Didelez:

Maximum Likelihood and Semiparametric Estimation in Logistic Models with Incomplete Covariate Data

Sonderforschungsbereich 386, Paper 110 (1998)

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

Maximum Likelihood and Semiparametric Estimation in Logistic Models with Incomplete Covariate Data

By Vanessa Didelez

University of Munich, Institute of Statistics,

Ludwigstr. 33, D-80539 Munich, Germany

SUMMARY

Maximum likelihood estimation of regression parameters with incomplete covariate information usually requires a distributional assumption about the concerned covariates which implies a source of misspecification. Semiparametric procedures avoid such assumptions at the expense of efficiency. A simulation study is carried out to get an idea of the performance of the maximum likelihood estimator under misspecification and to compare the semiparametric procedures with the maximum likelihood estimator when the latter is based on a correct assumptions.

KEY WORDS: Logistic regression; Maximum likelihood; EM algorithm; Missing covariates; Missing data; Semiparametric efficient.

1. Introduction

The problem of coping with incomplete information in the covariates when estimating a regression parameter is common in applied work. A simple solution is given by the complete case analysis where all incomplete cases are discarded. The resulting complete case estimator, however, is obviously inefficient and not generally consistent under the missing at random assumption (MAR), which excludes dependence of the observability of a covariate on its unobserved value (Rubin, 1976). Within the more sophisticated methods two main approaches can be distinguished and will be compared in this paper. These are the parametric one which consists in specifying

the covariate distribution and thus allows for likelihood inference and the semiparametric one which avoids any distributional assumption about the covariates, only the original regression model is specified.

Full parametric procedures have been proposed for instance by Little (1992), Blackhurst and Schluchter (1989), Ibrahim (1990), and Ibrahim and Weisberg (1992). Usually the distributional assumption concerns the conditional distribution of the incomplete covariate given a subset of or all the other covariates and the response variable. The resulting maximum likelihood estimator is asymptotically efficient if the specification of this conditional covariate distribution is correct. Misspecification is likely to occur when restrictive assumptions are inevitable as for example when one of the involved variables is continuous. Little (1992) and Ibrahim and Weisberg (1992) assume in this situation a Gaussian covariate distribution. There is no apparent reason why this standard assumption should be correct and nothing is known so far about its 'robustness' against misspecification. In our simulation study we investigate the behaviour of a maximum likelihood estimator which assumes a Gaussian covariate distribution while the true distribution is Student or χ^2 representing serious violations of the standard assumption.

Semiparametric procedures have been intensively investigated in the last years. For the situation of two–stage case–control studies Breslow and Cain (1988) propose a pseudoconditional likelihood approach yielding a consistent estimator under the MAR assumption. It has been shown (Cain and Breslow, 1988; Vach and Illi, 1997) that in the situation of a logistic regression model this turns out to be a simple modification of the complete case estimator. Another approach is to use the empirical distribution or nonparametric kernel estimates to estimate the unknown distribution. This has been proposed by Pepe and Flemming (1991) and Carroll and Wand (1991) in the context of mismeasured covariates but the resulting estimators for the regression parameter are only consistent if the missing mechanism is MAR and does not depend on the response variable. Reilly and Pepe (1995) apply the same idea in order to estimate the score function for incomplete observations. Their so–called mean score estimator is consistent under MAR. Robins et al. (1994) and Robins et al. (1995) address the performance of such semiparametric estimators by considering

the lower variance bound of any regular semiparametric estimator. The estimator that attains this variance bound, however, usually depends on the unknown covariate distribution. For rather general cases they describe adaptive semiparametric efficient estimators which are feasible without this knowledge. The simulation study conducted here gives an idea of the gain in efficiency of this estimator compared with the mean score and the pseudoconditional likelihood methods. In addition, it allows to assess the performance of the semiparametric efficient estimator compared to the maximum likelihood estimator with correct assumption about the covariate distribution for finite sample size.

The outline of the paper is as follows. We restrict ourselves to the situation of a logistic regression model which is mostly used in practice when the response is binary. In Section 2 we describe this model and the missing situation to which we apply the different estimators which in turn are presented in Section 3. The considered estimators are the complete case, the Breslow–and–Cain, the mean score, the semiparametric efficient, and the maximum likelihood estimator. The simulation designs are given in Section 4, the results of the simulation study in Section 5. Finally, we discuss the obtained results.

2. The Model

We compare the different approaches to estimate a regression parameter along the special case of a logistic regression. Let Y denote a binary response variable, X_1 a completely observed binary covariate and X_2 an incompletely observed continuous covariate. The logistic regression model is given by the assumption that

$$\Pr(Y = 1|X_1 = x_1, X_2 = x_2; \beta) = \frac{\exp\{\beta_0 + \beta_1 x_1 + \beta_2 x_2\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \beta_2 x_2\}}, \tag{1}$$

where $\beta = (\beta_0, \beta_1, \beta_2)^\top$ is the parameter vector to be estimated. For ease of notation we also write $\Pr(y|x_1, x_2; \beta)$ instead of $\Pr(Y = y|X_1 = x_1, X_2 = x_2; \beta)$.

The considered missing situation can be described as follows. Let M be an indicator variable indicating if X_2 is observable ($M = 1$) or not ($M = 0$). The missing mechanism is assumed to satisfy the MAR assumption, that is $\Pr(M = 1|y, x_1, x_2) = \Pr(M = 1|y, x_1) = q(y, x_1)$. These conditional probabilities for complete observations are assumed to be bounded away from zero. The MAR assumption allows us

to assume that unobserved values of X_2 have the same conditional distribution as the observed values. Let now (Y^i, X_1^i, X_2^i, M^i), $i = 1, \ldots, N$, be an independent sample from (Y, X_1, X_2, M). With $\mathcal{V} = \{i | m^i = 1\}$ the observed data is given by $\{(y^i, x_1^i, x_2^i, m^i) | i \in \mathcal{V}\}$ and $\{(y^i, x_1^i, m^i) | i \in \{1, \ldots, N\} \backslash \mathcal{V}\}$. The likelihood generating (Y, X_1, X_2, M) is

$$\begin{aligned} L(\beta, \theta) &= f_{X_1}(x_1|\alpha) f_{M|Y,X_1}(m|y, x_1; \gamma) \left\{ \Pr(y|x_1, x_2; \beta) f_{X_2|X_1}(x_2|x_1; \xi) \right\}^m \\ &\quad \left\{ \int \Pr(y|x_1, z; \beta) f_{X_2|X_1}(z|x_1; \xi)\, dz \right\}^{1-m}, \end{aligned} \tag{2}$$

where $\theta = (\alpha, \gamma, \xi)$. The parameters α, γ and ξ refer to the marginal distribution of X_1, to the conditional distribution of M given Y and X_1 which is binomial with probabilities $q(y, x_1)$, and to the conditional distribution of X_2 given X_1. Maximizing (2) in β is obviously not possible without knowledge of $f_{X_2|X_1}$ whereas knowledge about f_{X_1} and the missing mechanism is not required as far as the latter is MAR. The parametric approach which will be proposed in the next section consists in specifying $f_{X_2|X_1}$ up to the unknown parameter ξ which is assumed to be finite and then maximizing (2) simultaneously in β and ξ. The semiparametric approach takes θ as an 'infinite dimensional' parameter with values in the set of the corresponding densities.

A special role will be played by the observable response rates given by

$$\frac{V(y, x_1)}{N(y, x_1)} = \hat{q}(y, x_1), \qquad y, x_1 \in \{0, 1\}, \tag{3}$$

with frequencies $N(y, x_1) = \#\{i \in \{1, \ldots, N\} | y^i = y, x_1^i = x_1\}$ and $V(y, x_1) = \#\{i \in \mathcal{V} | y^i = y, x_1^i = x_1\}$. They can be regarded as estimates of $q(y, x_1)$, $y, x_1 \in \{0, 1\}$. Note that this straightforward estimation is only possible if Y and X_1 are discrete.

3. The Estimators

3.1 Complete case analysis

The complete case analysis consists in applying complete data methods to the reduced data set $\{(y^i, x_1^i, x_2^i) | i \in \mathcal{V}\}$, i.e. it maximizes

$$L^{CC}(\beta) = \prod_{i \in \mathcal{V}} \Pr(y^i | x_1^i, x_2^i; \beta).$$

The resulting estimator will be denoted by $\hat{\beta}^{CC}$. As shown by Zhao et al. (1996) it is consistent if the missingness is conditionally independent of Y given X_1 and X_2 but it may be biased under MAR. Obviously the complete case estimator is in general not efficient since it ignores the information in $\{(y^i, x_1^i) | i \in \{1, \ldots, N\} \backslash \mathcal{V}\}$.

3.2 Maximum likelihood estimation

Following Ibrahim and Weisberg (1992) the considered maximum likelihood estimator is computed under the assumption that the conditional distribution of X_2 given X_1 is Gaussian. This is parametrized as follows: let $\mu_x = \mathrm{E}(X_2 | X_1 = x)$, $x \in \{0, 1\}$, denote the means which depend on X_1 and σ^2 the variance which is independent of X_1, i.e. we have in (2) $\xi = (\mu_0, \mu_1, \sigma^2)$. The likelihood to be maximized is given by

$$\begin{aligned} L^{ML}(\beta, \xi) &= \prod_{i \in \mathcal{V}} \left[\Pr(y^i | x_1^i, x_2^i; \beta) f_{X_2|X_1}(x_2^i | x_1^i; \xi) \right] \\ &\quad \prod_{j \in \overline{\mathcal{V}}} \left[\int \Pr(y^j | x_1^j, z; \beta) f_{X_2|X_1}(z | x_1^j; \xi)\, dz \right], \end{aligned}$$

where $\overline{\mathcal{V}} = \{1, \ldots, N\} \backslash \mathcal{V}$ and $f_{X_2|X_1}(\cdot | x_1)$ is the density of the Gaussian distribution with parameters μ_{x_1} and σ^2. In general, maximization of $L^{ML}(\beta, \xi)$ has to be carried out numerically due to the integration in the third sum. This can partly be simplified by using the EM algorithm (Dempster et al., 1977) which is easy to apply when the considered model is an exponential family. In our special case, the joint conditional distribution of Y and X_2 given X_1 constitutes an exponential family as one can easily check. Still, the E–step involves numerical integration in order to compute the expectations with respect to the distribution of X_2 given Y and X_1 with density

$$f_{X_2|Y,X_1}(x_2 | y, x_1; \xi, \beta) = \frac{\Pr(y | x_1, x_2; \beta) f_{X_2|X_1}(x_2 | x_1; \xi)}{\int \Pr(y | x_1, z; \beta) f_{X_2|X_1}(z | x_1; \xi)\, dz}. \tag{4}$$

In our simulation the denominator is approximated by a 10 point Gaussian quadrature in analogy to Ibrahim and Weisberg (1992).

3.3 Semiparametric estimation

In this section we first present some specific semiparametric estimators which leave the unknown distributions in (2) completely unrestricted and which are consistent under the MAR assumption. Their relation to the parametric maximum likelihood

estimator is discussed. After that, a general class of semiparametric estimators is introduced which contains the semiparametric efficient estimator.

3.3.1 Corrected complete case estimator

The complete case estimator may be biased under the MAR assumption. By considering the bias factor Vach and Illi (1997) show that in the special case of a logistic regression model a simple correction is given by

$$\begin{aligned} \hat{\beta}_0^{CCC} &= \hat{\beta}_0^{CC} + \log \frac{\hat{q}(0,0)}{\hat{q}(1,0)}, \qquad \hat{\beta}_2^{CCC} = \hat{\beta}_2^{CC} \\ \hat{\beta}_1^{CCC} &= \hat{\beta}_1^{CC} + \log \frac{\hat{q}(1,0)\hat{q}(0,1)}{\hat{q}(0,0)\hat{q}(1,1)}. \end{aligned} \tag{5}$$

Note that this estimator uses the incomplete observations since the correction terms use (3) and therefore the additional knowledge about the frequencies $N(y, x_1)$.

Cain and Breslow (1988) derive $\hat{\beta}^{CCC}$ as a special case of a pseudoconditional likelihood approach in a more general setting where they prove the asymptotic normality (Breslow and Cain, 1988).

In case that all covariates are discrete and a saturated model for the covariate distribution is assumed $\hat{\beta}^{CCC}$ is identical to the ML estimator (Vach and Illi, 1997). It follows that in our case of a continuous X_2 $\hat{\beta}^{CCC}$ can be derived as a 'nonparametric' maximum likelihood estimator in the following sense. Let $\mathcal{X}$ denote the observed values of X_2 and $\xi = (\xi(x_1, x_2)|\ x_1 \in \{0,1\},\ x_2 \in \mathcal{X})$ discrete conditional probabilities for $X_2 = x_2$ given $X_1 = x_1$. By assuming that $f_{X_2|X_1}$ is the density of an arbitrary conditional distribution of X_2 given X_1 putting mass only on the observed values of X_2 we get $\hat{\beta}^{CCC}$ by maximizing the likelihood (2) in β and ξ. In the univariate case this nonparametric procedure leads to the empirical distribution as estimator of the underlying continuous one.

3.3.2 Mean score estimator

As shown by Robins et al. (1995) the contribution of an incomplete observation to the total score function is given by the derivation of the logarithm of (2) with respect to β which can be writen as

$$E\left(\frac{\partial}{\partial\beta}\log \Pr(Y|X_1, X_2; \beta)\middle| Y = y, X_1 = x_1\right)$$

$$= \int \frac{\partial}{\partial \beta} \log \left[\Pr(y|x_1, x_2; \beta) \right] f_{X_2|Y,X_1}(x_2|y, x_1) dx_2, \tag{6}$$

evaluated at the unknown true conditional density $f_{X_2|Y,X_1}$. Under the MAR assumption a consistent estimate of $f_{X_2|Y,X_1}$ can be based on the complete cases. Reilly and Pepe (1995) choose the empirical conditional distribution leading to

$$\sum_{i \in \mathcal{V}(y,x_1)} \frac{1}{V(y,x_1)} \frac{\partial}{\partial \beta} \log \Pr(y|x_1, x_2^i; \beta) \tag{7}$$

as an estimator for (6) where $\mathcal{V}(y, x_1) = \{i \in \mathcal{V} | y^i = y \wedge x_1^i = x_1\}$. As shown by the authors, replacing the unknown contribution of an incomplete observation to the total score function by (7) leads to a weighted sum of the contributions of the complete cases which motivates the name of the mean score method. The estimated total score function is given by

$$\sum_{i \in \mathcal{V}} \left(\frac{N(y^i, x_1^i)}{V(y^i, x_1^i)} \right) \frac{\partial}{\partial \beta} \log \Pr(y^i | x_1^i, x_2^i; \beta).$$

Computation of the corresponding estimator $\hat{\beta}^{ms}$ as root of the above expression is straightforward. Reilly and Pepe (1995) show that it is consistent and asymptotically normal.

A similar idea is proposed by Pepe and Fleming (1991) and Carroll and Wand (1991). Note that expression (6) can be rewritten as

$$\int \frac{\partial}{\partial \beta} \log \left[\Pr(y|x_1, x_2; \beta) \right] \frac{\Pr(y|x_1, x_2; \beta) f_{X_2|X_1}(x_2|x_1)}{\int \Pr(y|x_1, z; \beta) f_{X_2|X_1}(z|x_1) dz} \, dx_2. \tag{8}$$

The authors propose to substitute $f_{X_2|X_1}$ in (8) by a nonparametric density estimator. Since this estimator has to be based on the complete cases it is only consistent if the missing mechanism is MAR and does not depend on the response variable. It follows that the resulting estimator of β, too, is only consistent under this restrictive condition. A detailed discussion can be found in Robins et al. (1995).

Note that (6) or (8) are identical to the expectation of the loglikelihood for a complete observation with respect to (4), i.e. the conditional distribution of X_2 given Y and X_1. The idea of Reilly and Pepe (1995) and Pepe and Fleming (1991) can therefore be viewed as approximation of the maximum likelihood estimation by estimating the E–step and performing only one iteration of the EM algorithm.

3.3.3 Semiparametric efficient estimation

Robins et al. (1994) propose a class of semiparametric estimators which depend on two functions: With K denoting the dimension of the regression parameter the first one, $h : \mathbb{R}^2 \to \mathbb{R}^K$, is a function of the covariates and the second one, $\varphi : \mathbb{R}^2 \to \mathbb{R}^K$, is a function of the completely observed variables. The corresponding estimator $\hat{\beta}(h, \varphi)$ is given as solution of the following equation system

$$\sum_{i=1}^{N} \left(\frac{m^i h(x_1^i, x_2^i) \varepsilon^i(\beta)}{q(y^i, x_1^i)} - \frac{(m^i - q(y^i, x_1^i))\varphi(y^i, x_1^i)}{q(y^i, x_1^i)} \right) = 0, \tag{9}$$

where $\varepsilon^i(\beta) = y^i - \mathrm{E}(Y^i | x_1^i, x_2^i; \beta)$. Under regularity conditions and under MAR $\hat{\beta}(h, \varphi)$ is consistent and asymptotically normal. If the unknown missing mechanism in (9) is replaced by (3) the resulting estimator of β will be denoted by $\widehat{\hat{\beta}}(h, \varphi)$.

The main interest of Robins et al. (1994) concerns the derivation of an estimator which is semiparametric efficient. They show that the proposed class contains an estimator $\hat{\beta}(h_{eff}, \varphi_{eff})$ that attains the lower variance bound with the functions h_{eff} and φ_{eff} given as follows. The first is the solution of the functional equation

$$h(x_1, x_2) = t(x_1, x_2) \Big[\frac{\partial}{\partial \beta} \mu(x_1, x_2; \beta^0) + E_{Y|X_1,X_2} \big\{ (q(Y, X_1)^{-1} - 1) \\ \cdot E_{X_2|Y,X_1} \left(h(X_1, X_2) \varepsilon(\beta^0) | Y, X_1 = x_1 \right) \varepsilon(\beta^0) | X_1 = x_1, X_2 = x_2 \big\} \Big] \tag{10}$$

with $t(x_1, x_2) = \{ \mathrm{E}(\frac{\varepsilon(\beta^0)^2}{q(Y, X_1)} | X_1 = x_1, X_2 = x_2) \}^{-1}$ and β^0 as true value of the regression parameter. The function φ^h that minimizes the asymptotic variance of $\hat{\beta}(h, \varphi^h)$ for a general function h is given as conditional expectation

$$\varphi^h(y, x_1) = \mathrm{E}(h(X_1, X_2) \varepsilon(\beta) | Y = y, X_1 = x_1).$$

It follows that $\varphi_{eff} = \varphi^{h_{eff}}$. The authors further show that $\hat{\beta}(h, \varphi^h)$ is asymptotically equivalent to $\widehat{\hat{\beta}}(h, 0)$ calling the latter a pseudo complete case estimator since it uses the incomplete observations only to estimate the response rates (3).

In order to get closed expressions for φ_{eff} and h_{eff} we can make use of the fact that Y is discrete. Let $\mathcal{Y}$ denote the finite set of possible realizations of Y. Then taking expectation of (10) with respect to the conditional distribution of X_2 given Y and X_1 leads to a closed expression for each $\varphi_{eff}(y_0, \cdot)$, $y_0 \in \mathcal{Y}$. These are imputed in (10) to get h_{eff}. Both functions obviously still depend on the unspecified distribution of

X_2 given Y and X_1, on the true value β^0 of the regression parameter, and on the missing mechanism q. Estimators $\hat{h}_{eff}$ and $\hat{\varphi}_{eff}$ with the property that $\widehat{\hat{\beta}}(\hat{h}_{eff}, \hat{\varphi}_{eff})$ is asymptotically equivalent to $\hat{\beta}(h_{eff}, \varphi_{eff})$ can for example be obtained in the following way. The conditional distribution of X_2 given Y and X_1 is estimated by the corresponding empirical one and the unknown β^0 can be replaced by any consistent estimator even an inefficient one. In our simulation study we choose the mean score estimator since it is easy to compute. Finally, q is replaced by (3). Robins et al. (1994) show the desired asymptotic equivalence of the resulting estimator $\hat{\beta}^{eff}$ to the semiparametric efficient estimator.

Note that the semiparametric estimators proposed above are elements of the class just defined which can be seen as follows. If we choose $h = h^F_{eff}$ as the optimal function for complete data and if $q(y, x_1) \equiv q$ where q is a constant then $\hat{\beta}^{CC} = \hat{\beta}(h^F_{eff}, 0)$. In contrast, $\widehat{\hat{\beta}}(h^F_{eff}, 0)$ is consistent for general MAR mechanisms and identical to the mean score estimator. Furthermore, one can find a function h^{BC} such that $\widehat{\hat{\beta}}(h^{BC}, 0)$ is asymptotically equivalent to the estimator proposed by Breslow and Cain (1988) which is in our case the corrected complete case estimator. But neither $\widehat{\hat{\beta}}(h^F_{eff}, 0)$ nor $\widehat{\hat{\beta}}(h^{BC}, 0)$ are in general semiparametric efficient.

4. Simulation Designs

The simulation study presented here compares the proposed estimators for small sample size. A similar study has been carried out by Robins et al. (1994) with a large sample size and without including the maximum likelihood estimator. Other studies (Zhao and Lipsitz, 1992; Vach, 1994) consider only discrete covariates where the problem of misspecification, which is of special interest here, does not occur.

The different simulation designs are given by varying the type of missing mechanism, the slope of the conditional distribution of X_2 given X_1, the dependence between these covariates, and the regression parameter. The chosen missing mechanisms can be read off Table 1. The first mechanism means missing completely at random (MCAR) since the missingness is independent of Y and X_1. The second depends only on X_1 (MDX) and the third only on Y (MDY). Consequently, MDXY means that the mechanism depends on both, Y and X_1. Note that the MCAR mechanism

leads to a greater over all missing rate than the other mechanisms. This has to be taken into account when interpreting the results.

Table 1:
The missing mechanisms and the corresponding probabilities $q(y, x_1)$.

	$q(0,0)$	$q(1,0)$	$q(0,1)$	$q(1,1)$
MCAR	0.3	0.3	0.3	0.3
MDX	0.8	0.8	0.3	0.3
MDY	0.8	0.3	0.8	0.3
MDXY	0.8	0.3	0.3	0.8

The conditional distribution of X_2 given X_1 is either Gaussian, or $t(6)$ representing a symmetric but heavy–tailed distribution, or $\chi^2(1)$ representing a non–symmetric distribution. These choices are rather meant as archetypes than as being realistic. With respect to the dependence between the covariates we consider two choices for $\mu_x = \mathrm{E}(X_2|X_1 = x)$. In the case $\mu_0 = \mu_1 = 0$ the covariates are independent, in the case $\mu_0 = -1, \mu_1 = 1$ they are dependent.

To keep the number of parameter constellations limited we let β_2 take the values $\{-1.5, 0, 1.5\}$ whereas β_0 and β_1 are kept fixed as $\beta_0 = 0$ and $\beta_1 = 1$.

The covariate X_1 follows a Bernoulli distribution with $\Pr(X_1 = 1) = 0.5$. The sample size is chosen to be $N = 200$. For each of the resulting 72 designs 1000 samples are generated using Turbo Pascal 7.0.

5. Results

In order to compare the estimators we compute the estimated relative mean squared errors which are the ratios of the Monte Carlo mean squared error of the semiparametric efficient estimator and that of the considered one. This will simply be called relative MSE. Robins et al. (1994) consider instead the estimated relative efficiencies, i.e. the ratio of the Monte Carlo variances. This is not sensible here since the sample size is considerably smaller and hence bias is not negligible. Also, for this reason we additionally compute the means of the observed biases which we will simply call bias.

5.1 Comparison of the semiparametric efficient estimator and the ML estimator

We first discuss the bias of the semiparametric efficient and the ML estimator and after that the relative MSE of the latter one. Therein we distinguish the cases where the ML estimator is based on a correct assumption about the covariate distribution and where this assumption is wrong.

5.1.1 Bias of the semiparametric efficient estimator

The bias of the semiparametric efficient estimator can be read off Table 2. In case that X_2 has no influence, i.e. $\beta_2 = 0$, the bias of all three components $\hat{\beta}_0^{eff}$, $\hat{\beta}_1^{eff}$ and $\hat{\beta}_2^{eff}$ is negligible for each missing mechanisms and each covariate distribution since it is always in absolute value smaller than 0.08.

The case $\beta_2 \neq 0$ is more serious especially concerning the estimation of β_2 when the covariate distribution is not Gaussian. Here, the bias of $\hat{\beta}_2$ is for $\mu_0 = \mu_1$ often, and for $\mu_0 \neq \mu_1$ and any covariate distribution nearly always in absolute value larger than 0.1, and even larger for the χ^2 and Student covariate distribution. In both latter situations the bias of $\hat{\beta}_2$ is in absolute value roughly about 0.24 for the MCAR and about 0.12 for the other missing mechanisms. The bias of $\hat{\beta}_0$ and $\hat{\beta}_1$ is for $\beta_2 \neq 0$ still relatively small. The bias of $\hat{\beta}_0$ is in absolute value smaller than 0.1 besides some exceptions for the χ^2 covariate distribution whereas the one of $\hat{\beta}_1$ can be larger than 0.1 for $\mu_0 \neq \mu_1$ and also for the χ^2 and Student covariate distribution. The largest bias has an absolute value of 0.26 and can be observed for three designs, two of them within the χ^2 covariate distribution.

An additional aspect concerns the direction of the bias. The estimation of β_0 has nearly always a negative one especially for $\mu_0 = \mu_1$ and the Gaussian covariate distribution. In contrast to this, the bias of $\hat{\beta}_1^{eff}$ is in general positive for $\mu_0 = \mu_1$. The direction of the bias of $\hat{\beta}_2^{eff}$ depends on the true value: it is negative for $\beta_2 = -1.5$ and positive for $\beta_2 = 1.5$.

Table 2:
Bias of the semiparametric efficient estimator proposed by Robins et al. (1994).

covariate distrib. = Gaussian		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{eff}$	$\hat{\beta}_1^{eff}$	$\hat{\beta}_2^{eff}$	$\hat{\beta}_0^{eff}$	$\hat{\beta}_1^{eff}$	$\hat{\beta}_2^{eff}$
MCAR, $\beta_2 =$	-1.5	*-0.02*	*0.08*	*-0.13*	*-0.07*	*0.21*	*-0.18*
	0	*-0.01*	*0.03*	*-0.01*	*-0.02*	*0.07*	*-0.02*
	1.5	*-0.00*	*0.08*	*0.19*	*0.07*	*-0.07*	*0.24*
MDX, $\beta_2 =$	-1.5	*-0.00*	*0.05*	*-0.08*	*-0.02*	*0.11*	*-0.10*
	0	*-0.01*	*0.02*	*-0.01*	*-0.01*	*0.03*	*-0.00*
	1.5	*-0.02*	*0.07*	*0.09*	*0.01*	*-0.00*	*0.10*
MDY, $\beta_2 =$	-1.5	*-0.02*	*0.05*	*-0.09*	*-0.05*	*0.11*	*-0.09*
	0	*-0.01*	*0.02*	*0.00*	*-0.02*	*0.06*	*-0.01*
	1.5	*-0.03*	*0.04*	*0.08*	*0.01*	*-0.02*	*0.13*
MDXY, $\beta_2 =$	-1.5	*-0.03*	*0.08*	*-0.09*	*-0.05*	*0.13*	*-0.08*
	0	*-0.02*	*0.04*	*-0.01*	*-0.01*	*0.03*	*0.01*
	1.5	*-0.01*	*0.07*	*0.08*	*0.02*	*0.03*	*0.14*

covariate distrib. $= \chi^2$		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{eff}$	$\hat{\beta}_1^{eff}$	$\hat{\beta}_2^{eff}$	$\hat{\beta}_0^{eff}$	$\hat{\beta}_1^{eff}$	$\hat{\beta}_2^{eff}$
MCAR, $\beta_2 =$	-1.5	*-0.08*	*0.10*	*-0.26*	*-0.10*	*0.23*	*-0.23*
	0	*-0.00*	*0.05*	*0.01*	*0.03*	*-0.02*	*0.02*
	1.5	*0.04*	*0.07*	*0.19*	*0.13*	*-0.16*	*0.26*
MDX, $\beta_2 =$	-1.5	*-0.02*	*0.07*	*-0.09*	*-0.08*	*0.15*	*-0.12*
	0	*0.00*	*0.03*	*0.02*	*0.01*	*-0.01*	*0.01*
	1.5	*0.03*	*-0.00*	*0.08*	*0.04*	*-0.05*	*0.11*
MDY, $\beta_2 =$	-1.5	*-0.08*	*0.06*	*-0.12*	*-0.13*	*0.19*	*-0.15*
	0	*-0.01*	*0.02*	*-0.01*	*-0.04*	*0.07*	*-0.02*
	1.5	*0.00*	*0.06*	*0.11*	*0.02*	*-0.03*	*0.13*
MDXY, $\beta_2 =$	-1.5	*-0.06*	*0.13*	*-0.12*	*-0.07*	*0.15*	*-0.10*
	0	*-0.02*	*0.06*	*0.01*	*-0.02*	*0.07*	*-0.01*
	1.5	*0.01*	*0.07*	*0.11*	*0.00*	*0.01*	*0.15*

covariate distrib. = student		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{eff}$	$\hat{\beta}_1^{eff}$	$\hat{\beta}_2^{eff}$	$\hat{\beta}_0^{eff}$	$\hat{\beta}_1^{eff}$	$\hat{\beta}_2^{eff}$
MCAR, $\beta_2 =$	-1.5	*-0.02*	*0.11*	*-0.19*	*-0.08*	*0.26*	*-0.22*
	0	*0.00*	*0.04*	*-0.00*	*-0.03*	*0.07*	*-0.02*
	1.5	*-0.01*	*0.09*	*0.15*	*0.10*	*-0.08*	*0.26*
MDX, $\beta_2 =$	-1.5	*0.01*	*0.01*	*-0.11*	*-0.03*	*0.13*	*-0.12*
	0	*-0.01*	*0.04*	*-0.00*	*0.01*	*0.00*	*0.01*
	1.5	*-0.01*	*0.05*	*0.10*	*0.04*	*-0.02*	*0.13*
MDY, $\beta_2 =$	-1.5	*-0.01*	*0.03*	*-0.09*	*-0.05*	*0.12*	*-0.11*
	0	*-0.01*	*0.02*	*0.00*	*-0.00*	*0.01*	*0.01*
	1.5	*-0.04*	*0.04*	*0.10*	*0.01*	*-0.02*	*0.12*
MDXY, $\beta_2 =$	-1.5	*-0.02*	*0.09*	*-0.10*	*-0.09*	*0.20*	*-0.11*
	0	*0.00*	*0.02*	*-0.01*	*-0.01*	*0.03*	*0.01*
	1.5	*-0.03*	*0.11*	*0.13*	*0.00*	*0.04*	*0.15*

5.1.2 Bias of the the ML estimator

As can bee seen from Table 3, the bias of the ML estimator is very similar to the

one of the semiparametric efficient one for the Gaussian covariate distribution.

Table 4 shows the bias of the ML estimator in the situations where the distributional assumptions are wrong, i.e. for the χ^2 and Student covariate distribution. Here, we can observe a surprisingly small bias for several parameter constellations. If $\beta_2 = 0$ the bias is in absolute value smaller than 0.08 for each missing mechanism and regardless of the dependence between the covariates. Thus, the wrong assumption about the covariate distribution does not appear to affect the consistency of the ML estimator when this covariate has no influence.

If $\beta_2 \neq 0$ and X_2 follows the considered χ^2 distribution the bias is clearly affected. Especially the estimation of β_2 in the presence of a missing mechanism that depends on the response variable (MDY and MDXY) appears to be inconsistent since the bias takes values about 0.5 and even larger ones. The bias when estimating β_0 and β_1 is in general larger than 0.1 if $\mu_0 = \mu_1$ and even larger if $\mu_0 \neq \mu_1$. In contrast, the bias of $\hat{\beta}_2^{ML}$ is negligible given one of the other missing mechanisms (MCAR or MDX) whereas it is in some cases about 0.2 for $\hat{\beta}_0^{ML}$ and $\hat{\beta}_1^{ML}$. The largest observed absolute bias among the designs with χ^2 distribution is 1.03.

If the covariate distribution is Student we can observe the same bias pattern as for the Gaussian covariate distribution with a slight general tendency to extremer values and a clear tendency to extremer values for the estimation of β_1 and β_2 in the special case of $\mu_0 = \mu_1$, $\beta_2 \neq 0$ and a MDXY missing mechanism. With this last exception, the results are also similar to those of the semiparametric efficient estimator for the Student distribution. The largest observed absolute bias among the designs with Student covariate distribution is 0.27.

5.1.3 Relative MSE of the ML estimator with correct assumptions

The results presented in Table 3 allow a direct comparison of the semiparametric efficient and the parametric efficient estimation. Since the bias is similar for both estimators the relative MSE essentially reflects the gain in efficiency due to the additional parametric assumption.

Table 3:
Relative MSE and bias of the maximum likelihood estimator with incomplete data assuming a gaussian covariate distribution. (The bias is given in brackets in italics.)

covariate distrib. = Gaussian		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\beta}_2^{ML}$	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\beta}_2^{ML}$
MCAR, $\beta_2 =$	-1.5	1.05 *(-0.02)*	1.05 *(0.07)*	1.04 *(-0.13)*	1.05 *(-0.06)*	1.04 *(0.19)*	1.03 *(-0.17)*
	0	1.00 *(-0.01)*	1.00 *(0.03)*	1.00 *(-0.01)*	1.01 *(-0.02)*	1.01 *(0.07)*	1.00 *(-0.02)*
	1.5	1.04 *(-0.00)*	1.05 *(0.08)*	1.02 *(0.19)*	1.06 *(0.06)*	1.05 *(-0.07)*	1.07 *(0.22)*
MDX, $\beta_2 =$	-1.5	1.01 *(-0.00)*	1.04 *(0.05)*	1.01 *(-0.08)*	1.00 *(-0.02)*	1.01 *(0.10)*	1.01 *(-0.10)*
	0	1.00 *(-0.01)*	1.00 *(0.02)*	1.00 *(-0.01)*	1.00 *(-0.01)*	1.00 *(0.03)*	1.00 *(-0.00)*
	1.5	1.00 *(-0.02)*	1.03 *(0.07)*	1.01 *(0.09)*	1.01 *(0.01)*	1.03 *(0.01)*	1.04 *(0.10)*
MDY, $\beta_2 =$	-1.5	1.07 *(-0.02)*	1.07 *(0.04)*	1.10 *(-0.09)*	1.07 *(-0.04)*	1.04 *(0.10)*	1.05 *(-0.09)*
	0	1.00 *(-0.01)*	0.99 *(0.03)*	0.99 *(0.00)*	1.01 *(-0.02)*	1.01 *(0.06)*	1.00 *(-0.01)*
	1.5	1.05 *(-0.02)*	1.06 *(0.04)*	1.04 *(0.08)*	1.03 *(0.02)*	1.04 *(-0.00)*	1.10 *(0.12)*
MDXY, $\beta_2 =$	-1.5	1.06 *(-0.02)*	1.07 *(0.06)*	1.07 *(-0.08)*	1.06 *(-0.03)*	1.07 *(0.10)*	1.03 *(-0.07)*
	0	1.00 *(-0.02)*	1.00 *(0.04)*	0.99 *(-0.01)*	1.02 *(0.00)*	1.02 *(0.02)*	1.01 *(0.01)*
	1.5	1.04 *(-0.01)*	1.04 *(0.06)*	1.04 *(0.08)*	1.06 *(0.03)*	1.07 *(0.01)*	1.16 *(0.12)*

At first, one can say that both estimators are nearly equal for $\beta_2 = 0$, i.e. when the incompletely observed covariate has no effect on the response, since the relative MSE lies between 0.99 and 1.01. Furthermore, the gain in efficiency is generally only modest for the MCAR and MDX missing mechanisms amounting to less than 1.08 in the first case and to less than 1.05 in the second and being even smaller when $\mu_0 = \mu_1$. If $\beta_2 \neq 0$ and the missing mechanism is MDY or MDXY we can observe that in more than half of the designs the relative MSE is greater than 1.05 reaching the maxima of 1.10 and 1.16, respectively, for $\beta_2 = 1.5$. The missing mechanisms which depend on the response variable may therefore be those where the maximum likelihood estimator truly outperforms the semiparametric efficient one.

5.1.4 Relative MSE of the ML estimator with wrong assumptions

Despite the wrong distributional assumption there are some situations where the maximum likelihood estimator performs at least as good as the semiparametric efficient one with respect to the relative MSE. If $\beta_2 = 0$ the relative MSEs of all three components, $\hat{\beta}_0^{ML}$, $\hat{\beta}_1^{ML}$, and $\hat{\beta}_2^{ML}$, are at least 1 for both covariate distributions and almost every missing mechanism while at the same time the bias is always very small as we have seen above. Thus, in these situations the maximum likelihood estimator seems neither inconsistent nor inefficient. Exceptions are given by the design with the MDXY mechanism and the χ^2 distribution where for example the relative MSE

of $\hat{\beta}_2$ is only 0.72 ($\mu_0 = \mu_1$) and 0.63 ($\mu_0 \neq \mu_1$) which constitutes a serious loss of efficiency although the bias is still very small.

Table 4:
Relative MSE and bias of the maximum likelihood estimator with incomplete data falsely assuming a Gaussian covariate distribution. (The bias is given in brackets in italics.)

covariate distrib. $= \chi^2$		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\beta}_2^{ML}$	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\beta}_2^{ML}$
MCAR, $\beta_2 =$	-1.5	1.16 *(0.16)*	1.15 *(0.08)*	1.64 *(-0.08)*	1.14 *(0.27)*	1.20 *(-0.13)*	1.31 *(-0.10)*
	0	1.02 *(-0.01)*	0.99 *(0.05)*	1.02 *(0.01)*	1.13 *(0.03)*	1.14 *(-0.02)*	1.05 *(0.02)*
	1.5	0.89 *(-0.19)*	0.81 *(0.16)*	1.30 *(0.07)*	1.28 *(-0.25)*	1.01 *(0.44)*	1.34 *(0.07)*
MDX, $\beta_2 =$	-1.5	1.03 *(0.06)*	1.01 *(0.22)*	1.19 *(-0.01)*	1.15 *(0.04)*	1.01 *(0.08)*	1.15 *(-0.07)*
	0	1.00 *(0.00)*	1.00 *(0.03)*	1.00 *(0.02)*	1.02 *(0.01)*	1.06 *(-0.01)*	1.00 *(0.01)*
	1.5	1.04 *(-0.03)*	0.83 *(-0.03)*	1.12 *(0.05)*	1.12 *(-0.09)*	0.77 *(0.33)*	1.16 *(0.04)*
MDY, $\beta_2 =$	-1.5	1.26 *(0.15)*	1.10 *(0.03)*	0.68 *(0.51)*	0.66 *(0.69)*	0.54 *(-1.03)*	0.93 *(0.42)*
	0	1.01 *(-0.01)*	1.00 *(0.02)*	1.01 *(0.01)*	1.16 *(-0.01)*	1.13 *(0.03)*	1.14 *(0.00)*
	1.5	0.79 *(-0.03)*	0.90 *(0.10)*	0.25 *(0.71)*	0.83 *(0.26)*	0.52 *(-0.50)*	0.84 *(0.27)*
MDXY, $\beta_2 =$	-1.5	0.73 *(-0.03)*	1.08 *(0.16)*	1.27 *(-0.06)*	0.71 *(0.43)*	1.07 *(-0.25)*	1.08 *(0.02)*
	0	0.98 *(-0.02)*	0.93 *(0.07)*	0.72 *(0.03)*	0.89 *(0.04)*	1.03 *(-0.05)*	0.63 *(0.06)*
	1.5	0.81 *(-0.17)*	0.26 *(0.73)*	0.44 *(0.48)*	0.67 *(0.37)*	0.96 *(-0.10)*	0.34 *(0.77)*

covariate distrib. = Student		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\beta}_2^{ML}$	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\beta}_2^{ML}$
MCAR, $\beta_2 =$	-1.5	1.00 *(-0.02)*	1.00 *(0.13)*	1.03 *(-0.18)*	1.11 *(-0.02)*	1.14 *(0.14)*	1.09 *(-0.20)*
	0	1.00 *(0.00)*	1.00 *(0.04)*	1.00 *(-0.00)*	1.04 *(-0.03)*	1.06 *(0.07)*	1.02 *(-0.02)*
	1.5	1.00 *(-0.01)*	1.01 *(0.09)*	1.03 *(0.15)*	1.19 *(0.04)*	1.08 *(0.03)*	1.12 *(0.22)*
MDX, $\beta_2 =$	-1.5	0.99 *(0.01)*	0.99 *(0.04)*	1.01 *(-0.10)*	1.01 *(-0.02)*	1.05 *(0.07)*	1.02 *(-0.11)*
	0	1.00 *(-0.01)*	1.00 *(0.04)*	1.00 *(-0.00)*	1.01 *(0.01)*	1.03 *(0.00)*	1.01 *(0.01)*
	1.5	0.99 *(-0.01)*	0.98 *(0.07)*	1.01 *(0.10)*	1.02 *(0.03)*	1.08 *(0.07)*	1.02 *(0.12)*
MDY, $\beta_2 =$	-1.5	0.98 *(-0.05)*	1.02 *(0.05)*	1.01 *(-0.09)*	1.08 *(-0.00)*	1.14 *(0.02)*	1.16 *(-0.08)*
	0	1.00 *(-0.01)*	0.99 *(0.02)*	1.01 *(0.00)*	1.03 *(-0.00)*	1.04 *(0.01)*	1.03 *(0.01)*
	1.5	0.96 *(-0.07)*	1.02 *(0.06)*	1.07 *(0.09)*	1.11 *(-0.02)*	1.02 *(0.09)*	1.16 *(0.10)*
MDXY, $\beta_2 =$	-1.5	0.91 *(-0.07)*	0.83 *(0.21)*	0.88 *(-0.16)*	1.13 *(-0.01)*	1.15 *(0.05)*	1.18 *(-0.03)*
	0	1.00 *(0.00)*	1.00 *(0.02)*	0.97 *(-0.01)*	1.05 *(0.01)*	1.07 *(-0.02)*	0.99 *(0.03)*
	1.5	0.88 *(-0.08)*	0.79 *(0.23)*	0.87 *(0.19)*	1.02 *(0.02)*	1.03 *(0.02)*	0.82 *(0.27)*

If $\beta_2 \neq 0$ and X_2 is distributed according to the considered χ^2 distribution the maximum likelihood estimator performs fairly well for the MCAR and MDX missing mechanisms. The relative MSE is often greater than 1 and nearly always greater than 0.8. Taking the bias into account, it follows that the good results are mainly due to a small variance of the maximum likelihood estimator. But if additionally the missing mechanism depends on the response variable the results are truly bad. The smallest observed relative MSE amounts to 0.25 and the largest absolute bias to 1.03 occurring for the MDY mechanism.

If the covariate distribution is Student the maximum likelihood estimator performs nearly as well as for the Gaussian. The designs where the relative MSE takes values smaller than 1 are given when the covariates are independent and the missing mechanism is not MCAR. It reaches its minimum of 0.79 for the MDXY mechanism and $\beta_2 = 1.5$. If, in contrast, $\mu_0 \neq \mu_1$ the relative MSE is in most cases even greater for the Student than for the Gaussian covariate distribution. This may suggest that the maximum likelihood estimator is still appropriate for dependent covariates because it makes a correct assumption about the dependence structure although the distributional assumption is wrong.

5.2 Loss of information due to missing values

The comparison of the semiparametric efficient estimator with the complete data maximum likelihood estimator denoted by $\hat{\beta}^{full}$ allows to assess the general loss of information of the semiparametric approach which is due to the missing values (measured by the relative MSE of the complete data maximum likelihood estimator). The simulation results are shown in Table 5.

First we discuss the information loss concerning the estimation of β_0 and β_1. An important result is that this is very small if $\beta_2 = 0$ and the covariates are independent ($\mu_0 = \mu_1$) as can be observed for each missing mechanism and each covariate distribution. But in the other situations, i.e. if $\beta_2 \neq 0$ or $\mu_0 \neq \mu_1$, one has to reckon with a considerable loss of information due to the missing values since the relative MSE is clearly greater than 1. If the covariates are not independent ($\mu_0 \neq \mu_1$) the relative MSE is not smaller even for the designs whith $\beta_2 = 0$. Especially for the MDY and MDXY missing mechanisms it decreases with increasing β_2. The simultaneity of $\beta_2 \neq 0$ and $\mu_0 \neq \mu_1$ mainly affects the estimation of β_0 which then often has a greater relative MSE than for independent covariates. Concerning $\hat{\beta}_1^{full}$ this can only be observed with the Student covariate distribution. Another difference between these two components is that the MDX mechanism results in smaller relative MSEs of $\hat{\beta}_0^{full}$ whereas there is no obvious effect of the non–MCAR mechanisms on $\hat{\beta}_1^{full}$. A general result is that the relative MSEs of $\hat{\beta}_0^{full}$ and $\hat{\beta}_1^{full}$ are greater for the χ^2 covariate distribution than for the Gaussian or Student while being sim-

ilar for the latter two. Additionally, the relative MSEs are greatest for the MCAR mechanism being roughly about 2.7. But this is mainly due to the higher global missing rate of this mechanism.

The last aspect can also be observed for $\hat{\beta}_2^{full}$ where the relative MSE is about 5 for the MCAR and about 2.5 for the other missing mechanisms. The dependence or independence of the covariates does not seem to affect the estimation of β_2 nor does the influence of X_2. This means that if $\beta_2 = 0$ the relative MSE is not generally smaller, with some exceptions for $\mu_0 = \mu_1$ and X_2 following a Gaussian or Student distribution. But here we already observe greater relative MSEs for the χ^2 distribution when $\mu_0 = \mu_1$ regardless of the true value of β_2. In the other situations the estimation is not clearly affected by the different covariate distributions although the maximal relative MSEs of more than 6 only occur within the Student and χ^2 distribution (and MCAR mechanism). The results for the MCAR mechanism are interesting because the asymptotic relative efficiency of the efficient complete data estimator and the complete case estimator equals 3.3 since both are consistent and the missing rate is 0.3. Thus, for finite sample size (N=200) and certain parameter constellations the relative MSE of the semiparametric efficient estimator may be greater than the asymptotic MSE of the complete case estimator.

Table 5:
Relative MSE and bias of the complete data maximum likelihood estimator. (The bias is given in brackets in italics.)

covariate distrib. = Gaussian		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{full}$	$\hat{\beta}_1^{full}$	$\hat{\beta}_2^{full}$	$\hat{\beta}_0^{full}$	$\hat{\beta}_1^{full}$	$\hat{\beta}_2^{full}$
MCAR, $\beta_2 =$	-1.5	2.19 *(-0.02)*	2.34 *(0.04)*	5.03 *(-0.02)*	2.79 *(0.01)*	3.72 *(0.01)*	5.80 *(-0.03)*
	0	1.14 *(-0.00)*	1.12 *(0.01)*	4.79 *(-0.01)*	2.15 *(-0.01)*	2.48 *(0.03)*	3.99 *(-0.01)*
	1.5	2.44 *(-0.01)*	2.50 *(0.04)*	5.37 *(-0.05)*	2.52 *(0.01)*	2.38 *(0.01)*	5.49 *(0.07)*
MDX, $\beta_2 =$	-1.5	1.13 *(0.00)*	1.43 *(0.04)*	1.99 *(-0.04)*	1.31 *(-0.00)*	1.88 *(0.04)*	2.31 *(-0.05)*
	0	1.02 *(-0.01)*	1.04 *(0.02)*	1.89 *(-0.00)*	1.29 *(-0.01)*	1.41 *(0.03)*	1.75 *(-0.00)*
	1.5	1.15 *(-0.01)*	1.56 *(0.05)*	2.00 *(0.05)*	1.29 *(-0.00)*	1.52 *(0.03)*	1.87 *(0.06)*
MDY, $\beta_2 =$	-1.5	1.50 *(-0.00)*	1.44 *(0.03)*	2.63 *(-0.04)*	1.93 *(-0.02)*	1.84 *(0.06)*	2.23 *(-0.05)*
	0	1.04 *(-0.00)*	1.03 *(0.02)*	2.24 *(0.01)*	1.55 *(-0.01)*	1.77 *(0.05)*	2.35 *(-0.00)*
	1.5	1.48 *(-0.01)*	1.49 *(0.02)*	2.36 *(0.04)*	1.35 *(0.01)*	1.62 *(0.03)*	2.56 *(0.06)*
MDXY, $\beta_2 =$	-1.5	1.58 *(-0.01)*	1.98 *(0.03)*	2.76 *(-0.04)*	1.97 *(-0.01)*	2.45 *(0.04)*	2.21 *(-0.04)*
	0	1.05 *(-0.00)*	1.08 *(0.01)*	2.55 *(-0.00)*	1.68 *(0.00)*	1.83 *(0.01)*	2.78 *(0.00)*
	1.5	1.52 *(0.00)*	1.90 *(0.02)*	2.55 *(0.04)*	1.50 *(0.03)*	1.46 *(-0.02)*	3.71 *(0.07)*

covariate distrib. $= \chi^2$		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{full}$	$\hat{\beta}_1^{full}$	$\hat{\beta}_2^{full}$	$\hat{\beta}_0^{full}$	$\hat{\beta}_1^{full}$	$\hat{\beta}_2^{full}$
MCAR, $\beta_2 =$	-1.5	3.14 *(-0.01)*	3.32 *(0.01)*	6.19 *(-0.06)*	3.25 *(-0.02)*	3.27 *(0.06)*	5.13 *(-0.06)*
	0	1.16 *(-0.01)*	1.12 *(0.03)*	5.44 *(0.01)*	2.77 *(0.00)*	2.84 *(0.01)*	5.99 *(0.00)*
	1.5	2.60 *(0.00)*	2.46 *(0.04)*	4.13 *(0.06)*	4.00 *(0.02)*	3.28 *(-0.01)*	6.31 *(0.06)*
MDX, $\beta_2 =$	-1.5	1.28 *(-0.00)*	2.21 *(0.00)*	2.29 *(-0.05)*	1.36 *(-0.04)*	1.55 *(0.08)*	2.00 *(-0.07)*
	0	1.03 *(0.00)*	1.05 *(0.02)*	2.10 *(0.01)*	1.27 *(0.02)*	1.40 *(-0.02)*	2.04 *(0.01)*
	1.5	1.21 *(0.02)*	1.39 *(0.01)*	1.75 *(0.06)*	1.37 *(0.01)*	1.37 *(-0.01)*	1.61 *(0.07)*
MDY, $\beta_2 =$	-1.5	1.97 *(-0.02)*	1.40 *(0.04)*	3.29 *(-0.04)*	2.67 *(-0.02)*	2.34 *(0.05)*	3.14 *(-0.06)*
	0	1.04 *(-0.00)*	1.03 *(0.01)*	2.40 *(0.00)*	1.59 *(-0.00)*	1.73 *(0.02)*	2.73 *(0.00)*
	1.5	1.45 *(0.02)*	1.60 *(0.03)*	2.20 *(0.06)*	1.84 *(0.04)*	1.68 *(-0.03)*	2.86 *(0.06)*
MDXY, $\beta_2 =$	-1.5	1.93 *(-0.01)*	2.79 *(0.02)*	3.09 *(-0.04)*	2.08 *(0.00)*	2.51 *(0.03)*	2.32 *(-0.05)*
	0	1.06 *(-0.00)*	1.08 *(0.03)*	2.68 *(0.01)*	1.53 *(-0.01)*	1.75 *(0.02)*	2.74 *(0.00)*
	1.5	1.58 *(0.02)*	2.31 *(0.00)*	2.74 *(0.04)*	2.01 *(0.03)*	1.66 *(-0.00)*	3.40 *(0.05)*

covariate distrib. = student		$\mu_0 = \mu_1 = 0$			$\mu_0 = -1, \mu_1 = 1$		
		$\hat{\beta}_0^{full}$	$\hat{\beta}_1^{full}$	$\hat{\beta}_2^{full}$	$\hat{\beta}_0^{full}$	$\hat{\beta}_1^{full}$	$\hat{\beta}_2^{full}$
MCAR, $\beta_2 =$	-1.5	2.34 *(-0.01)*	2.62 *(0.05)*	4.78 *(-0.07)*	2.92 *(-0.00)*	4.00 *(0.06)*	5.77 *(-0.06)*
	0	1.10 *(0.00)*	1.14 *(0.02)*	4.45 *(-0.00)*	1.89 *(-0.01)*	2.28 *(0.02)*	3.83 *(-0.00)*
	1.5	2.38 *(0.00)*	2.51 *(0.03)*	6.00 *(0.04)*	3.25 *(0.02)*	2.78 *(-0.01)*	6.76 *(0.05)*
MDX, $\beta_2 =$	-1.5	1.17 *(0.01)*	1.63 *(0.01)*	2.22 *(-0.05)*	1.35 *(-0.01)*	1.89 *(0.05)*	2.27 *(-0.06)*
	0	1.02 *(-0.01)*	1.04 *(0.04)*	1.90 *(-0.01)*	1.31 *(0.01)*	1.47 *(-0.00)*	2.12 *(0.01)*
	1.5	1.13 *(-0.01)*	1.55 *(0.04)*	2.31 *(0.05)*	1.27 *(0.02)*	1.65 *(0.01)*	2.10 *(0.07)*
MDY, $\beta_2 =$	-1.5	1.57 *(0.01)*	1.39 *(0.02)*	2.18 *(-0.05)*	2.01 *(-0.01)*	1.86 *(0.06)*	2.29 *(-0.06)*
	0	1.01 *(0.01)*	1.02 *(0.01)*	2.40 *(0.00)*	1.35 *(-0.00)*	1.53 *(0.02)*	2.24 *(0.00)*
	1.5	1.56 *(-0.02)*	1.49 *(0.03)*	2.57 *(0.04)*	1.30 *(0.01)*	1.52 *(0.00)*	2.53 *(0.06)*
MDXY, $\beta_2 =$	-1.5	1.65 *(-0.00)*	2.14 *(0.03)*	2.65 *(-0.05)*	2.34 *(-0.02)*	2.94 *(0.06)*	2.59 *(-0.05)*
	0	1.03 *(0.01)*	1.06 *(-0.01)*	2.52 *(-0.01)*	1.44 *(-0.01)*	1.60 *(0.02)*	2.54 *(-0.00)*
	1.5	1.59 *(-0.01)*	2.11 *(0.04)*	2.86 *(0.06)*	1.37 *(0.01)*	1.34 *(-0.01)*	3.08 *(0.08)*

The bias of the complete data maximum likelihood estimator is very small for all designs, as expected, i.e. in absolute value smaller than 0.09. But one should note that the largest values occur when estimating β_2 in case $\beta_2 \neq 0$. Therefore, these seem to be the 'difficult' situations.

5.3 Performance of the semiparametric estimators

In this section, we discuss the performance of the complete case, the corrected complete case, and the mean score estimators compared with the semiparametric efficient one. The results of the simulation study are not given in details.

5.3.1 The complete case estimator

For the designs where the complete case estimator is inconsistent we get that the bias of $\hat{\beta}_0^{CC}$ is always less than -1 for both missing mechanisms that depend on

the response variable whereas $\hat{\beta}_1^{CC}$ is inconsistent only for the MDXY mechanism showing a bias of more than 2. But even when $\hat{\beta}^{CC}$ is consistent the relative MSE is severely affected by discarding the incomplete cases, it often takes values between 0.55 and 0.8.

5.3.2 The corrected complete case estimator

The corrected complete case estimator produces nearly identical results as the semiparametric efficient one. The relative MSEs are nearly always between 0.99 and 1.00, exceptions arising only for the non–Gaussian covariate distributions when the missing mechanism depends on the response variable and $\beta_2 \neq 0$. But even then the relative MSE is at least 0.98. Concerning the bias we can observe the same pattern as for the semiparametric efficient estimator with a slight tendency to a greater bias of $\hat{\beta}_2^{CCC}$ for the missing mechanisms that depend on the response variable.

5.3.3 The mean score estimator

The mean score estimator is clearly dominated by the semiparametric efficient estimator. The relative MSE is almost always smaller than 1.00. The worst result is a relative MSE of 0.67 but in most cases it is still at least 0.8 and even greater than 0.9 for the MCAR missing mechanism. The main difficulty seems to concern the estimation for the MDX mechanism especially for $\mu_0 \neq \mu_1$. Here, the relative MSEs are roughly about 0.8.

Although the results are similar for the different covariate distributions it can be observed that in case of a non-MCAR mechanism, $\beta_2 = 0$ and $\mu_0 \neq \mu_1$ the relative MSE of all three components is in any case greater for the Gaussian covariate distribution than for the others.

6. Discussion

The main result of the simulation study concerns the performance of the ML estimator compared to the semiparametric efficient one proposed by Robins et al. (1994). On the one hand, we have seen that in the situation of a correct assumption about the covariate distribution and finite sample size the gain in efficiency by ML estima-

tion is only modest. On the other hand, this parametric approach can lead to serious bias if the assumed covariate distribution is 'far away' from the true one, where 'far away' means χ^2 instead of Gaussian. The Student distribution is in contrast similar enough to the Gaussian for the bias of the ML estimator to be negligible, at least for a sample size of 200. However, simulations with a sample size of 1000, which are not reported here, show a more serious bias of the ML estimator given a Student covariate distribution. As conclusion we propose the semiparametric efficient estimation as a very good alternative to the parametric approach. Despite its semiparametric efficiency being an asymptotic property, the performance appears to be satisfying also for finite sample size.

Another interesting result has been obtained for the corrected complete case estimator. It strongly supports the supposition that in the special case of a logistic regression where all variables except the incomplete one are discrete the estimator of Breslow and Cain (1988) is semiparametric efficient since it appears to be equivalent to the semiparametric efficient estimator of Robins et al. This has also been confirmed by simulations with a sample size of 1000 yielding nearly always identical results for both estimators. However, we have to restrict this result to the logistic regression model since it has been shown by Robins et al. (1994) that $\hat{\beta}^{CCC}$ is in general not semiparametric efficient.

A point which has not been addressed in this paper but that has to be taken into account is the possible misspecification of the missing mechanism. The discussed semiparametric approaches need an estimation of the observation probabilities which is given by (3). For continuous Y or X_1 there is no such straightforward procedure. Instead, a parametric model for the missing mechanism has to be assumed. As shown by Zhao et al. (1996) the correctness of this model is crucial in assuring the consistency of the semiparametric estimators.

References

Blackhurst, D.W., Schluchter, M.D. (1989): Logistic regression with a partially observed covariate. *Comm. in Statist. – Simulation and Computation* **18**, 163 – 177.

Breslow, N.E., Cain, K.C. (1988): Logistic regression for two–stage case–control data. *Biometrika* **75**, 11 – 20.

Cain, K.C., Breslow, N.E. (1988): Logistic regression analysis and efficient design for two–stage studies. *Am. J. of Epidem.* **128**, 1198 – 1206.

Carroll, R.J., Wand, M.P. (1991): Semiparametric estimation in logistic measurement error models. *JRSS B* **53**, 573 – 585.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977): Maximum likelihood estimation from incomplete data via the EM–algorithm. *JRSS B* **39**, 1 – 38.

Ibrahim, J.G. (1990): Incomplete data in generalized linear models. *JASA* **85**, 765 – 769.

Ibrahim, J.G., Weisberg, S. (1992): Incomplete data in generalized linear models with continuous covariates. *Austr. J. of Statist.* **34**, 461 – 470.

Little, J.A. (1992): Regression with missing X's: A review. *JASA* **87**, 1227 – 1237.

Pepe, M.S., Fleming, T.R. (1991): A nonparametric method for dealing with mismeasured covariate data. *JASA* **86**, 108 – 113.

Reilly, M., Pepe, M. (1995): A mean score method for missing and auxiliary covariate data in regression models. *Biometrika* **82**, 299 – 314.

Robins, J.M., Hsieh, F., Newey, W. (1995): Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. *JRSS B* **57**, 409 – 424.

Robins, J.M., Rotnitzky, A., Zhao, L.P. (1994): Estimation of regression coefficients when some regressors are not always observed. *JASA* **89**, 846 – 866.

Rubin, D.B. (1976): Inference and missing data. *Biometrika* **63**, 581 – 592.

Vach, W. (1994): *Logistic regression with missing values in the covariates.* Springer–Verlag, New York.

Vach, W., Illi, S. (1997): Biased estimation of adjusted odds ratios from incomplete covariate data due to the violation of the missing at random assumption. *Biometrical Journal* **39**, 13 –28

Zhao, L.P., Lipsitz, S. (1992): Design and analysis of two–stage designs. *Stat. in Med.* **11**, 769 – 782.

Zhao, L.P., Lipsitz, S., Lew, D. (1996): Regression analysis with missing covariate data using estimating equations. *Biometrics* **52**, 1165 – 1182.