Scheid:

A Selection Model for Bivariate Normal Data, with a Flexible Nonparametric Missing Model and a Focus on Variance Estimates

Sonderforschungsbereich 386, Paper 285 (2002)

A Selection Model for Bivariate Normal Data, with a Flexible Nonparametric Missing Model and a Focus on Variance Estimates

S. Scheid

August 2, 2002

Abstract

Nonignorable nonresponse is a common problem in bivariate or multivariate data. Here a selection model for bivariate normal distributed data (Y_1, Y_2) is proposed. The missingness of Y_2 is supposed to depend on its own values. The model for missingness describes the probability of nonresponse in dependency of Y_2 itself and it is chosen nonparametrically to allow flexible patterns. We try to get a reasonable estimate for the expectation and especially for the variance of Y_2. Estimation is done by data augmentation and computation by common sampling methods.

Keywords: bivariate normal distributed data, variance estimate, selection model, nonparametric missing model, data augmentation

1 Introduction

In some applications beside the expectation itself the estimated variance of a random variable is of particular interest. This may, for example, be the case in a pharmacological study if we ask for reliability of a certain treatment. A further example is given by income studies of a population where the variability defines the gaps between income classes. In this article we propose a model for bivariate Gaussian variables (Y_1, Y_2), where informative missingness of Y_2 occurs. Flexible missing models shall be allowed with the aim to estimate expectations and variances correctly. Missingness is assumed to depend on Y_2. The joint distribution of (Y_1, Y_2, R), where R measures whether Y_2 is observed, is factorized according to a selection model. For the definition of selection models see Little and Rubin (1987), p. 218ff. The model that measures whether Y_2 is missing is chosen nonparametrically. It is a Bayesian Model that specifies a function in Y_2 with piecewise constant function values. Neighbored values are smoothed by an underlying first order random walk. Nonparametric models of this type are treated extensively in Fahrmeir and Lang (2001). The estimation of the selection model is done by the help of data augmentation.

In the literature selection models are treated by numberous authors. The term selection model dates, at least, back to to the well-known sample selection model of Heckman (1976). A Bayesian Tobit Model is developed in Cowles, Carlin and Connett (1996). Models for non-ignorable missing categorical response in longitudinal data situations have been treated e.g. by Conaway (1993), Park and Brown (1994) and Baker (1995), the case of general multivariate categorical data e.g. by Fitzmaurice, Clifford and Heath (1996) and Park (1998). For log-linear models see e.g. Chambers and Welsch (1993), for transitional models for binary data see Albert (2000) and for generalized linear mixed models (GLMMs) see Ibrahim, Chen and Lipsitz (2001). Nonignorable missing continuous (Gaussian) response data in longitudinal data in case of monotone missingness is treated by Diggle and Kenward (1994). Kenward (1998) presents a sensitivity analysis, while Troxel, Harrington and Lipsitz (1998) and Troxel, Lipsitz and Harrington (1998) treat the nonmonotone case. An extensive treatment of selection models in the context of linear mixed models is given in Verbeke and Molenberghs (2000). As far as we know, nonparametric methods for the missing model have not been developed.

In the following section, the model will be described. The third section deals with computational aspects. Finally, an implementation of the model is tested in Section 4.

2 Model

We assume the random vector $Y = (Y_1, Y_2)^t$ to be Gaussian with expectation $\mu = (\mu_1, \mu_2)^t$ and covariance matrix Σ. Whereas Y_1 is always observed, Y_2 may be missing. The random indicator variable R shall take the value 0 if Y_2 is missing and 1 (Y_2 is reported) else. The joint distribution can be factorized as

$$P(Y_1, Y_2, R) = P(Y_1, Y_2)P(R \mid Y_1, Y_2).$$

This common type of factorization is called a selection model (Rubin 1987). Suppose that missingness depends on Y_2 and is independent of Y_1 given Y_2, i.e.,

$$P(R \mid Y_1, Y_2) = P(R \mid Y_2) \ .$$

Usually, one is interested in an estimate of the expectation μ_2 of Y_2 and, especially, in the case where the probability of $R = 0$ is monotone in Y_2 itself. Here, for example, a logit model for $P(R \mid Y_2)$ with a predictor that is linear in Y_2 can be chosen. We want to focus now on the case where the probability of $R = 0$ (Y_2 is missing) may for example be high for extreme (i.e. for high and low) values of Y_2 and low in the midrange of Y_2. A logit model with a linear predictor Y_2 is no longer appropriate in this case. As an illustration, imagine an income panel of only two time points with missing response on the second time point. In econometrics a common perspective is that people with very low and people with very high income refuse to answer about their income. Moreover, it seems more likely that missingness depends on the actual income or at least on a function of the actual income than on the previous one. To answer the question how many people are poor, not only the expected income but also the variance is of interest. Note that for our application it is important, that after transformation (e.g. logarithmization) we can assume Y_1, Y_2 to be of the assumed distribution, i.e., Gaussian.

To address our problem we choose a logit model for R with a smooth function in Y_2

$$P_{\mu,\Sigma,\Theta}(Y_1, Y_2, R) = P_{\mu,\Sigma}(Y_1, Y_2)P_\Theta(R \mid Y_2) \qquad \Theta = (\beta, \sigma^2) \ ,$$

$$P_\Theta(R = 0 \mid Y_2) = \text{logit}^{-1}(f(Y_2)) = \frac{1 + \exp(f(Y_2))}{\exp(f(Y_2))}$$

where $f(Y_2)$ is explained in the further context.

Since our missing model depends on values of Y_2 that are actually not observed, common estimation routines fail. To solve this problem we choose a technique called data augmentation (see Tanner (1991)).

Since data augmentation is a Bayesian method, we first have to choose appropriate prior distributions for the unknown parameters. Furthermore, we want to specify the introduced function in Y_2, $f(Y_2)$.

For the expectation of μ we choose a diffuse prior,

$$P(\mu) \propto \text{const.}$$

A noninformative prior for Σ (in the sense of Jeffrey; Rüger (1999)) is the limiting distribution of an inverse Wishart distribution

$$W^{-1}(m, \Lambda), m \to -1, \ \Lambda^{-1} \to 0 \ .$$

Together, μ and Σ follow (see e.g. Schafer (1997))

$$P(\mu, \Sigma) \propto \mid \Sigma \mid^{-(\frac{3}{2})} \ .$$

To construct the smooth function (compare Fahrmeir and Lang (2001)), we choose an interval $I = [a, b)$, which shall contain the main part of the observations of Y_2.

This interval is now devided into $k - 1$ small intervals of the same length, i.e.,

$$I = [a, b) = [a = a_1, a_2) \cup [a_2, a_3) \cup \ldots \cup [a_{K-1}, a_K = b) \ .$$

Note: for estimation in Section 4 we chose a= -2.5, b=2.5, k=100. Let I_0 be the interval $(-\infty, a_1)$,

$$I_i = [a_i, a_{i+1}) \text{ for } i = 1, \ldots, K - 1$$

and

$$I_K = [a_K, \infty) \ .$$

Our smooth function consists of a logit model that specifies a different function value for each interval I_i. To make the estimation of the function values $f(y_2)$,

$y_2 \in I_i$, $i = 0, ..., K$ stable, neighbored function values are assumed to be similar. This fact is introduced by an informative prior for the coefficients β_i (that address the function values) of indicator functions that measure whether an observation lies in the interval I_i, or not. To be more concrete we specify the linear predictor of our logit model as

$$\eta_i = \beta_0(y_{2i} \in I_0) + \beta_1(y_{2i} \in I_1) + \cdots + \beta_K(y_{2i} \in I_K)$$

where y_{2i} denotes the $i-$th observation of Y_2.

The model logit $^{-1}(\eta)$ depends on the coefficients β_i. For the prior distribution of the β_i's we choose a random walk. In our computations we choose a first order random walk. This random walk can be expressed due to the conditional distributions of β_i given the other β's as

$$\begin{aligned} P(\beta_i \mid \beta_{i-1}, \beta_{i+1}) &\sim N(\tilde{\mu}_i, \sigma^2) \ , \\ \tilde{\mu}_i &= \frac{\tilde{\mu}_{i-1} + \tilde{\mu}_{i+1}}{2} \quad \text{for } i = 1, \ldots, K-1 \ , \\ \tilde{\mu}_0 &= \tilde{\mu}_1 \ , \\ \tilde{\mu}_K &= \tilde{\mu}_{K-1} \ . \end{aligned}$$

We assume $\mu_0 \propto$ const. The prior distribution $P(\beta \mid \sigma^2)$ for $\beta = \{\beta_0, \beta_1, ..., \beta_K\}$ still depends on σ^2. A common choice for $P(\sigma^2)$ is an inverse Gamma distribution.

We assume $P(\beta \mid \sigma^2)P(\sigma^2)$ to be independent of $P_{\mu,\Sigma,\Theta}(Y_1, Y_2, R)$ and independent of $P(\mu, \Sigma)$. On the other hand, $P(\mu, \Sigma)$ shall be independent of $P(\beta \mid \sigma^2)P(\sigma^2)$ and $P_{\mu,\Sigma,\Theta}(R \mid Y_2)$.

The joint distribution now can be written as

$$P_{\mu,\Sigma}(Y_1, Y_2)P_\Theta(R \mid Y_2)P(\mu, \Sigma)P(\beta \mid \sigma^2)P(\sigma^2) \ .$$

To get estimates for μ and Σ, values for the missing Y_2 are imputed and the standard estimates

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix}$$

$$\widehat{\Sigma}_k = \frac{1}{n}\sum_{i=1}^{n} (y_i - \widehat{\mu}_i)(y_i - \widehat{\mu}_i)^t$$

with

$$y_i = \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix}$$

are applied to the observations of (Y_1, Y_2). By the data augmentation procedure, randomly drawn values are computed. We can repeat the imputation procedure

(l times) and take the means of the single estimates for μ and Σ as new estimates themselves:

$$\widehat{\mu} = \frac{1}{l}\sum_{k=1}^{l}\widehat{\mu}_k$$

$$\widehat{\Sigma} = \frac{1}{l}\sum_{k=1}^{l}\widehat{\Sigma_k} \ .$$

Data augmentation consists of two steps that are applied consecutively. That is, the I(mputation)–Step where values for the missing values are drawn given the observed values and actual parameter estimates. In the other step, called P(robability)–Step we calculate new estimates for our parameters, as we draw from the posteriors, using the new imputed values beside the observed ones.

In our case this looks as follows:

- I–Step

  Draw from $P(Y_2 \mid y_{1i}, R = 0, \mu, \Sigma, \beta, \sigma^2)$ for all missing y_{2i}.

- P–Step

  Draw $\beta, \sigma^2 \sim P(\beta, \sigma^2 \mid Y_1, Y_2, R, \mu, \Sigma \propto P_\Theta(R \mid Y_2)P(\beta, \sigma^2)$ (see independence assumptions from above).

  Draw $\mu, \Sigma \sim P(\mu, \Sigma \mid Y_1, Y_2, R, \beta, \sigma^2 \propto P_{\mu,\Sigma}(Y_1, Y_2)P(\mu, \Sigma)$ (see independence assumptions from above).

3 Computation

We start by describing the P–Step in which we draw from the posteriori distributions. Further, we assume that first starting values for our missing y_{2i}'s are chosen, for example, simply the mean of the observed y_{2i}'s.

As $P(\beta, \sigma^2 \mid Y_1, Y_2, R, \mu, \Sigma) \propto P_\Theta(R \mid Y_2)P(\beta, \sigma^2)$ and $P(\mu, \Sigma \mid Y_1, Y_2, R, \beta, \sigma^2) \propto P_{\mu,\Sigma}(Y_1, Y_2)P(\mu, \Sigma)$ do not depend on each other, from the posteriors of β, σ^2 and μ, Σ can be drawn separately.

Since our prior for Σ is from a conjugate family, drawing from the posteriors of μ and Σ can be done straightforwardly as

$$\Sigma \mid Y_1, Y_2 \sim W^{-1}(n-2, S^{-1}),$$

where S^{-1} denotes the empirical covariance matrix of the completed data (y_{1i}, y_{2i}), $i = 1, \ldots, n$, and

$$\mu \mid \Sigma, Y_1, Y_2 \sim N(\widehat{\mu}, \Sigma),$$

where $\widehat{\mu}$ denotes the empirical mean of the $(y_{1i}, y_{2i})^t$'s, with the actual imputed values for the missing y_{2i}'s.

To draw $P(\beta, \sigma^2 \mid Y_1, Y_2, R, \mu, \Sigma)$ we use Gibbs–sampling, combined with Metropolis steps to update the β_i's.

The following steps are applied several times until we can assume that the values for β and σ^2 correspond to the target distribution (examine for example the autocorrelation).

1. Values for the conditional posterior distribution of σ^2 given values for all β_i can be drawn staightforwardly. If the priori of σ^2 is $IG(a, b)$, than the postrior is $IG(a', b')$ with

   $a' = a + K$

   $b' = b + \frac{1}{2}(\beta_0^2 + \beta_k^2 + 2\sum_{i=1}^{k-1} \beta_i^2 - 2\sum_{i=0}^{k-1} \beta_i \beta_{i+1})$.

   For our computation we chose quite informative values for the inverse Gamma distribution of the variance, σ^2, i.e. a= 5, b= 3. This seems to be appropriate as very low or high probabilities in logit models are very unlikely to be measured empirically correct.

2. Make proposals for $(\beta_i, \dots, \beta_{i+k})$, $i = 1, \dots, n-k$ (in our implementation we chose $k = 2$), starting with $i = 1$ and continuouing consecutively, accept $(\beta_i, \dots, \beta_{i+k})^{\text{new}}$ at each step with probability

$$\frac{P_\Theta(R \mid Y_2) P((\beta_{i-1}, \dots, \beta_{i+k+1})^{\text{new}}, \sigma^2)}{P_\Theta(R \mid Y_2) P((\beta_{i-1}, \dots, \beta_{i+k+1})^{\text{old}}, \sigma^2)} \wedge 1.$$

   $P((\beta_{i-1}, \dots, \beta_{i+k+1})^{\text{new}}, \sigma^2)$ denotes the full conditional distribution in the case of our first order random walk (compare Knorr-Held (1999)).

As a proposal distribution for the $(\beta_i, \dots, \beta_{i+k}), i = 1, \dots, n-k$ we choose a random walk with multivariate normal changes with mean $(0, ..., 0)$ and a covariance matrix with no correlation and small variances. After a pre–run we calculate the empirical variances of our accepted β's and continue with a normal distribution with mean $(0, ..., 0)$ and the empirical covariances.

The I–Step is computed in the following way:

For each missing y_{2i} we draw $Y_2^{\text{prop.}} \sim P_{\mu,\Sigma}(Y_2 \mid Y_1)$ and $R \sim P_\Theta(R \mid Y_2^{\text{prop.}})$ with the actual values for the parameters μ, Σ and Θ. For the same y_{2i} repeat drawing $Y_2^{\text{prop.}}$ and R until $R = 0$. If succeeded we choose $Y_2^{\text{prop.}}$ as our new imputation for y_{2i} and continue with the next missing y_{2i}.

4 Validation

Depending on the data situation it might occur that more than one parameter setting will describe the data situation in an appropriate manner. In these cases the original distribution would be unable to be reconstructed exactly. Our Bayesian estimation will then be more or less the mean of the distributions that may fit to the data.

From a pragmatic point of view we want to see what we can expect about our model in realistic situations. To do so, we created data sets, where the part of missing y_{2i}'s in the mean does not exceed 50%. The actual percentage depends on the missing function. The following six functions were chosen to create the data:

$P_\Theta(R = 0|Y_2)$ =logit$^{-1}(\eta)$, where

1.	$\eta = -1.5 + 0.75 \cdot Y_2^2$	
2.	$\eta = -1.5 + (3.0 \exp(Y_2^2) - 1.0)/53.6))$	
3.	$\eta = -1.5 \cos(3.14159 \cdot Y_2/2)$	$Y_2 \in [-2, 2]$
	$\eta = -1.5$	$Y_2 \in (-\infty, -2)$
	$\eta = 1.5$	$Y_2 \in (2, \infty)$
4.	$\eta = -9/8 + 9/16Y_2^2 + 2/3Y_1$	
5.	$\eta = -9/8 + 3/4(3.0 \exp(Y_2^2) - 1.0)/53.6)) + 2/3Y_1$	
6.	$\eta = -9/8 \cos(3.14159 \cdot Y_2/2) + 2/3Y_1$	$Y_2 \in [-2, 2]$
	$\eta = -9/8 + 2/3Y_1$	$Y_2 \in (-\infty, -2)$
	$\eta = 9/8 + 2/3Y_1$	$Y_2 \in (2, \infty)$

For the set–up, we chose functions that show low values for the middle range and high values for the extremes. As nonparametric function estimates are often sensitive at the edges, the chosen functions show a different behavior. Function 1 rises in quadratic terms, function 2 in exponential terms, whereas function 3 does not rise at the very edges. Functions 4 to 6 are the same as functions 1 to 3, but with an additional linear term. For functions 4 to 6, the estimation of $E(Y_2)$ in addition is of special interest.

The functions are illustrated below.

For the the normal distribution the parameters

$\mu_1 = \mu_2 = 0$

$\text{Var}(Y_1) = \text{Var}(Y_2) = 1$, and

$\text{Cov}(Y_1, Y_2) = 0.5, 0.7, 0.9$ were chosen. We draw samples of the sizes

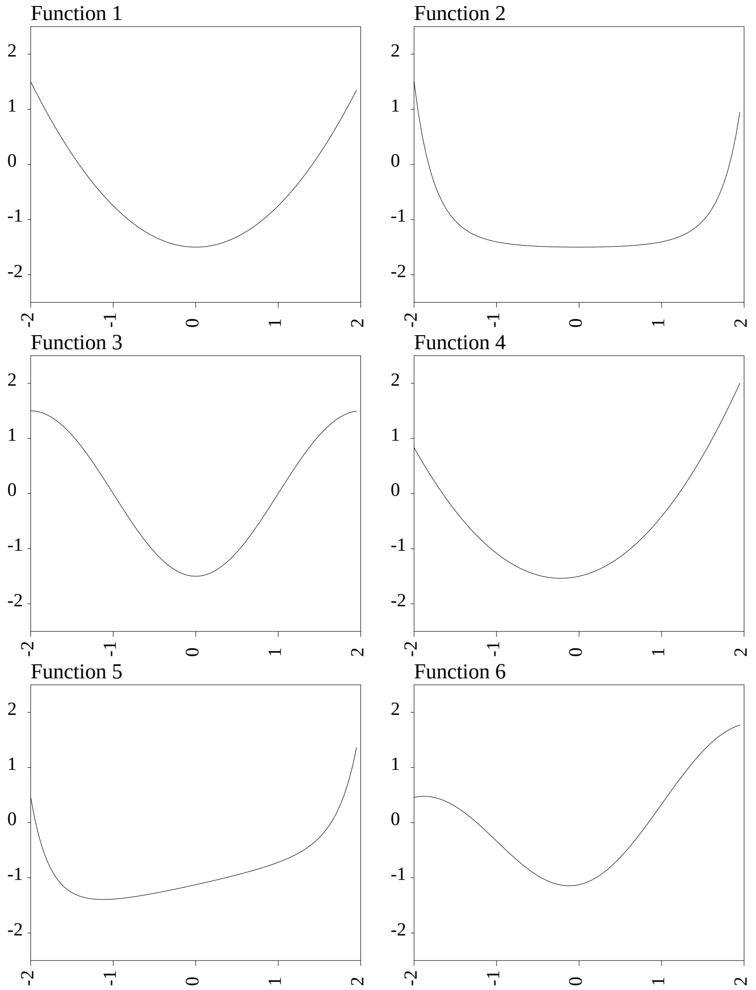
$N = 400, 700, 1000$ or 2000.

For each setting we draw ten replications. This number is related to the time the computations needed. Nevertheless, we can get some impression on the results to be expected.

The means of the missing percentages of y_{2i}'s over the replications are shown in the Table below.

To get a comparison we additionally estimated the distribution of the bivariate normal distribution under the assumption of an underlying MAR mechanism. We used the same algorithm, but accepted each value drawn for Y_2 in the I-Step. The computation of the missing model is thus not required.

The estimation results are illustrated in the Tables 4.2 to 4.13 which show the means and the standard deviations of the the interesting parameters of the ten replications for each setting.



		Function 1	Function 2	Function 3	Function 4	Function 5	Function 6
$N = 400$	$\rho = 0.5$	33	24	41	33	30	42
$N = 400$	$\rho = 0.7$	32	25	41	32	30	42
$N = 400$	$\rho = 0.9$	32	25	41	33	31	41
$N = 700$	$\rho = 0.5$	33	24	41	32	30	43
$N = 700$	$\rho = 0.7$	33	25	40	34	31	42
$N = 700$	$\rho = 0.9$	34	24	42	32	30	44
$N = 1000$	$\rho = 0.5$	34	24	40	32	30	41
$N = 1000$	$\rho = 0.7$	34	25	39	32	30	43
$N = 1000$	$\rho = 0.9$	33	23	40	34	32	43
$N = 2000$	$\rho = 0.5$	33	25	41	33	29	43
$N = 2000$	$\rho = 0.7$	33	25	41	33	30	42
$N = 2000$	$\rho = 0.9$	33	25	41	33	31	42

Table 4.1: **Mean percentage of missing Y_2.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	1.05	0.488	0.00102	0.459	0.184	0.0763
$N = 400$	$\rho = 0.7$	0.844	0.597	0.0234	0.202	0.135	0.061
$N = 400$	$\rho = 0.9$	1.51	1.1	-0.0328	0.898	0.319	0.079
$N = 700$	$\rho = 0.5$	1.11	0.505	-0.0267	0.432	0.181	0.046
$N = 700$	$\rho = 0.7$	1.15	0.764	0.014	0.282	0.14	0.0544
$N = 700$	$\rho = 0.9$	0.975	0.879	-0.00576	0.0914	0.0609	0.0352
$N = 1000$	$\rho = 0.5$	0.892	0.454	0.0128	0.251	0.112	0.0477
$N = 1000$	$\rho = 0.7$	0.985	0.677	0.00254	0.182	0.0907	0.046
$N = 1000$	$\rho = 0.9$	1.18	0.996	-0.0202	0.171	0.0981	0.0321
$N = 2000$	$\rho = 0.5$	0.784	0.397	-0.0115	0.195	0.0877	0.0444
$N = 2000$	$\rho = 0.7$	0.95	0.678	0.000277	0.102	0.0563	0.0214
$N = 2000$	$\rho = 0.9$	0.986	0.89	-0.00882	0.0827	0.051	0.021

Table 4.2: **Results for Function 1 using the selection model.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.608	0.341	0.00357	0.0387	0.0337	0.0394
$N = 400$	$\rho = 0.7$	0.648	0.502	0.0157	0.0513	0.0622	0.0561
$N = 400$	$\rho = 0.9$	1.2	0.987	-0.0208	0.781	0.314	0.0632
$N = 700$	$\rho = 0.5$	0.61	0.328	0.00314	0.0451	0.0441	0.0168
$N = 700$	$\rho = 0.7$	0.7	0.57	0.0135	0.0657	0.0702	0.0416
$N = 700$	$\rho = 0.9$	0.805	0.791	-0.0065	0.0586	0.0502	0.0307
$N = 1000$	$\rho = 0.5$	0.62	0.346	-0.00166	0.0227	0.0226	0.0433
$N = 1000$	$\rho = 0.7$	0.659	0.514	0.0137	0.0471	0.0424	0.0249
$N = 1000$	$\rho = 0.9$	0.93	0.881	-0.0163	0.114	0.0841	0.0272
$N = 2000$	$\rho = 0.5$	0.613	0.328	-0.00127	0.0249	0.0179	0.0281
$N = 2000$	$\rho = 0.7$	0.661	0.528	-0.00279	0.0258	0.0247	0.0197
$N = 2000$	$\rho = 0.9$	0.808	0.797	-0.00633	0.0588	0.0438	0.0199

Table 4.3: **Results for Function 1 assuming MAR.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	1.1	0.542	-0.0315	0.312	0.129	0.0683
$N = 400$	$\rho = 0.7$	1.09	0.731	-0.00992	0.31	0.123	0.0797
$N = 400$	$\rho = 0.9$	1.01	0.89	0.0275	0.0974	0.0653	0.0669
$N = 700$	$\rho = 0.5$	1.08	0.508	-0.0168	0.24	0.116	0.088
$N = 700$	$\rho = 0.7$	1.06	0.707	-0.0163	0.155	0.0974	0.0467
$N = 700$	$\rho = 0.9$	1.03	0.919	0.0257	0.0916	0.0633	0.0506
$N = 1000$	$\rho = 0.5$	1.17	0.553	0.0182	0.25	0.103	0.0391
$N = 1000$	$\rho = 0.7$	1.01	0.693	0.000303	0.107	0.0709	0.046
$N = 1000$	$\rho = 0.9$	1.04	0.922	0.00247	0.115	0.0789	0.0303
$N = 2000$	$\rho = 0.5$	1.14	0.542	0.0134	0.198	0.0767	0.0298
$N = 2000$	$\rho = 0.7$	1.02	0.708	-0.00175	0.115	0.0714	0.0274
$N = 2000$	$\rho = 0.9$	1	0.901	0.0154	0.0695	0.05	0.0209

Table 4.4: **Results for Function 2 using the selection model.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.725	0.414	-0.0106	0.0399	0.0572	0.0435
$N = 400$	$\rho = 0.7$	0.809	0.618	-0.0083	0.0721	0.0508	0.055
$N = 400$	$\rho = 0.9$	0.889	0.832	0.0244	0.0518	0.0474	0.0597
$N = 700$	$\rho = 0.5$	0.753	0.394	-0.00341	0.0398	0.0496	0.0662
$N = 700$	$\rho = 0.7$	0.779	0.58	-0.0179	0.0753	0.0739	0.0378
$N = 700$	$\rho = 0.9$	0.914	0.862	0.0235	0.0647	0.0547	0.0485
$N = 1000$	$\rho = 0.5$	0.748	0.406	0.00809	0.0307	0.033	0.0203
$N = 1000$	$\rho = 0.7$	0.781	0.586	0.00334	0.0303	0.0371	0.0333
$N = 1000$	$\rho = 0.9$	0.903	0.855	0.00265	0.0814	0.0674	0.0271
$N = 2000$	$\rho = 0.5$	0.752	0.394	0.00322	0.0287	0.0228	0.0179
$N = 2000$	$\rho = 0.7$	0.794	0.602	-0.000947	0.0464	0.05	0.0177
$N = 2000$	$\rho = 0.9$	0.882	0.84	0.0136	0.0453	0.041	0.0191

Table 4.5: **Results for Function 2 assuming MAR.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.919	0.435	0.0368	0.615	0.25	0.0778
$N = 400$	$\rho = 0.7$	1.06	0.675	-0.0283	0.491	0.192	0.054
$N = 400$	$\rho = 0.9$	1.1	0.915	-0.0155	0.438	0.211	0.0576
$N = 700$	$\rho = 0.5$	1.08	0.502	-0.0161	0.363	0.155	0.0832
$N = 700$	$\rho = 0.7$	0.972	0.655	0.000319	0.255	0.132	0.0466
$N = 700$	$\rho = 0.9$	1.05	0.919	0.00092	0.124	0.0727	0.0463
$N = 1000$	$\rho = 0.5$	1.04	0.524	-0.00513	0.307	0.136	0.0424
$N = 1000$	$\rho = 0.7$	1.21	0.773	0.00129	0.404	0.172	0.054
$N = 1000$	$\rho = 0.9$	1.15	0.953	0.00719	0.399	0.189	0.0193
$N = 2000$	$\rho = 0.5$	1.16	0.574	-0.00374	0.196	0.0767	0.0235
$N = 2000$	$\rho = 0.7$	0.987	0.689	0.00312	0.158	0.084	0.0273
$N = 2000$	$\rho = 0.9$	1.02	0.897	-0.00125	0.189	0.108	0.0242

Table 4.6: **Results for Function 3 using the selection model.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.575	0.326	-0.00548	0.0907	0.0765	0.0391
$N = 400$	$\rho = 0.7$	0.594	0.48	-0.00812	0.0814	0.0625	0.0309
$N = 400$	$\rho = 0.9$	0.818	0.789	-0.0084	0.222	0.148	0.0457
$N = 700$	$\rho = 0.5$	0.58	0.318	-0.0078	0.0615	0.0489	0.032
$N = 700$	$\rho = 0.7$	0.619	0.488	0.00162	0.051	0.0615	0.0237
$N = 700$	$\rho = 0.9$	0.795	0.794	0.00645	0.061	0.0507	0.0361
$N = 1000$	$\rho = 0.5$	0.573	0.337	-0.00166	0.0223	0.0298	0.0145
$N = 1000$	$\rho = 0.7$	0.639	0.531	-0.00114	0.0442	0.0613	0.0397
$N = 1000$	$\rho = 0.9$	0.818	0.803	0.0109	0.183	0.13	0.0141
$N = 2000$	$\rho = 0.5$	0.583	0.34	0.00216	0.0263	0.0219	0.0149
$N = 2000$	$\rho = 0.7$	0.609	0.496	0.00326	0.0427	0.0412	0.0216
$N = 2000$	$\rho = 0.9$	0.776	0.771	0.00158	0.108	0.0826	0.0172

Table 4.7: **Results for Function 3 assuming MAR.**

		$\widehat{VarY_2}$	$\widehat{Cov(Y_1,Y_2)}$	$\widehat{EY_2}$	$\sigma(\widehat{VarY_2})$	$\sigma(\widehat{Cov(Y_1,Y_2)})$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.953	0.421	-0.0394	0.402	0.167	0.0887
$N = 400$	$\rho = 0.7$	1.22	0.771	-0.00154	0.449	0.221	0.0796
$N = 400$	$\rho = 0.9$	1.27	1	0.0189	0.589	0.271	0.0695
$N = 700$	$\rho = 0.5$	0.808	0.411	-0.0409	0.305	0.144	0.0429
$N = 700$	$\rho = 0.7$	1.22	0.796	0.0119	0.234	0.0955	0.0652
$N = 700$	$\rho = 0.9$	0.941	0.848	-0.0134	0.174	0.118	0.0337
$N = 1000$	$\rho = 0.5$	0.762	0.4	-0.0335	0.237	0.115	0.0479
$N = 1000$	$\rho = 0.7$	1.06	0.728	-0.00397	0.17	0.0823	0.0336
$N = 1000$	$\rho = 0.9$	1.04	0.92	0.00296	0.068	0.0387	0.0247
$N = 2000$	$\rho = 0.5$	0.831	0.434	0.000309	0.147	0.0677	0.0314
$N = 2000$	$\rho = 0.7$	0.924	0.661	-0.0253	0.0748	0.0439	0.0414
$N = 2000$	$\rho = 0.9$	1.04	0.922	0.00456	0.138	0.082	0.0351

Table 4.8: **Results for Function 4 using the selection model.**

		$\widehat{VarY_2}$	$\widehat{Cov(Y_1,Y_2)}$	$\widehat{EY_2}$	$\sigma(\widehat{VarY_2})$	$\sigma(\widehat{Cov(Y_1,Y_2)})$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.621	0.311	-0.084	0.0515	0.0455	0.0376
$N = 400$	$\rho = 0.7$	0.706	0.571	-0.0472	0.0766	0.0862	0.0438
$N = 400$	$\rho = 0.9$	1	0.897	-0.00673	0.413	0.236	0.0601
$N = 700$	$\rho = 0.5$	0.613	0.33	-0.0883	0.0309	0.0467	0.0346
$N = 700$	$\rho = 0.7$	0.713	0.572	-0.0474	0.0573	0.0465	0.0388
$N = 700$	$\rho = 0.9$	0.785	0.766	-0.0334	0.108	0.0919	0.0352
$N = 1000$	$\rho = 0.5$	0.611	0.344	-0.0872	0.034	0.0402	0.0264
$N = 1000$	$\rho = 0.7$	0.677	0.543	-0.0547	0.0265	0.028	0.0194
$N = 1000$	$\rho = 0.9$	0.835	0.815	-0.0217	0.0444	0.0319	0.0233
$N = 2000$	$\rho = 0.5$	0.624	0.343	-0.0691	0.0235	0.0272	0.0208
$N = 2000$	$\rho = 0.7$	0.655	0.519	-0.0682	0.0291	0.0298	0.0274
$N = 2000$	$\rho = 0.9$	0.85	0.827	-0.0179	0.0874	0.0663	0.0292

Table 4.9: **Results for Function 4 assuming MAR.**

		$\overline{\widehat{VarY_2}}$	$\overline{\widehat{Cov(Y_1, Y_2)}}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(\widehat{Cov(Y_1, Y_2)})$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.985	0.459	-0.0256	0.391	0.166	0.0778
$N = 400$	$\rho = 0.7$	1.07	0.706	-0.000414	0.224	0.103	0.0817
$N = 400$	$\rho = 0.9$	1.14	0.953	0.00441	0.307	0.177	0.0348
$N = 700$	$\rho = 0.5$	1.06	0.507	0.00878	0.334	0.129	0.0469
$N = 700$	$\rho = 0.7$	1.07	0.716	0.0121	0.178	0.106	0.0399
$N = 700$	$\rho = 0.9$	1.03	0.913	0.0216	0.0823	0.0574	0.0455
$N = 1000$	$\rho = 0.5$	0.858	0.43	-0.0146	0.188	0.0908	0.0547
$N = 1000$	$\rho = 0.7$	0.951	0.654	0.00662	0.184	0.126	0.0336
$N = 1000$	$\rho = 0.9$	0.983	0.884	-0.00771	0.155	0.105	0.0359
$N = 2000$	$\rho = 0.5$	1.03	0.516	0.00508	0.155	0.0661	0.0235
$N = 2000$	$\rho = 0.7$	1.01	0.705	-0.00186	0.0723	0.0401	0.0221
$N = 2000$	$\rho = 0.9$	0.964	0.876	0.00228	0.063	0.0431	0.0202

Table 4.10: **Results for Function 5 using the selection model.**

		$\overline{\widehat{VarY_2}}$	$\overline{\widehat{Cov(Y_1, Y_2)}}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(\widehat{Cov(Y_1, Y_2)})$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.738	0.383	-0.0658	0.0579	0.0564	0.0408
$N = 400$	$\rho = 0.7$	0.815	0.599	-0.0524	0.0889	0.0621	0.0597
$N = 400$	$\rho = 0.9$	0.985	0.887	-0.0173	0.232	0.15	0.0296
$N = 700$	$\rho = 0.5$	0.754	0.401	-0.0426	0.0408	0.0295	0.0215
$N = 700$	$\rho = 0.7$	0.801	0.595	-0.034	0.057	0.0649	0.0217
$N = 700$	$\rho = 0.9$	0.9	0.85	-0.0036	0.0651	0.0526	0.0407
$N = 1000$	$\rho = 0.5$	0.731	0.384	-0.0633	0.0369	0.0294	0.0271
$N = 1000$	$\rho = 0.7$	0.761	0.566	-0.036	0.0496	0.0622	0.0229
$N = 1000$	$\rho = 0.9$	0.89	0.837	-0.023	0.135	0.1	0.0354
$N = 2000$	$\rho = 0.5$	0.764	0.41	-0.0521	0.0259	0.0181	0.0226
$N = 2000$	$\rho = 0.7$	0.781	0.595	-0.0438	0.0256	0.0245	0.0224
$N = 2000$	$\rho = 0.9$	0.868	0.826	-0.0161	0.0505	0.0398	0.0193

Table 4.11: **Results for Function 5 assuming MAR.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.987	0.47	0.0146	0.301	0.133	0.111
$N = 400$	$\rho = 0.7$	0.94	0.607	0.0262	0.431	0.207	0.0696
$N = 400$	$\rho = 0.9$	1.15	0.933	-0.00154	0.48	0.228	0.0535
$N = 700$	$\rho = 0.5$	1.09	0.534	0.00678	0.282	0.117	0.0429
$N = 700$	$\rho = 0.7$	1.01	0.703	-0.00182	0.256	0.132	0.0369
$N = 700$	$\rho = 0.9$	1.01	0.889	-0.000183	0.245	0.133	0.0535
$N = 1000$	$\rho = 0.5$	0.95	0.46	0.00366	0.249	0.113	0.0603
$N = 1000$	$\rho = 0.7$	0.999	0.693	-0.0162	0.217	0.107	0.0323
$N = 1000$	$\rho = 0.9$	1.07	0.931	0.0257	0.234	0.128	0.0252
$N = 2000$	$\rho = 0.5$	0.888	0.465	-0.0193	0.157	0.0718	0.0371
$N = 2000$	$\rho = 0.7$	0.951	0.677	-0.0109	0.131	0.0762	0.0321
$N = 2000$	$\rho = 0.9$	0.945	0.86	-0.0105	0.123	0.0788	0.0191

Table 4.12: **Results for Function 6 using the selection model.**

		$\overline{\widehat{VarY_2}}$	$\overline{Cov(\widehat{Y_1, Y_2})}$	$\overline{\widehat{EY_2}}$	$\sigma(\widehat{VarY_2})$	$\sigma(Cov(\widehat{Y_1, Y_2}))$	$\sigma(\widehat{EY_2})$
$N = 400$	$\rho = 0.5$	0.626	0.334	-0.0912	0.0817	0.0444	0.0524
$N = 400$	$\rho = 0.7$	0.649	0.49	-0.0641	0.0558	0.0747	0.0465
$N = 400$	$\rho = 0.9$	0.892	0.829	-0.0388	0.249	0.161	0.0506
$N = 700$	$\rho = 0.5$	0.68	0.374	-0.103	0.0499	0.0331	0.0507
$N = 700$	$\rho = 0.7$	0.703	0.559	-0.0939	0.0623	0.0651	0.0216
$N = 700$	$\rho = 0.9$	0.839	0.803	-0.0307	0.155	0.105	0.0499
$N = 1000$	$\rho = 0.5$	0.648	0.342	-0.094	0.0487	0.0461	0.0153
$N = 1000$	$\rho = 0.7$	0.705	0.555	-0.078	0.0533	0.0435	0.025
$N = 1000$	$\rho = 0.9$	0.892	0.845	-0.00426	0.153	0.105	0.0232
$N = 2000$	$\rho = 0.5$	0.649	0.359	-0.113	0.0441	0.0355	0.0248
$N = 2000$	$\rho = 0.7$	0.711	0.555	-0.0739	0.0431	0.0371	0.0277
$N = 2000$	$\rho = 0.9$	0.811	0.788	-0.0387	0.0819	0.0638	0.0183

Table 4.13: **Results for Function 6 assuming MAR.**

5 Results and Conclusions

The estimates for $\text{Var}(Y_2)$ using the selection model lie around the origin value of 1. Taking a closer look at the estimates for function 2, all lie above 1. If in contrast we use the MAR estimator, the estimates seem to be biased downwards. The estimates for $\text{E}(Y_2)$ using the selection model, lie around zero for all functions. In contrast, for functions 4 to 6, the MAR estimates all lie below zero.

Concluding, in case of an informative random missing process R the proposed method of data augmentation is demonstrated to be superiour to estimation under the MAR assumption.

References

Albert, P. S. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data, *Biometrics* (56): 602–608.

Baker, S. G. (1995). Marginal regression for repeated binary data with outcome subject to non-ignorable non-response, *Biometrics* **51**: 1042–1052.

Chambers, R. L., and Welsch, A. H. (1993). Log-linear models for survey data with non-ignorable non-response, *Journal of the Royal Statistical Society, Series B* **55**(1): 157–170.

Conaway, M. R. (1993). Non-ignorable non-response models for time-ordered categorical variables, *Applied Statistics* **42**: 105–115.

Cowles, M. K., Carlin, B. P., and Connett, J. E. (1996). Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness, *Journal of the American Statistical Association* **91**(433): 86–98.

Diggle, P. J., and Kenward, M. G. (1994). Informative dropout in longitudinal data analysis, *Applied Statistics* **43**: 49–94.

Fahrmeir, L., and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on markov random field priors, *Applied Statistics* **50**(2): 201–220.

Fitzmaurice, G. M., Clifford, P., and Heath, A. F. (1996). Logistic regression models for binary panel data with attrition, *Journal of the Royal Statistical Society, Series A* **159**(2): 249–263.

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 Economic and Social Measurement* **5**: 475–492.

Ibrahim, J. G., Chen, M.-H., and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable, *Biometrika* **88**(2): 551–564.

Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity, *Statistics in Medicine* **17**: 2723–2732.

Knorr-Held, L. (1999). Conditional prior proposals in dynamic models, *Scandinavian Journal of Statistics* **26**: 129–144.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

Park, T. (1998). An approach to categorical data with nonignorable nonresponse, *Biometrics* **54**(4): 1579–1590.

Park, T., and Brown, M. B. (1994). Models for categorical data with nonignorable nonresponse, *Journal of the American Statistical Association* **89**(425): 44–52.

Rüger, B. (1999). *Test- und Schätztheorie. Band I: Grundlagen*, Oldenbourg, München.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*, Chapman and Hall, London.

Tanner, M. A. (1991). *Tools for Statistical Inference*, Springer–Verlag, New York.

Troxel, A. B., Harrington, D. P., and Lipsitz, S. R. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values, *Applied Statistics* **47**(3): 425–438.

Troxel, A. B., Lipsitz, S. R., and Harrington, D. P. (1998). Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data, *Biometrika* **85**(3): 661–672.

Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*, Springer, New York.