Ziegler, Kastner:

Solving Generalised Estimating Equations With Missing Data Using Pseudo Maximum Likelihood Estimation Is Equivalent to Complete Case Analysis

Sonderforschungsbereich 386, Paper 128 (1998)

Solving Generalised Estimating Equations With Missing Data Using Pseudo Maximum Likelihood Estimation Is Equivalent to Complete Case Analysis

A. Ziegler * C. Kastner **

September 2, 1998

Abstract

Arminger and Sobel (1990) proposed an approach to estimate mean- and covariance structures in the presence of missing data. These authors claimed that their method based on Pseudo Maximum Likelihood (PML) estimation may be applied if the data are missing at random (MAR) in the sense of Little and Rubin (1987). Rotnitzky and Robins (1995), however, stated that the PML approach may yield inconsistent estimates if the data are (MAR). We show that the adoption of the PML approach for mean- and covariance structures to mean structures in the presence of missing data as proposed by Ziegler (1994) is identical to the complete case (CC) estimator. Nevertheless, the PML approach has the computational advantage in that the association structure remains the same.

Keywords: Correlated Data Analysis, Generalised Estimating Equations, Marginal Models, Missing Data, Pseudo Maximum Likelihood

1 Introduction

Arminger and Sobel (1990) introduced an approach to estimate mean- and covariance structures in the presence of missing data. Their method is based on Pseudo Maximum Likelihood (PML) estimation proposed by Gourieroux, Monfort and Trognon (1984). The approach of Arminger and Sobel (1990) was adopted by Ziegler (1994) to the analysis of mean structures including the generalised estimating equations (GEE) first proposed by Liang and Zeger (1986). For an overview s. e.g. Ziegler, Kastner and Blettner (1998). Arminger and Sobel (1990) claimed that their approach may be applied if the data are missing at random (MAR) in the sense of Little and Rubin (1987). Rotnitzky and Robins (1995), however, stated that the PML approach with missing data may lead

*Medical Centre for Methodology and Health Research, Institute of Medical Biometry and Epidemiology, Philipps–University of Marburg, Bunsenstr. 3, 35033 Marburg, Germany, ziegler@mailer.uni-marburg.de

**Institute of Statistics, LMU München, Ludwigstr. 33, 80539 München, Germany, kchris@stat.uni-muenchen.de

to inconsistent parameter estimates when the data are MAR but not missing completely at random (MCAR) in Little and Rubin's (1987) sense. The proof of their statement was left to the reader.

In this paper we show that the PML estimator for incomplete data as proposed by Arminger and Sobel (1990) and adopted by Ziegler (1994) to mean structure models is in fact equivalent to the complete case (CC) estimator. For this purpose we use a simple example for estimating the mean structure. Our results also hold for mean- and covariance structure models.

The outline of the paper is as follows. In section 2 we introduce the PML model using the multivariate normal density to construct the Pseudo-log-likelihood function. In section 3 the PML approach for missing data is introduced. Finally, in section 4 the PML technique is illustrated using a response vector with two responses and two missing observations.

2 The Pseudo Maximum Likelihood Approach

Suppose that the correctly specified marginal mean of the $T \times 1$ response vector y_i of individual i given the $n \times p$ matrix of explanatory variables X_i is given by

$$E(y_i|X_i) = \mu(X_i\beta) = \mu_i,$$

where μ is the response function and β is the $p \times 1$ unknown parameter vector of interest. The true conditional covariance matrix $\Omega_i(X_i)$ of y_i given X_i is unknown and need not be parameterised in β. A "good guess" of the true covariance matrix is used instead. This covariance matrix denoted by Σ_i may depend on the explanatory variables X_i as well as on the parameter vector β and an additional association parameter α which is treated as nuisance. Σ_i is usually called working covariance matrix.

Regardless of the true distribution of y_i given X_i we choose the multivariate normal distribution to estimate the parameter vector β. Hence, we assume that

$$y_i|X_i \sim N\Big(\mu(X_i\beta), \Sigma_i\Big).$$

A solution $\hat{\beta}$ is found by maximizing the kernel of the Pseudo-log-likelihood function

$$l(\beta) = -\sum_{i=1}^{n} l_i(\beta) = -\sum_{i=1}^{n} \Big(y_i - \mu_i\Big)' \Sigma_i^{-1} \Big(y_i - \mu_i\Big) = -\sum_{i=1}^{n} \mathrm{tr}\,\Big(\Sigma_i^{-1} S_i\Big),$$

where $S_i = (y_i - \mu_i)(y_i - \mu_i)'$. This corresponding score equations for β are given by

$$0 = \sum_{i=1}^{n} D_i' \Sigma_i^{-1} \Big(y_i - \mu_i\Big)$$

which are identical to GEE for the mean structure.

To account for the probably incorrectly specified distribution of y_i given X_i, the so-called robust variance matrix (s. e.g. Ziegler et al., 1998) is used to estimate the covariance matrix of $\hat{\beta}$ Gourieroux et al. (1984).

If missing data are present, a classical approach is to use the CC estimator. This takes into account only those observations t within an individual i that are completely observed for all explanatory variables used to model the marginal mean of the response y_{it} at time t. In addition, y_{it} also has to be available. This approach is applicable, if the data are MCAR. One computational problem arises in the currently available implementations (s. e.g. Ziegler and Grömping, 1998) of this approach. Rows that suffer from missing data are internally deleted. Hence, covariance structures Σ_i that require the exact positions of the entire observations cannot be estimated adequately: If, for example, observations at time points 1 and 3 are available for an autoregressive model of order 1 (AR1), these are treated as observations 1 and 2 by the currently available commercial programs. These observations contribute with a factor ϱ instead of ϱ^2 to an estimate of the correlation. Thus, the structural change usually effects the interpretation of the association parameter α which is usually underestimated in these situations (s. e.g. Ziegler and Grömping, 1998). It also may effect the efficiency of the estimate of β.

3 Pseudo Maximum Likelihood Estimation With Missing Data

Arminger and Sobel (1990) have shown that the PML method based on the normal theory may be applied, even if missing data in the explanatory variables and / or the response variables are present. In the following we assume w.l.o.g. that both the dependent and the explanatory variables are either missing or present.

Analogously to Arminger and Sobel (1990) we define a matrix $\tilde{K}_i$ that selects the variables observed for the ith sample element. The number of rows in $\tilde{K}_i$ is equal to the length of observed units u; the number of columns in $\tilde{K}_i$ is T. Hence, the number of rows in $\tilde{K}_i$ can vary for each individual. For computation of first and second derivatives it is useful to make the number of rows equal to T. For this purpose a row of zeros is filled in for each unobserved variable. The new $T \times T$ matrix is denoted by K_i, and the matrix $I - K_i$ by $\bar{K}_i$ where I is the $T \times T$ unit matrix.

Using the same arguments as Arminger and Sobel (1990) for mean- and covariance structure models, Ziegler (1994) has shown for mean structure models that the estimator based on the kernel of the Pseudo-log-likelihood function

$$\tilde{l}(\beta) = -\sum_{i=1}^{n} \operatorname{tr}\left(\Sigma_i^{\star -1} S_i^\star\right)$$

with

$$\Sigma_i^\star = K_i \Sigma_i K_i + \bar{K}_i$$

and

$$S_i^\star = K_i S_i K_i + \bar{K}_i$$

is proportional to the estimator obtained from $l(\beta)$ using the incomplete responses.

4 Illustration

In this section we illustrate the results obtained in the previous section using observations from a single person. This individual was observed at time points 1 and 3, while the data for time points 2 and 4 are missing.

For this person, $\tilde{K}_i$ is a 2×4 matrix, while K_i and $\bar{K}_i$ are 4×4 matrices:

$$\tilde{K}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad K_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \bar{K}_i = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$S_i^\star$ and $\Sigma_i^\star$ are given by

$$S_i^\star = \begin{pmatrix} (y_{i1}-\mu_{i1})^2 & 0 & (y_{i1}-\mu_{i1})(y_{i3}-\mu_{i3}) & 0 \\ 0 & 1 & 0 & 0 \\ (y_{i3}-\mu_{i3})(y_{i1}-\mu_{i1}) & 0 & (y_{i3}-\mu_{i3})^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\Sigma_i^\star = \begin{pmatrix} \sigma_1^2 & 0 & \sigma_{13} & 0 \\ 0 & 1 & 0 & 0 \\ \sigma_{13} & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Simple algebra yields

$$\Sigma_i^{\star -1} = \begin{pmatrix} \frac{\sigma_3^2}{\sigma_1^2\sigma_3^2-\sigma_{13}^2} & 0 & \frac{-\sigma_{13}}{\sigma_1^2\sigma_3^2-\sigma_{13}^2} & 0 \\ 0 & 1 & 0 & 0 \\ \frac{-\sigma_{13}}{\sigma_1^2\sigma_3^2-\sigma_{13}^2} & 0 & \frac{\sigma_1^2}{\sigma_1^2\sigma_3^2-\sigma_{13}^2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

so that the kernel of the Pseudo-log-likelihood function of this individual is given by

$$\begin{aligned} \tilde{l}_i(\beta) &= -\operatorname{tr}\left(\Sigma_i^{\star -1} S_i^\star\right) = \frac{1}{\sigma_1^2\sigma_3^2-\sigma_{13}^2} \times \\ &\quad \left(\sigma_3^2(y_{i1}-\mu_{i1})^2 - 2\sigma_{13}(y_{i1}-\mu_{i1})(y_{i3}-\mu_{i3}) + \sigma_1^2(y_{i3}-\mu_{i3})^2\right) + 2 \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \tilde{l}_i(\beta)}{\partial \beta} &= \frac{-1}{\sigma_1^2\sigma_3^2-\sigma_{13}^2} \\ &\quad \left[2\sigma_3^2(y_{i1}-\mu_{i1})\left(-\frac{\partial \mu_{i1}}{\partial \beta}\right)\right. \\ &\quad -2\sigma_{13}\left((y_{i1}-\mu_{i1})\left(-\frac{\partial \mu_{i3}}{\partial \beta}\right) + \left(-\frac{\partial \mu_{i1}}{\partial \beta}\right)(y_{i3}-\mu_{i3})\right) \\ &\quad \left. +2\sigma_1^2(y_{i3}-\mu_{i3})\left(-\frac{\partial \mu_{i3}}{\partial \beta}\right)\right] = \end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{2}{\sigma_1^2\sigma_3^2-\sigma_{13}^2}\left[\sigma_3^2(y_{i1}-\mu_{i1})\frac{\partial\mu_{i1}}{\partial\beta}\right. \\
& -\sigma_{13}\left((y_{i1}-\mu_{i1})\frac{\partial\mu_{i3}}{\partial\beta}+\frac{\partial\mu_{i1}}{\partial\beta}(y_{i3}-\mu_{i3})\right) \\
& \left.+\sigma_1^2(y_{i3}-\mu_{i3})\frac{\partial\mu_{i3}}{\partial\beta}\right]
\end{aligned}
$$

is the responding score function of this individual. In matrix notation we get

$$
2\,D_i'\Sigma_i^{\star-1}\begin{pmatrix} y_{i1}-\mu_{i1} \\ 0 \\ y_{i3}-\mu_{i3} \\ 0 \end{pmatrix}, \qquad (1)
$$

where $D_i = \partial\mu_i/\partial\beta'$. Equation (1) shows that the CC estimating equations and the PML estimating equations are proportional yielding identical estimators. Hence, the PML approach garantuees consistent parameter estimates for data that are MCAR but not MAR.

Acknowledgment

For partial research support, the first author thanks the Sonderfoschungsbereich 386 at the Ludwig-Maximilians University of Munich, where he was a visiting researcher in July 1998. This work was supported by the Deutsche Forschungsgemeinschaft.

References

Arminger, G. and Sobel, M. E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data, *Journal of the American Statistical Association* **85**: 195–203.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory, *Econometrica* **52**: 681–700.

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

Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*, Wiley, New York.

Rotnitzky, A. G. and Robins, J. M. (1995). Semiparametric estimation of models for means and covariances in the presence of missing data, *Scandinavian Journal of Statistics* **22**: 323–333.

Ziegler, A. (1994). *Verallgemeinerte Schätzgleichungen zur Analyse korrelierter Daten*, PhD thesis, Fachbereich Statistik, Universität Dortmund.

Ziegler, A. and Grömping, U. (1998). The generalised estimating equations: A comperison of procedures available in commercial software packages, *Biometrical Journal* **40**: 245–260.

Ziegler, A., Kastner, C. and Blettner, M. (1998). The generalised estimating equations: An annotated bibliography, *Biometrical Journal* **40**: 115–139.