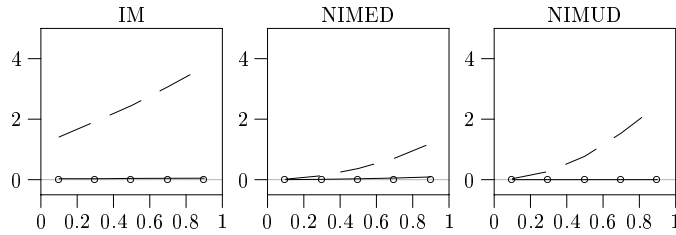


Figure 1: MSE of the full observed time points depending on the correlation. Investigation unit one with 250 subjects and 30% of missing values. line: PMM; dashed line: CC; circles: AC.

incomplete time points, as one can see in Figure 2, AV and CC deliver only under low correlation and IM assumption estimates with low MSE. If missingness depends on  $Y_2$  (NIMED and NIMUD) AC and CC have a stable high MSE independent on the correlation. The share of missing values make it even worse, but it seems that the number of subjects is not relevant for the height of the MSE in this case. PMM estimated the parameters of the incomplete time points to our complete satisfaction under IM. Even a low number of subjects with a high share of missing values didn't influence the MSE, it was always near to zero. In contrast the PMM did disappointingly on the estimates in the NIMED case. We notified, that the MSE will become close to zero only when the correlation grows to a particular strength. This was with 250 subjects about 0.3 and with 50 subjects about 0.7, independent on the share of missing values. Under this correlation barrier the PMM is even worse than the AC and CC model. A more positive idea of PMM we got under the NIMUD assumption, where the parameters were estimated via the EM algorithm. Under a low correlation we could see, that the PMM is of the same quality as the AC or CC estimates. But with increasing correlation PMM will get much better than the AC or the CC model. Even though we reached a MSE close to zero only under the very high correlation of 0.9.

Because of the very bad result in the NIMED case under low correlation structure in the data, we made the next investigation unit and tried to get better results, when we calculate the NIMED case with the EM algorithm of the NIMUD case. Here we just want to have look on the interesting incomplete time points. The statement we can make is that the EM algorithm is an advantage only in the case that we have a low correlation. In our model with 250 subjects and a share of missing values the correlation barrier is about 0.3, then the exact estimate method will be better than the asymptotic one (see Figure 3).

Another important question was what will happen, when we make the wrong assumption about the missing mechanism. The estimate of the complete time points is not affected by this assumption, so we have only a look on the incomplete time points. Starting with IM we could see (Figure 4) that the MSE of the PMM was increasing while we intensified the wrong assumption, nevertheless the MSE was always smaller than those of AC and CC. At the NIMED case we



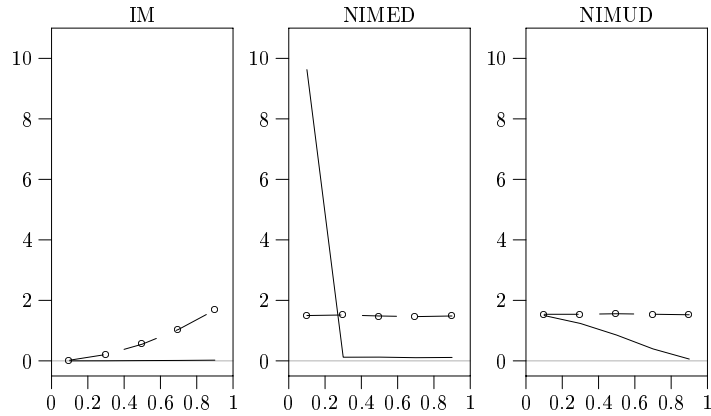


Figure 2: MSE of the incomplete time points depending on the correlation. Investigation unit one with 250 subjects and 30% of missing values. Line: PMM; dashed line: CC; circles: AC.

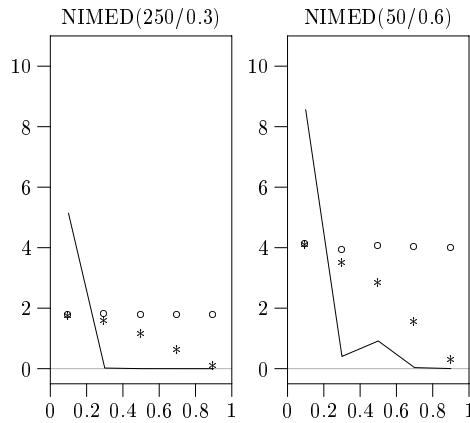


Figure 3: MSE of the incomplete time points depending on the correlation using the EM algorithm for NIMED situation. Investigation unit two compares simulations having 250 subjects and 30% of missing values with 50 subjects and 60% of missing values. Line: PMM; stars: PMM with EM algorithm; circles: AC.

notified a very strong reaction with respect to the wrong assumption. Already with a medium misspecification of the missing mechanism we had a worse MSE than AC and CC. For the NIMUD type of missing we can remark, that a wrong assumption have YEAR = 1994 only a positive effect on the MSE. As these simulations run under the fixed adjustments of  $n = 250$ ,  $\rho = 0.5$  and  $mis = 0.3$  we could proceed from the assumption that an increasing correlation or share of missing values have the same effects on the estimate quality like described above.

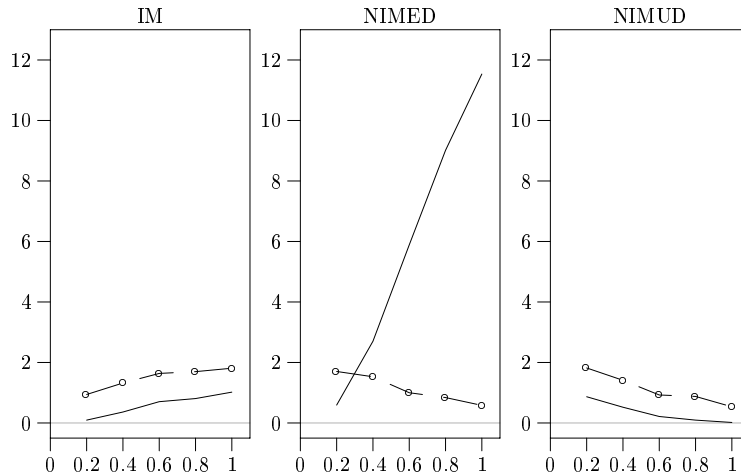


Figure 4: MSE of the incomplete time points depending on the extent of misspecification of the missing mechanism. Investigation unit three with 250 subjects and 30% of missing values. Line: PMM; dashed line: CC; circles: AC.

## 5 Discussion

Under the IM assumption, the PMM delivers very satisfying estimates. We could see that the PMM estimator is unbiased independent of the correlation structure of the data. Problems arise with nonignorable mechanisms. The PMM under NIMED is only after a correlation barrier better than AC and CC. So we also discussed the idea to use the EM algorithm for low correlations, what seems to be reasonable. Not perfect but better is PMM under the NIMUD assumption, and we notified that the MSE will come close to zero only under high correlation among the measure points. Nevertheless PMM is always better than AC or CC in this case.

Another approach to calculate the NIMED case would be to use the probit selection model (Little and Wang (1996)). Perhaps this estimator would be more stable for low correlation structures. The disadvantage of probit models is the difficulty to fit them, because  $\Lambda$  (from the assumption  $P(m = 1 | Y_1, Y_2, X) = f(CY_1 + \Lambda Y_2, X)$ ) has, together with the other parameters, to be estimated. This might be difficult because the information about  $\Lambda$ , delivered from the data, is very weak (Brown, 1990).

It should be mentioned, that in the literature selection models are expected to be more efficient when  $\Lambda$  is correct specified (Little, 1994). This fact could be observed during the simulations, but only for the complete observed time points, because then AC is equal to SM estimator.

With a small number of subjects we could see that the estimate became increasingly inefficient. This is a reason to think of changing the estimator. It seems that a Bayesian approach is the right choice for the problem. Little and Wang suggested the Bayesian analyses on Gibbs sampling.

Another source of error is the prior assumption about the missing mecha-

nism for choosing the right model. At first the analyst should get an idea of the correlation structure of the data. This is of course not an easy task because we only can use the complete cases for this analysis. The next step is to make oneself realize what kind of missing mechanism in the data appears. This could be managed by plausible causal connections which describe what missing mechanism is responsible for the dropouts. These first reflections are important to judge the estimator. If we are sure to have a high correlation among the measure points, then wrong assumption about the missing mechanism cannot have a huge effect on the estimates in the IM and NIMUD case. In the NIMED case we should be more careful. Here the estimator gets very quick worse than AC or CC estimator. Having shaky assumptions we should think of taking the EM algorithm at the price of efficiency.

This model holds many worthy extension possibilities. It would be desirable to handle the PMM with more different kind of data and missing patterns. So further work has to be done.

## References

- Brown, C. H. (1990). Protecting against nonrandomly missing data in longitudinal studies, *Biometrics* **46**: 143–157.
- Daniels, M. J. and Hogan, J. W. (1999). A new parameterization for pattern mixture models of longitudinal data with informative dropout, *Technical report*, Department of Statistics, Iowa State University, Department of Statistics, Iowa State University, 102G Snedecor Hall, Ames, IA 50011.
- Dempster, A. P. (1969). *Elements of Continuous Multivariate Analysis*, Addison-Wesley, Reading, Massachusetts.
- Ekholm, A. and Skinner, C. J. (1998). The Muscatine children’s obesity data reanalysed using pattern mixture models, *Applied Statistics* **47**: 251–263.
- Fairclough, D. L., Peterson, H. F., Cella, D. and Bonomi, P. (1998). Comparison of several model-based methods for analyzing incomplete quality of life data in cancer clinical trials, *Statistics in Medicine* **17**: 781–796.
- Hedeker, D. and Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies, *Psychological Methods* **2**: 64–78.
- Heumann, C., Fieger, A. and Kastner, C. (1998). C++ Utilities zur Implementierung statistischer Verfahren unter Berücksichtigung fehlender Werte, *SFB386 – Discussion paper 109*, Ludwig-Maximilians-Universität München.
- Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data, *Journal of the American Statistical Association* **88**: 125–133.

- Little, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data, *Biometrika* **81**: 471–483.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.
- Little, R. J. A. and Wang, Y.-X. (1996). Pattern-mixture models for multivariate incomplete data with covariates, *Biometrics* **52**: 98–111.
- Michiels, B., Molenberghs, G. and Lipsitz, S. R. (1999). Selection models and pattern-mixture models for incomplete data with covariates, *Biometrics* **55**: 978–983.
- Molenberghs, G., Michiels, B. and Kenward, M. G. (1998). Pseudo-Likelihood for Combined Selection and Pattern-Mixture Models for Incomplete Data, *Biometrical Journal* **40**: 557–572.
- Molenberghs, G., Michiels, B., Kenward, M. G. and Diggle, P. J. (1998). Monotone missing data and pattern-mixture models, *Statistica Neerlandica* **52**: 153–161.
- Rao, C. R. and Toutenburg, H. (1999). *Linear Models: Least Squares and Alternatives*, 2 edn, Springer, New York.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Sample Surveys*, Wiley, New York.
- Storck, S. (1999). *Pattern Mixture Modelle für fehlende Werte bei korrelierten Beobachtungen*, Diplomarbeit, Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80535 München, Germany.
- Stroustrup, B. (1997). *Die C++ Programmiersprache*, Addison-Wesley, Bonn.
- SunOS (1995). *SunOS Reference Manual*, Sun Microsystems, Mountain View.
- Woolson, R. F. and Clarke, W. R. (1984). Analysis of categorical incomplete longitudinal data, *Journal of the Royal Statistical Society, Series A* **147**: 87–99.