Hornsteiner, Hamerle, Michels:

Parametric versus Nonparametric Treatment of Unobserved Heterogeneity in Multivariate Failure Times

Sonderforschungsbereich 386, Paper 80 (1997)

Parametric versus Nonparametric Treatment of Unobserved Heterogeneity in Multivariate Failure Times

Ulrich Hornsteiner * , Alfred Hamerle * and Paul Michels **

Abstract

Two contrary methods for the estimation of a frailty model of multivariate failure times are presented. The assumed Accelerated Failure Time Model includes censored data, observed covariates and unobserved heterogeneity. The parametric estimator maximizes the marginal likelihood whereas the method which does not require distributional assumptions combines the GEE approach (Liang and Zeger, 1986) with the Buckley-James (1979) estimator for censored data. Monte Carlo experiments are conducted to compare the methods under various conditions with regard to bias and efficiency. The ML estimator is found to be rather robust against some misspecifications and both methods seem to be interesting alternatives in uncertain circumstances which lack exact solutions. The methods are applied to data of recurrent purchase acts of yogurt brands.

Key words: Multivariate failure times; accelerated failure time model; censored data; unobserved heterogeneity; generalized estimating equations; marginal maximum likelihood estimation; misspecification; simulation study; individual purchase behavior.

1 Introduction

Multivariate failure time data may occur in the context of economics, sociology, medicine or other sciences – either when we observe different but related elementary units and investigate the time it takes for each of them until a certain event takes place – or when we observe the units for a rather long space of time and are interested in the time intervals between events of various kinds or between recurrent events of the same kind. In addition to the proper handling of censored data, multivariate failure time data require appropriate models and methods to take into account that the failure times within one family, one unit or one "block" are correlated. Usually it is supposed that a set of variables which are constant within one block but differ from one block to the other and which are not observed (the "unobserved heterogeneity" or "frailty") influences the failure times and so is responsible for the correlation. Some standard texts on the statistical theory of failure time data include sections about multivariate failure time data such as Kalbfleisch and Prentice (1980), Lawless (1982), Andersen, Borgan, Gill and Keiding (1993) and Fahrmeir, Hamerle and Tutz (1996).

*Lehrstuhl für Statistik, Wirtschaftswissenschaftliche Fakultät, Universität Regensburg.
U. Hornsteiner is supported by the Deutsche Forschungsgemeinschaft, SFB 386.
**A.C. Nielsen GmbH, Frankfurt/Main.

There is a series of recently published articles on various treatments of unobserved heterogeneity in multivariate failure times, particularly of recurrent events (Nielsen, Gill, Andersen and Sørensen, 1992, Davies, 1993, dos Santos, Davies and Francis, 1995, Pickles and Crouchley, 1995, Haider and Davies, 1996). As nobody knows the real distribution of frailties, they all in principle make contributions to the controversy if it is better to model the frailty nonparametrically or to assume a perhaps incorrect family of distributions. The answers are partly contradictory, depend on the analysed datasets and on the decision if heterogeneity is treated as a nuisance (for better estimations of the regression parameters) or as a value of interest per se. A common conclusion is that in a particular application one should consider if it is more important to model the hazard nonparametrically and to assume a frailty distribution or vice versa (Pickles and Crouchley, 1995, p. 1458, dos Santos, Francis and Davies, 1995, p. 125).

Our intention in this paper is to refer to another approach (Hornsteiner and Hamerle, 1996) which treats both the frailty and the hazard nonparametrically by combining the "generalized estimating equations" (Liang and Zeger, 1986) with the handling of censored data like Buckley and James (1979) and to compare it with a fully parametric method, the marginal maximum likelihood estimation. For the latter we assume a normally distributed frailty and three versions of the hazard distribution. Some Monte Carlo experiments shall give additional answers to the question described above.

The remainder of the paper is organized as follows. In section 2 we present a general extension of an Accelerated Failure Time Model for recurrent events including unobserved heterogeneity. An estimation method which manages without any distributional or structural assumptions, the combined GEE/Buckley-James method, is described in section 3. In contrast, we derive the fully parametric procedure in section 4. The main part of the paper is the simulation study in section 5 where estimation results for simulated data according to various "true" distributions are reported. In section 6 the methods are applied to German household panel data on recurrent yogurt purchase behavior based on home scanners. Section 7 concludes with a summarizing discussion.

2 The model

We consider an extension of an Accelerated Failure Time Model for recurrent events. All the methods in this paper can also be applied to failure times which are correlated because of belonging to groups of different but related elementary units. The restriction to recurrent events is just for simplicity.

We have observed N elementary units ($n = 1, \ldots, N$) with a varying number of K_n spells per unit ($k = 1, \ldots, K_n$). The logarithm of every failure time

$$y_{nk} = \ln(T_{nk}) = x'_{nk}\beta + \sigma_\alpha \alpha_n + \sigma_\varepsilon \varepsilon_{nk}$$

depends linearly on a vector x_{nk} of P covariates (including a 1 for the intercept) – which may partly be constant within the block and partly vary from spell to spell – and a P-dimensional vector of regression parameters $\beta = (\beta_1, \ldots, \beta_P)'$, $p = 1, \ldots, P$.

The stochastic component consists of a frailty term α_n which absorbs non-observed covariates which are constant for all the spells of one unit and an error term ε_{nk}. The α_n are assumed to be independently and identically distributed. The distribution of ε_{nk} is assumed to be one of the distributions used in Accelerated Failure Time Models. More concrete specifications of the distributions including consequences of misspecifications will be the main topic of the simulation studies in section 5.

Instead of T_{nk} we observe

$$z_{nk} = \min(T_{nk}, c_{nk})$$

where c_{nk} is a censor value, together with an indicator variable

$$\delta_{nk} = \begin{cases} 1 & \text{if} \quad T_{nk} \leq c_{nk} \\ 0 & \text{if} \quad T_{nk} > c_{nk}. \end{cases}$$

In the case of recurrent events usually there is no censor value given for each spell but it is the limited total observation period C_n which is responsible for censoring. C_n may be determined by the design of the study or may result from death or drop-out of the observed unit. We assume for simplicity that the beginning of the observation period coincides with an event, in other words there are no left-censored spells. That yields

$$c_{nk} = C_n - \sum_{l=1}^{k-1} T_{nl} \quad \forall n = 1, \ldots, N \;\; \forall k = 1, \ldots, K_n.$$

Thus the last spell of each unit is censored almost surely.

3 The GEE/Buckley-James method

Recently published literature on modelling multivariate failure time data with unobserved heterogeneity varies a lot in distributional assumptions of both the frailty and the hazard and yields contradictory results on advantages and disadvantages of parametric and nonparametric approaches, see e.g. dos Santos, Davies and Francis (1995) and Pickles and Crouchley (1995).

For these reasons our aim was to develop an estimator of the regression parameters which is robust also if we can not be sure about the distributional assumptions of frailty and hazard and about the correlation structure of the failure times within one unit. The considerations resulted in a combination of the GEE approach for longitudinal data (Liang and Zeger, 1986) with the nonparametric Least Squares method of Buckley and James (BJ, 1979) for censored data (Hornsteiner and Hamerle, 1996).

The method is based on the *generalized estimating equations*

$$\sum_{n=1}^{N} X_n' V_n^{-1} (y_n^* - X_n \hat{\beta}) = 0,$$

where X_n is the matrix containing the lines x_{nk}', $k = 1, \ldots, K_n$, and $y_n^* = (y_{n1}^*, \ldots, y_{nK_n}^*)'$ consists of the *pseudo variables*

$$y_{nk}^* = \delta_{nk} \ln z_{nk} + (1 - \delta_{nk}) \hat{E}(y_{nk} \mid y_{nk} > \ln z_{nk}).$$

Thus for uncensored spells $y_{nk}^* = y_{nk}$, whereas for censored spells the conditional expectation of y_{nk} given survival up to the censor time is substituted by the nonparametric product limit estimator (Kaplan and Meier, 1958).

Furthermore $V_n = R_{K_n}(\gamma)/\phi$, where $R_{K_n}(\gamma)$ is a working correlation matrix, γ a vector that fully characterizes the correlation structure, and $1/\phi = Var(y_{nk}^*) := v$ is assumed to be constant. The most suitable working correlation structure when there is no further information about the residuals is the equicorrelation structure ("equ") with $\gamma = Corr(y^*_{nk}, y^*_{nl})$, and $c := Cov(y^*_{nk}, y^*_{nl}) = v \cdot \gamma$. An alternative for purpose of comparison is the very simple specification $\gamma = c = 0$, the independence structure ("ind").

We implemented an iterative algorithm which consists of three steps in each iteration. We get an initial estimation simply by $\hat{\beta}^{(0)} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$ where $\tilde{X}$ and $\tilde{y}$ solely include the uncensored spells. The u^{th} iteration ($u = 1, 2, \ldots$) is a sequence of the renewal of the Buckley-James pseudo variables, the moment estimation of γ and v, and finally a modified Fisher scoring for β,

$$\hat{\beta}^{(u)} = \hat{\beta}^{(u-1)} + \left(\sum_{n=1}^{N} X'_n \tilde{V}_n^{-1}(\hat{\beta}^{(u-1)}) \frac{\partial y^*_n}{\partial \beta}\right)^{-1} \left(\sum_{n=1}^{N} X'_n \tilde{V}_n^{-1}(\hat{\beta}^{(u-1)}) \left(y^*_n - X'_n \hat{\beta}^{(u-1)}\right)\right).$$

The algorithm has to be broken off successfully not only in the case of convergence but also if it has become stabilized in a loop of several values, see e.g. Miller (1981), p. 152 and Currie (1996). In this case the averages of the relevant values are taken as estimators.

The asymptotic covariance matrix of the parameter estimates and the standard error of the pth parameter, respectively, are estimated by the robust "sandwich" formula

$$\widehat{Cov}(\hat{\beta}) = \left(\sum_{n=1}^{N} X'_n \tilde{V}_n^{-1} \frac{\partial y^*_n}{\partial \beta}\right)^{-1} \left(\sum_{n=1}^{N} X'_n \tilde{V}_n^{-1} \widehat{Cov}(y^*_n) \tilde{V}_n^{-1} X_n\right) \left(\sum_{n=1}^{N} X'_n \tilde{V}_n^{-1} \frac{\partial y^*_n}{\partial \beta}\right)^{-1}$$

and

$$\hat{\sigma}(\hat{\beta}_p) = \sqrt{\left(\text{diag}(\widehat{Cov}(\hat{\beta}))\right)_p}$$

where $\widehat{Cov}(y^*_n) = (y^*_n - x'_n\beta)\,(y^*_n - x'_n\beta)'$.

4 Maximizing the marginal likelihood

In contrast to the nonparametric method of the previous section we now add some restrictive assumptions to the model of section 2 about the distributions of the error terms and their dependence structure. These assumptions are:

- The α_n are independent and normally distributed with mean 0 and variance 1.
- The ε_{nk} are independent and identically distributed by any known distribution.
- α_n and ε_{nk} are independent.

This enables us to "integrate out" the unknown heterogeneity and to maximize the likelihood which is then marginal relative to the heterogeneity deviation parameter.

Some comments on these assumptions are to be stated. The normal distribution assumption for the heterogeneity is not the usual one in the context of survival analysis. If a distributional assumption is made, it is generally that $\exp(\alpha_n)$ is for mathematical or computational reasons gamma distributed (see e.g. Lancaster and Nickell, 1980, Follmann and Goldberg, 1988, Meyer, 1990, Clayton, 1991, Schneider, 1997, Scheike and

Jensen, 1997), but nowhere arguments are given which are derived from the application context. Normally distributed frailties are considered by Kiefer (1983), Davies (1993), Crouchley and Pickles (1995), Pickles and Crouchley (1995), dos Santos, Davies and Francis (1995), Haider and Davies (1996).

Without loss of generality we have to fix mean and variance, as other values are absorbed by the intercept β_1 and the heterogeneity deviation σ_α, respectively.

Assuming the ε_{nk} to be independent is analogous to the equicorrelation structure in the GEE approach. The behavior of the estimators under alternative correlation structures (see Spieß, Nagl and Hamerle, 1997) is not a topic of this paper.

The distribution of ε_{nk} has to be specified for further computation and implementation what is equivalent to modelling the hazard rate given observed covariates and unobserved heterogeneity. There are rarely concrete reasons for a specific distribution but common settings are the normal, logistic or extreme value distribution which lead to the lognormal (N), the log-logistic (L) or the Weibull (W) model, respectively. We have implemented these three versions.

The log-likelihood has the form

$$l(\beta, \sigma_\alpha, \sigma_\varepsilon \mid X_1, \ldots, X_N, z_{11}, \ldots, z_{N,K_N}, \delta_{11}, \ldots, \delta_{N,K_N}) = \text{const}+$$
$$\sum_{n=1}^{N} \ln \int_{-\infty}^{\infty} \exp\left(\sum_{k=1}^{K_n} (\delta_{nk} \cdot \ln f_{nk}(z_{nk}) + (1-\delta_{nk}) \cdot \ln S_{nk}(z_{nk}))\right) \cdot \varphi(\alpha)\, d\alpha,$$

where the density f_{nk} and the survivor function S_{nk} are specified according to the model assumption (Fahrmeir, Hamerle and Tutz, 1996, pp. 310-312):

(N)

$$f_{nk}(z_{nk}) = \frac{1}{\sigma_\varepsilon z_{nk}} \varphi\left(\frac{\ln z_{nk} - x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon}\right),$$

$$S_{nk}(z_{nk}) = 1 - \Phi\left(\frac{\ln z_{nk} - x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon}\right);$$

(L)

$$f_{nk}(z_{nk}) = \frac{1}{\sigma_\varepsilon} \exp\left(\frac{-x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon}\right) \cdot z_{nk}^{\frac{1}{\sigma_\varepsilon}-1} \cdot \left(1 + \exp\left(\frac{-x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon}\right) \cdot z_{nk}^{\frac{1}{\sigma_\varepsilon}}\right)^{-2},$$

$$S_{nk}(z_{nk}) = \left(1 + \exp\left(\frac{-x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon}\right) \cdot z_{nk}^{\frac{1}{\sigma_\varepsilon}}\right)^{-1};$$

(W)

$$f_{nk}(z_{nk}) = \frac{1}{\sigma_\varepsilon} \exp\left(\frac{-x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon} - \exp\left(\frac{-x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon}\right) \cdot z_{nk}^{\frac{1}{\sigma_\varepsilon}}\right) \cdot z_{nk}^{\frac{1}{\sigma_\varepsilon}-1},$$

$$S_{nk}(z_{nk}) = \exp\left(-\exp\left(\frac{-x'_{nk}\beta - \sigma_\alpha \alpha}{\sigma_\varepsilon}\right) \cdot z_{nk}^{\frac{1}{\sigma_\varepsilon}}\right).$$

The appearing integral cannot be solved analytically. We approximate the log-likelihood and its deviations by Gauß-Hermite quadrature (Bock and Lieberman, 1970, Butler and Moffit, 1982, Spieß, 1995, Spieß and Hamerle, 1995). Applied to the present problem we have

$$\int_{-\infty}^{\infty} f(\alpha)\cdot\varphi(\alpha)\,d\alpha = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(\alpha)\exp(-\frac{\alpha^2}{2})\,d\alpha \approx \frac{1}{\sqrt{\pi}}\sum_{g=1}^{G} w_g f(\sqrt{2}h_g),$$

where G is the number of evaluation points h_g, and w_g is the weight given to the g^{th} evaluation point, $g = 1,\ldots,G$ (Stroud and Secrest, 1966, pp. 217-252).

To maximize the log-likelihood the Newton-Raphson method together with a line search method for global convergence is used (Dennis and Schnabel, 1983). Provided that the model is specified correctly and enough evaluation points are included, the maximum likelihood estimator is consistent and asymptotically normally distributed. As the required number of evaluation points is not discussed in this paper, for all simulations and empirical analyses the sufficient number of $G = 64$ is used.

5 Simulation studies

To study the properties of the developed estimation methods in finite samples under correct as well as incorrect specification, simulation and estimation programs have been written in SAS/IML, version 6 (SAS Institute Inc., 1989, 1990). The studies reported here intend to show the behaviors in large sample sizes ($N = 500$) when the distribution of the heterogeneity is varied. We do not discuss here other interesting topics such as the required number of evaluation points, the computation time, the behavior in small sample sizes, the behavior in the case of time-varying but wrongly as time-constant specified covariates, the dependence on the correct specification of the correlation structure, on the type of censoring, or on the censor rate.

In one running of the program $S = 200$ data sets are simulated. In every simulation, $N = 500$ blocks of covariates and failure times are produced. The design matrices consist of the column of ones and four stochastic regressors, two of them (x2 and x3) constant within one unit, the second two (x4 and x5) varying from spell to spell. In each of the two cases, one of the two is a metric, normally distributed variable having mean and variance one (x2 and x4), the other one is a dichotomous variable taking the two values zero and one with probability 0.5 each (x3 and x5). The regression coefficients $\beta_2,\ldots,\beta_5$ as well as the intercept β_1 are specified as 0.5 each.

The distribution of the heterogeneity α_n is varied (see later in the text) so that the effects on the estimation results can be studied. The distribution of the error term ε_{nk} is specified as normal. Misspecification concerning the hazard can be studied by applying the ML method specifying the log-logistic or Weibull model.

The total observation period C_n decides about the number of spells K_n and about the censoring of the last spell. Given the covariates and the parameters, the observation period has been chosen as $C_n \equiv C = 16.9$ so that the overall average number of spells resulted in about three per unit and the censor rate in about 0.33.

Finally, the estimation parts of the programs require convergence criteria, a maximum number of iterations, and – in the case of GEE/BJ estimation – the specification of the type of the working correlation matrix (ind or equ) as described in section 3, or

– in the case of ML estimation – the model specification (N, L or W), the number of evaluation points and initial values for the parameters to be estimated.

The output contains the mean of the estimated parameter vectors,

$$\overline{\hat{\beta}} = \frac{1}{S} \sum_{s=1}^{S} \hat{\beta}_s,$$

and the root mean squared error

$$\text{RMSE}(\hat{\beta}_p) = \left(\frac{1}{S-1} \sum_{s=1}^{S} \left(\hat{\beta}_{sp} - \beta_p \right)^2 \right)^{1/2}$$

and the standard deviation

$$s(\hat{\beta}_p) = \left(\frac{1}{S-1} \sum_{s=1}^{S} \left(\hat{\beta}_{sp} - \overline{\hat{\beta}}_p \right)^2 \right)^{1/2}$$

of each estimated parameter ($p = 1, \ldots, P$), estimated over the S simulations. The latter is used as quality criterion in comparison with the mean of the estimated standard deviations of $\hat{\beta}_p$,

$$\overline{\hat{\sigma}(\hat{\beta}_p)} = \frac{1}{S} \sum_{s=1}^{S} \left(\hat{\sigma}(\hat{\beta}_p) \right)_s .$$

In addition to the numerical assessment of these values we apply statistical tests to control normality and bias of the parameter estimates. The SAS procedure PROC UNIVARIATE provides us with the p-values of the Shapiro-Wilk statistics testing each component of the estimated parameter vector if it is a random sample from a normal distribution. To test if a bias can be explained by the random character of the simulations or if it is significant, t-tests for the null hypotheses $E(\hat{\beta}_p) = \beta_p$, $p = 1, \ldots, P$, are implemented in the IML program.

Table 1 shows the estimation results of the various methods in the case of correct specified frailty, that means the datasets are indeed generated by $\alpha_n \sim N(0, 1)$ with $\sigma_\alpha = \sqrt{0.05} = 0.2236$. As already mentioned, ε_{nk} and σ_ε have the same specifications. Thus the ML(N) estimator additionally specifies the hazard correctly whereas it is misspecified by the ML(L) and ML(W) estimators. As the standard deviations of the "standard" logistic and extreme value distributions differ from 1, the differences of the estimated σ_ε are systematic, and root mean squared errors concerning this parameter would be senseless. Moreover, the extreme value distribution is shifted from the expectation 0 which is absorbed by the intercept term.

The deviations of the estimated regression parameters from the true value 0.5 and particularly the p-values of the t-tests show that the GEE/BJ estimator using the independence assumption of nearly all parameters is clearly biased, whereas the hypothesis of unbiased estimators is not rejected on the 1% level of significance when we use the equicorrelation working structure. Of course, the correctly specified ML(N) estimator is not biased in this case of a relative large sample size. Misspecifying the hazard by the log-logistic model does not result in any biased parameters. However, the assumption of the Weibull model produces biases particularly of the parameters β_2 and β_3 concerning the block-constant covariates. This is not surprising when we consider that

Table 1: *Comparison of GEE/BJ and ML in the case of correct specification of the frailty but "correct" vs. "incorrect" specification of the hazard ($\alpha_n \sim N(0,1)$, $\varepsilon_{nk} \sim N(0,1)$, $\sigma_\alpha = \sigma_\varepsilon = \sqrt{0.05} = 0.2236$): Means, rmse and standard deviations of the parameter estimations, means of the estimated standard deviations and p-values over $S = 200$ simulations*

	GEE/BJ		ML		
	ind	equ	(N)	(L)	(W)
$\overline{\hat{\beta}_1} = 0.$	4566	4961	5005	4997	6111
$\overline{\hat{\beta}_2} = 0.$	5053	4990	4990	4992	4951
$\overline{\hat{\beta}_3} = 0.$	5066	4997	4994	4998	4946
$\overline{\hat{\beta}_4} = 0.$	5023	5016	4997	4997	4988
$\overline{\hat{\beta}_5} = 0.$	5006	4997	4981	4986	4977
$\overline{\hat{\sigma}_\alpha} = 0.$			2229	2234	2343
$\overline{\hat{\sigma}_\varepsilon} = 0.$			2229	1255	1857
$\overline{\hat{v}} = 0.$	0751	0751			
$\overline{\hat{c}} = 0.$		0369			
$\text{RMSE}(\hat{\beta}_1) = 0.$	0533	0266	0262	0263	-
$\text{RMSE}(\hat{\beta}_2) = 0.$	0167	0137	0136	0136	0148
$\text{RMSE}(\hat{\beta}_3) = 0.$	0308	0275	0274	0277	0278
$\text{RMSE}(\hat{\beta}_4) = 0.$	0114	0098	0088	0088	0092
$\text{RMSE}(\hat{\beta}_5) = 0.$	0175	0157	0148	0150	0174
$\text{RMSE}(\hat{\sigma}_\alpha) = 0.$			0119	0116	0156
$\text{RMSE}(\hat{\sigma}_\varepsilon) = 0.$			0060	-	-
$s(\hat{\beta}_1) = 0.$	0310	0264	0262	0263	0268
$s(\hat{\beta}_2) = 0.$	0158	0136	0136	0136	0140
$s(\hat{\beta}_3) = 0.$	0301	0275	0274	0277	0273
$s(\hat{\beta}_4) = 0.$	0112	0097	0088	0088	0091
$s(\hat{\beta}_5) = 0.$	0175	0157	0147	0149	0173
$s(\hat{\sigma}_\alpha) = 0.$			0119	0116	0114
$s(\hat{\sigma}_\varepsilon) = 0.$			0059	0033	0056
$\overline{\hat{\sigma}(\hat{\beta}_1)} = 0.$	0299	0292	0239	0240	0244
$\overline{\hat{\sigma}(\hat{\beta}_2)} = 0.$	0175	0170	0137	0137	0141
$\overline{\hat{\sigma}(\hat{\beta}_3)} = 0.$	0329	0320	0262	0263	0268
$\overline{\hat{\sigma}(\hat{\beta}_4)} = 0.$	0110	0099	0083	0083	0083
$\overline{\hat{\sigma}(\hat{\beta}_5)} = 0.$	0211	0186	0158	0158	0157
$\overline{\hat{\sigma}(\hat{\sigma}_\alpha)} = 0.$			0117	0117	0112
$\overline{\hat{\sigma}(\hat{\sigma}_\varepsilon)} = 0.$			0063	0040	0056
p-values H_0:					
$\mu_1 = \beta_1$ (0. ....)	0000	0368	8062	8550	-
$\mu_2 = \beta_2$ (0. ....)	0000	2784	2782	3911	0000
$\mu_3 = \beta_3$ (0. ....)	0023	8626	7665	9065	0055
$\mu_4 = \beta_4$ (0. ....)	0039	0173	6095	6776	0601
$\mu_5 = \beta_5$ (0. ....)	6158	7949	0677	1816	0564

the normal and the logistic distributions are more similar compared to each other than to the extreme value distribution.

Despite of these findings, the deviations of the parameter estimates from the true values are so moderate that the root mean squared errors in no case differ much from the standard deviations. They both are well estimated by the ML method independent from the hazard specification whereas they are overestimated by the GEE/BJ procedure (see also Hornsteiner and Hamerle, 1996, pp. 8-11).

In no case we found any significant deviations of the regression parameter estimates from the normal distribution.

Finally, the various methods can be compared to each other with respect to efficiency. Starting from the GEE/BJ estimator using independence assumption there is an obvious decrease of the rmse applying any of the other methods considered. The rmse of the GEE/BJ(equ), the ML(N) and the ML(L) estimators of the parameters concerning block-constant covariates are nearly identical. We observe little additional gain of efficiency applying the ML(N) or ML(L) method on the parameters β_4 and β_5 which correspond to the spell-varying covariates. On the other hand, misspecifying the hazard by the Weibull model results in slightly higher rmse for some parameters.

To study the behavior of the ML estimator under misspecified frailty distribution in comparison to the GEE/BJ estimator, we simulated data sets under the same conditions as in table 1 except the frailty distribution. On the background of the application in section 6, there was the idea of a bimodal frailty. We could interpret this as follows. The population is separated into two parts. We cannot observe to which part a certain unit belongs but this membership influences the risk of failure. Additionally we assume an overlayed continuous variable which also has influence but is unobserved.

For the simulations in table 2 we assume two parts of the same size that means a symmetric bimodal heterogeneity distribution. To be more concrete, it is a mixture of two normal distributions, each taken by a probability of 0.5, one of them has mean 0.2, the other -0.2, both have a standard deviation of 0.1. When σ_α is specified as 1, the resulting distribution has mean 0 and standard deviation $\sqrt{0.05}$.

Again, the GEE/BJ estimator using the independence assumption is clearly biased. But now the improvement by the equicorrelation assumption is not as satisfying as it was in the case of normally distributed frailty.

Surprisingly, all the three ML specifications result in estimates which seem to be unbiased for nearly all parameters concerning the t-tests although the ML(N) estimator assumes an incorrectly distributed frailty and the ML(L) and the ML(W) estimators assume both an incorrect frailty and an incorrect hazard. The Shapiro-Wilk tests again do not reject the hypotheses of normally distributed estimates. The standard deviations and the root mean squared errors are again well estimated by the ML specifications and overestimated by the GEE/BJ method. With respect to efficiency, the GEE/BJ(equ) estimator performs much better than the independence version and nearly as well as all the ML estimators.

For the simulations in table 3 we also assumed a separation into two parts – in contrast to table 2 asymmetrically and without overlayed continuous distribution. Exactly, the heterogeneity is created by a Bernoulli distribution with $p = 0.9472$, and the resulting 0/1 variable is for purpose of comparison shifted by $-p$ so that the resulting variable has mean 0 and standard deviation $\sqrt{0.05}$. The intention was a distribution which differs from the theoretical normal assumption particularly by curtosis.

Table 2: *Comparison of GEE/BJ and ML in the case of "incorrect" specification of the frailty I: Mixture of two normal distributions (see text for further explanation),* $\varepsilon_{nk} \sim N(0,1)$, $\sigma_\varepsilon = \sqrt{0.05} = 0.2236$*: Means, rmse and standard deviations of the parameter estimations, means of the estimated standard deviations and p-values over* $S = 200$ *simulations*

	GEE/BJ		ML		
	ind	equ	(N)	(L)	(W)
$\overline{\hat{\beta}_1} = 0.\ ....$	4547	4892	4958	4959	6058
$\overline{\hat{\beta}_2} = 0.\ ....$	5069	5022	5020	5020	4985
$\overline{\hat{\beta}_3} = 0.\ ....$	5092	5028	5021	5020	4985
$\overline{\hat{\beta}_4} = 0.\ ....$	5025	5024	5004	5003	4997
$\overline{\hat{\beta}_5} = 0.\ ....$	5027	5034	5013	5014	5007
$\overline{\hat{\sigma}_\alpha} = 0.\ ....$			2227	2243	2345
$\overline{\hat{\sigma}_\varepsilon} = 0.\ ....$			2232	1256	1860
$\overline{\hat{v}} = 0.\ ....$	0738	0739			
$\overline{\hat{c}} = 0.\ ....$		0321			
$\mathrm{RMSE}(\hat{\beta}_1) = 0.\ ....$	0530	0280	0259	0260	-
$\mathrm{RMSE}(\hat{\beta}_2) = 0.\ ....$	0171	0139	0138	0138	0141
$\mathrm{RMSE}(\hat{\beta}_3) = 0.\ ....$	0304	0274	0269	0271	0269
$\mathrm{RMSE}(\hat{\beta}_4) = 0.\ ....$	0104	0098	0089	0090	0097
$\mathrm{RMSE}(\hat{\beta}_5) = 0.\ ....$	0179	0165	0159	0161	0175
$\mathrm{RMSE}(\hat{\sigma}_\alpha) = 0.\ ....$			0100	0099	0146
$\mathrm{RMSE}(\hat{\sigma}_\varepsilon) = 0.\ ....$			0065	-	-
$s(\hat{\beta}_1) = 0.\ ....$	0276	0259	0255	0257	0259
$s(\hat{\beta}_2) = 0.\ ....$	0157	0137	0136	0137	0140
$s(\hat{\beta}_3) = 0.\ ....$	0290	0272	0268	0270	0269
$s(\hat{\beta}_4) = 0.\ ....$	0101	0096	0089	0090	0097
$s(\hat{\beta}_5) = 0.\ ....$	0177	0162	0159	0160	0175
$s(\hat{\sigma}_\alpha) = 0.\ ....$			0099	0099	0097
$s(\hat{\sigma}_\varepsilon) = 0.\ ....$			0065	0037	0062
$\overline{\hat{\sigma}(\hat{\beta}_1)} = 0.\ ....$	0287	0284	0239	0240	0243
$\overline{\hat{\sigma}(\hat{\beta}_2)} = 0.\ ....$	0169	0169	0137	0137	0142
$\overline{\hat{\sigma}(\hat{\beta}_3)} = 0.\ ....$	0316	0317	0262	0264	0269
$\overline{\hat{\sigma}(\hat{\beta}_4)} = 0.\ ....$	0109	0098	0083	0083	0083
$\overline{\hat{\sigma}(\hat{\beta}_5)} = 0.\ ....$	0210	0186	0158	0158	0157
$\overline{\hat{\sigma}(\hat{\sigma}_\alpha)} = 0.\ ....$			0119	0118	0113
$\overline{\hat{\sigma}(\hat{\sigma}_\varepsilon)} = 0.\ ....$			0063	0040	0057
p-values H_0:					
$\mu_1 = \beta_1$ (0. ....)	0000	0000	0199	0242	-
$\mu_2 = \beta_2$ (0. ....)	0000	0265	0418	0400	1448
$\mu_3 = \beta_3$ (0. ....)	0000	1493	2765	2941	4174
$\mu_4 = \beta_4$ (0. ....)	0006	0005	5322	6818	6459
$\mu_5 = \beta_5$ (0. ....)	0328	0035	2664	2332	5518

Table 3: *Comparison of GEE/BJ and ML in the case of "incorrect" specification of the frailty II: Two strongly different clusters (see text for further explanation),* $\varepsilon_{nk} \sim N(0,1)$, $\sigma_\varepsilon = \sqrt{0.05} = 0.2236$*: Means, rmse and standard deviations of the parameter estimations, means of the estimated standard deviations and p-values over* $S = 200$ *simulations*

	GEE/BJ		ML		
	ind	equ	(N)	(L)	(W)
$\overline{\hat{\beta}_1} = 0.\ ....$	4295	5100	5051	5049	6103
$\overline{\hat{\beta}_2} = 0.\ ....$	5087	5047	5011	5019	4966
$\overline{\hat{\beta}_3} = 0.\ ....$	5186	5101	5062	5070	5014
$\overline{\hat{\beta}_4} = 0.\ ....$	5028	4999	5009	5009	4998
$\overline{\hat{\beta}_5} = 0.\ ....$	5017	4982	5001	5000	4994
$\overline{\hat{\sigma}_\alpha} = 0.\ ....$			2723	2705	2652
$\overline{\hat{\sigma}_\varepsilon} = 0.\ ....$			2176	1223	1838
$\overline{\hat{v}} = 0.\ ....$	1212	1216			
$\overline{\hat{c}} = 0.\ ....$		1607			
$\mathrm{RMSE}(\hat{\beta}_1) = 0.\ ....$	0855	0301	0263	0261	-
$\mathrm{RMSE}(\hat{\beta}_2) = 0.\ ....$	0331	0174	0164	0164	0161
$\mathrm{RMSE}(\hat{\beta}_3) = 0.\ ....$	0559	0317	0299	0299	0289
$\mathrm{RMSE}(\hat{\beta}_4) = 0.\ ....$	0139	0121	0087	0088	0094
$\mathrm{RMSE}(\hat{\beta}_5) = 0.\ ....$	0270	0217	0168	0167	0174
$\mathrm{RMSE}(\hat{\sigma}_\alpha) = 0.\ ....$			0520	0504	0459
$\mathrm{RMSE}(\hat{\sigma}_\varepsilon) = 0.\ ....$			0086	-	-
$s(\hat{\beta}_1) = 0.\ ....$	0483	0284	0258	0257	0263
$s(\hat{\beta}_2) = 0.\ ....$	0320	0167	0164	0163	0157
$s(\hat{\beta}_3) = 0.\ ....$	0527	0300	0292	0291	0289
$s(\hat{\beta}_4) = 0.\ ....$	0136	0121	0086	0087	0094
$s(\hat{\beta}_5) = 0.\ ....$	0270	0216	0168	0167	0174
$s(\hat{\sigma}_\alpha) = 0.\ ....$			0182	0185	0193
$s(\hat{\sigma}_\varepsilon) = 0.\ ....$			0062	0036	0061
$\overline{\hat{\sigma}(\hat{\beta}_1)} = 0.\ ....$	0510	0300	0267	0266	0256
$\overline{\hat{\sigma}(\hat{\beta}_2)} = 0.\ ....$	0325	0201	0154	0154	0148
$\overline{\hat{\sigma}(\hat{\beta}_3)} = 0.\ ....$	0609	0371	0300	0298	0291
$\overline{\hat{\sigma}(\hat{\beta}_4)} = 0.\ ....$	0137	0114	0083	0083	0082
$\overline{\hat{\sigma}(\hat{\beta}_5)} = 0.\ ....$	0266	0210	0157	0157	0156
$\overline{\hat{\sigma}(\hat{\sigma}_\alpha)} = 0.\ ....$			0114	0115	0116
$\overline{\hat{\sigma}(\hat{\sigma}_\varepsilon)} = 0.\ ....$			0058	0037	0054
p-values H_0:					
$\mu_1 = \beta_1$ (0. ....)	0000	0000	0058	0081	-
$\mu_2 = \beta_2$ (0. ....)	0001	0001	3280	0956	0025
$\mu_3 = \beta_3$ (0. ....)	0000	0000	0028	0008	5051
$\mu_4 = \beta_4$ (0. ....)	0036	9482	1241	1684	7965
$\mu_5 = \beta_5$ (0. ....)	3694	2525	9321	9789	6391

Many of the results of tables 1 and 2 are repeated with the following exceptions. There is are clear difference in the behavior between the intercept and the parameters concerning block-constant covariates on the one hand and the parameters concerning spell-varying covariates on the other hand. The formers are estimated with a bias which is distinct in the GEE/BJ cases and moderate in the ML cases, the latters are estimated without bias throughout. The standard deviations of the parameters of spell-varying covariates in this case are not overestimated by the GEE/BJ(equ) method. In terms of rmse, the ML estimates are all more efficient than the GEE/BJ(equ) estimates.

6 Application to yogurt purchase behavior

The described methods are applied to data collected by the A.C. Nielsen purchase and TV panel Single Source. In 1994, 6000 panel households scanned the barcodes of their purchases of fast moving consumer goods using electronic scanners. In a subsample of 864 households also TV behavior is measured by so-called people meters by collecting data on who is watching what program at what time. This is an excellent basis for measuring the effects of TV advertising on sales of particular brands. The data have been analysed by Michels and Brüne (1996) using binary logistic regression models for the dependent variables $y_i = 1$ if a household buys brand no. i (given a purchase in the product category), $y_i = 0$ if the household buys one of the other brands, $i = 1, \ldots, 4$ (analyses of four brands have been reported). The independent variables in this logistic regression are (all of them categorized to dichotomous indicators)

- the age of the primary shopper
- the household net income
- "children in the household"
- exposure to the spots of brand no. i
- exposure to spots of competitors of brand no. i
- "household bought brand no. i in the preceding purchase act".

This method is a modified and improved version of the so-called STAS concept (Short Term Advertising Strength) of Jones (1995) where "exposure to the spots of brand no. i" is the only variable that explains the purchase probability of brand no. i. In practice, this is the variable of main interest also in the extended logistic analysis. The corresponding odds ratio is interpreted as success of advertising.

From a statistical point of view, there are several approaches that could be undertaken to form a more precise model for the very precise data.

i) It should be taken into account that the purchase acts of one household over the year are correlated responses. Marginal ML estimation or GEE techniques would be appropriate to the correlated binary response (Spieß, 1995, Spieß and Hamerle, 1995).

ii) It is somewhat unsatisfactory to estimate four binary models of four brands using the same data base. In reality, every purchase act is a multinomial response variable – the categories are the brands. Qualified methods are the multinomial logistic model, see e.g. Fahrmeir, Hamerle and Tutz (1996, p. 262-271), and the extension of GEE techniques on correlated multinomial responses, see e.g. Miller, Davis and Landis (1993) and Fieger, Heumann and Kastner (1996).

iii) Panel models for correlated responses usually assume equi-distant observation of the variables in time and do not take into account that different periods of time pass until one or the other brand is bought. We think that there is an interesting difference between two households X and Y, both of them only watching TV spots of brand no. i, X buys brand no. i (and only i) once a month whereas Y buys yogurt twice a week changing between i and j. So we are in the field of failure time analysis methods.

For clarification, in this paper we are far from solving all the listed problems and from estimating a model of correlated failure times in multi-categorical states including an effect of advertising. But we try some steps in this direction by applying the model of section 2 and the estimation methods of sections 3 and 4 on the yogurt purchase data. As it is a model for recurrent events of the same kind, it does not allow for the analysis of multiple states i.e. the simultaneous treatment of all interesting brands. So we have to estimate as many models as we have brands in the dataset – comparable to the STAS analysis. The other, more serious lack is that the estimation methods presented in this paper do not take into account covariates which vary during the spells – they may be time-varying only in the sense of varying from spell to spell. The problem is that watching TV spots produces a covariate which varies decisively during the spell between two purchase acts. So the analyses in this paper only include the covariates age, income, and children in the household, do not give evidence of the effect of advertising, but can be seen as some preparatory work. Further research is already pursued to handle models of recurrent events and multi-state-multiple-spell models with unobserved heterogeneity and time-varying covariates.

The analyses are presented in two steps. Firstly, we do not consider brands at all. Instead, we try to explain the sequence of yogurt purchase acts by a household on the whole by the covariates

A1 $= 1$ if the primary shopper is younger than 40, else $= 0$,

E1 $= 1$ if the household net income is less than 3000 DM, else $= 0$,

KI $= 1$ if there are children in the household, else $= 0$.

36 of the 864 TV households did not buy yogurt at all in the reported year, the other 828 households had 20071 yogurt purchase acts. As we do not deal with left-censored spells, we drop for each household the time from January, 1^{st}, 1994, until the date of the first yogurt purchase. The spells from the last purchase act in the year until January, 1^{st}, 1995, are included as right-censored spells. The number of days between every two yogurt purchases of a household are the lengths of the not-censored spells. So each purchase enters into the analysis as the beginning of a spell. Additionally we include the 36 households that did not buy yogurt at all as censored spells with an observed

Table 4: *Sizes of datasets and censor rates*

brand	number of households (a)	purchase acts (this brand) (b)	households without purchases of yogurt/ this brand (c)	number of spells d=b+c	censor rate e=a/d (%)
all brands	864	20071	36	20107	4
A	864	2132	488	2620	33
B	864	1596	483	2079	42

length of $C_n = c_{n1} = 366$ days. That yields a total of 20107 spells, 864 of them (4%) are censored (see table 4).

The other kind of analysis is done brand by brand. The forming of spells is the same as described above but only purchases of brand no. i are relevant. We report the results of two brands, referred to as brand A and brand B (corresponding to Michels and Brüne, 1996). For numbers of purchases, spells, and censor rates see table 4.

In table 5, there are given the results of the first step (all brands): the estimations of the regression parameters and their standard errors, in the case of ML estimation the standard deviations of the heterogeneity and the residuals with their standard errors, and the maximum value of the log-likelihood, in the case of GEE/BJ estimation the moment estimation of the variance and the covariance, and p-values testing the hypotheses $\beta_p = 0$ for all included regression parameters. CO means the intercept.

In the analysis of all yogurt purchases the estimated parameters are of roughly the same values for all the five methods, except the higher estimated standard errors in the case of GEE/BJ estimation. From the simulation study we know that this is at least partly due to a general over-estimation of the standard errors which is characteristic of this method. Further, the equicorrelation working matrix in GEE/BJ estimation results in absolutely higher estimated effects in comparison to the independence structure. For this, there is no corresponding observation in the simulations. But the equicorrelation assumption seems to be advisable as the heterogeneity deviation parameter in all ML estimations is significantly different from zero. On the other hand, the heterogeneity is moderate with approximately the same deviance as the residual error. The effects of age and net income are significant on the 5% level in all cases (except the independence working matrix). The children indicator is significant using ML and specifying the lognormal or the Weibull model, but is not significant in the log-logistic model or using GEE/BJ estimation. This observing is consistent as if we do not know anything about the distributions, the nonparametric method should not give a more exact statement than any of the parametric methods. On the other hand, if there were any reasons for assuming the lognormal or the Weibull model in this case, we could claim a significant effect of having children despite the GEE/BJ result.

Because unobserved heterogeneity is proved to be present and the simulations showed that the GEE/BJ method with independence working matrix does not work well in this case, we neglect the GEE/BJ(ind) column for the substantial interpretation of the regression parameters. The GEE/BJ(equ) results may be interpreted as follows. The expected interval length between two yogurt purchases of a household with all covariates equalling zero – that means the primary shopper older than 40 years, having

Table 5: *Estimation results – all brands*

		GEE/BJ		ML		
		ind	equ	(N)	(L)	(W)
$\hat{\beta}$	CO	2.03	2.51	2.56	2.48	3.17
$[\hat{\sigma}(\hat{\beta})]$		[0.0430]	[0.0451]	[0.0356]	[0.0341]	[0.0435]
	A1	0.123	0.143	0.221	0.105	0.204
		[0.0599]	[0.0701]	[0.0454]	[0.0476]	[0.0455]
	E1	0.0787	0.180	0.224	0.193	0.117
		[0.0810]	[0.0820]	[0.0675]	[0.0468]	[0.0471]
	KI	−0.0129	−0.0822	−0.163	−0.0303	−0.179
		[0.0593]	[0.0660]	[0.0421]	[0.0422]	[0.0383]
$\hat{\sigma}_\alpha$		.	.	0.819	0.733	0.985
$[\hat{\sigma}(\hat{\sigma}_\alpha)]$		.	.	[0.0208]	[0.0192]	[0.0199]
$\hat{\sigma}_\varepsilon$		.	.	0.893	0.499	0.853
$[\hat{\sigma}(\hat{\sigma}_\varepsilon)]$		.	.	[0.00471]	[0.00308]	[0.00460]
$\hat{v}$		1.15	1.15	.	.	.
$\hat{c}$		.	0.398	.	.	.
p-values	A1	0.040	0.041	0.000	0.027	0.000
	E1	0.331	0.028	0.001	0.000	0.013
	KI	0.827	0.213	0.000	0.474	0.000
l		.	.	−64863	−64678	−66271

no children and a net income of more than 3000 DM – is about $e^{2.51} = 12.3$ days. Positive regression parameters cause lengthened purchase intervals, that means rarer purchase acts. Here, being younger or earning less have a multiplicative effect of about $e^{0.143} = 1.15$ or $e^{0.180} = 1.20$, respectively, on the interval lengths. On the other hand, the interval-shortening effect of having children is estimated as $e^{-0.0822} = 0.92$. The three ML columns show rather similar regression parameter estimations. Note that the intercept estimated by the Weibull model is not directly comparable to the other methods because of the shifted expectation of the distribution. The children effect is $e^{-0.0303} = 0.97$ in the log-logistic model but $e^{-0.179} = 0.84$ in the Weibull model.

The data situation changes of course when we analyze particular brands. More than half the households do not buy brand A or B all year long; the households that do buy brand A or B at least once the year in average contribute only about four to five purchase acts, and censor rates are about one third (as in the simulation studies) or higher.

There is much more unobserved heterogeneity than in the analysis of all yogurt purchases (see $\hat{\sigma}_\alpha$ in tables 6 - 8). Almost every household buys yogurt. Thus the fact of being a yogurt eater does not discriminate the sample very much. Having a high affinity to a certain brand is a much stronger classification of households reflecting their likes and dislikes. This interpretation would point to a dichotomous variable which is not observed, and this was the incentive to the kind of simulation studies reported in tables 2 and 3. In spite of a possibly misspecified frailty distribution the results of these simulations encourage us not to reject completely the ML methods in the brand-specific analysis but only to interpret the results with some caution. The regression parameters

Table 6: *Estimation results – brand A*

		GEE/BJ		ML		
		ind	equ	(N)	(L)	(W)
$\hat{\beta}$	CO	3.37	4.95	6.54	6.49	6.91
$[\hat{\sigma}(\hat{\beta})]$		[0.353]	[0.260]	[0.157]	[0.142]	[0.148]
	A1	−0.299	−0.0494	0.253	0.310	0.302
		[0.515]	[0.386]	[0.0916]	[0.0816]	[0.0885]
	E1	0.854	0.364	2.09	2.13	2.15
		[0.505]	[0.450]	[0.132]	[0.128]	[0.121]
	KI	0.810	0.0570	0.426	0.411	0.405
		[0.485]	[0.374]	[0.0889]	[0.0789]	[0.0900]
$\hat{\sigma}_\alpha$		.	.	3.18	3.18	3.21
$[\hat{\sigma}(\hat{\sigma}_\alpha)]$		.	.	[0.102]	[0.0948]	[0.0962]
$\hat{\sigma}_\varepsilon$		.	.	1.14	0.634	1.02
$[\hat{\sigma}(\hat{\sigma}_\varepsilon)]$		.	.	[0.0208]	[0.0133]	[0.0191]
$\hat{v}$		4.30	3.59	.	.	.
$\hat{c}$		.	1.67	.	.	.
p-values	A1	0.561	0.898	0.006	0.000	0.001
	E1	0.091	0.419	0.000	0.000	0.000
	KI	0.095	0.879	0.000	0.000	0.000
l		.	.	−7502	−7473	−7583

are perhaps slightly over-estimated. On the other hand, because of the higher censor rates, the GEE/BJ method is expected to result in over-estimated standard errors and probably in not-recognized effects.

In the analysis of brand A (table 6) the regression parameter estimates of GEE/BJ estimation are absolutely lower in the equicorrelation than in the independence case, and all the GEE/BJ estimated parameters are not significant on the 5% level. On the other hand, the three ML specifications compared to each other produce nearly the same estimates of parameters and their standard errors, and the effects are all highly significant. The ML results suggest that households with older primary shoppers and particularly with higher net income buy brand A more often than others. As we already know, they generally buy more yogurt. But brand A is the favourite of childless households – having children lengthens the brand A purchase intervals by a factor of about $e^{0.41} \approx 1.5$. However, this result is not corroborated by the nonparametric method.

The data situation (number of spells and censor rate) is similar in the analysis of brand B, but there is another constellation of significant parameters (table 7). Age and income are not significant in GEE/BJ estimation as well as in ML estimation with lognormal model. It is the same case as the not clear children effect in the analysis of all yogurt purchases: The nonparametric method does not give a significant result when any of the parametric models does not. The children effect of brand B, on the other hand, is significant, independent of the method. Having children shortens the brand B purchase intervals to a third or a fourth of the time.

Table 7: *Estimation results – brand B*

		GEE/BJ		ML		
		ind	equ	(N)	(L)	(W)
$\hat{\beta}$	CO	5.11	6.20	7.82	7.88	8.27
$[\hat{\sigma}(\hat{\beta})]$		[0.566]	[0.356]	[0.259]	[0.203]	[0.212]
	A1	−0.00652	−0.0525	−0.00175	−0.378	−0.216
		[0.423]	[0.354]	[0.247]	[0.136]	[0.129]
	E1	0.141	0.205	0.559	0.384	0.464
		[0.585]	[0.515]	[0.379]	[0.154]	[0.210]
	KI	−1.42	−1.37	−1.39	−1.07	−1.07
		[0.571]	[0.448]	[0.226]	[0.145]	[0.197]
$\hat{\sigma}_\alpha$		.	.	2.71	2.70	2.37
$[\hat{\sigma}(\hat{\sigma}_\alpha)]$		.	.	[0.101]	[0.0945]	[0.0778]
$\hat{\sigma}_\varepsilon$		.	.	1.24	0.696	1.04
$[\hat{\sigma}(\hat{\sigma}_\varepsilon)]$		.	.	[0.0286]	[0.0180]	[0.0248]
$\hat{v}$		3.81	4.11	.	.	.
$\hat{c}$		.	1.67	.	.	.
p-values	A1	0.988	0.882	0.994	0.005	0.095
	E1	0.810	0.691	0.141	0.013	0.027
	KI	0.013	0.002	0.000	0.000	0.000
l		.	.	−5673	−5657	−5703

There are two reasons that encouraged us to do a re-analysis for brand A including interaction effects (which was not a topic of the extended STAS analysis). The first one is the common argument for interaction terms: The observed effects probably do not act independently. The other one is the hope that the interactions may explain a further part of the variability in the data, and unobserved heterogeneity will decrease.

The latter is not the case as the ML results in table 8 show. The estimated heterogeneity and residual deviations and the likelihood stay on the same level. The GEE/BJ procedure with equicorrelation matrix has not come to an end after 90 iterations, and the independence assumption does not give much evidence.

For interpretation, we keep to the ML results as they are rather uniform. Interactions with the children effect are not significant at all. The spell-lengthening effect of lower age is turned around in the case of lower income by the interaction A1*E1. So older childless households with higher income prefer brand A; it is bought least by older households with children and lower income.

7 Discussion

The nonparametric GEE/Buckley-James method has been compared to the fully parametric marginal maximum likelihood estimation by simulation studies in section 5. The results of these studies are valid for the specified conditions, especially the large sample size, the medium censor rate, the assumed distributions, the moderate heterogeneity and the kinds of covariates. These specifications have been made with the application

Table 8: *Estimation results – brand A – including interaction effects*

		GEE/BJ ind	GEE/BJ equ	ML (N)	ML (L)	ML (W)
$\hat{\beta}$	CO	3.31	.	6.55	6.53	6.93
$[\hat{\sigma}(\hat{\beta})]$		[0.379]	.	[0.165]	[0.147]	[0.157]
	A1	−0.321	.	0.356	0.382	0.323
		[0.872]	.	[0.120]	[0.104]	[0.110]
	E1	1.31	.	2.26	2.36	2.41
		[0.841]	.	[0.167]	[0.196]	[0.161]
	KI	1.28	.	0.546	0.496	0.465
		[0.651]	.	[0.135]	[0.115]	[0.143]
	A1*E1	1.91	.	−1.22	−1.68	−1.42
		[1.05]	.	[0.343]	[0.299]	[0.307]
	A1*KI	−0.422	.	−0.197	−0.124	0.160
		[1.07]	.	[0.190]	[0.169]	[0.188]
	E1*KI	−2.14	.	−0.270	−0.327	−0.439
		[1.07]	.	[0.263]	[0.267]	[0.268]
$\hat{\sigma}_\alpha$		.	.	3.21	3.23	3.24
$[\hat{\sigma}(\hat{\sigma}_\alpha)]$		.	.	[0.105]	[0.0970]	[0.102]
$\hat{\sigma}_\varepsilon$		.	.	1.14	0.633	1.02
$[\hat{\sigma}(\hat{\sigma}_\varepsilon)]$		.	.	[0.0208]	[0.0133]	[0.0191]
$\hat{v}$		4.37	.	.	.	.
$\hat{c}$		.	.	.	.	.
p-values	A1	0.712	.	0.003	0.000	0.003
	E1	0.120	.	0.000	0.000	0.000
	KI	0.050	.	0.000	0.000	0.001
	A1*E1	0.069	.	0.000	0.000	0.000
	A1*KI	0.694	.	0.299	0.461	0.397
	E1*KI	0.045	.	0.304	0.221	0.102
l		.	.	−7499	−7470	−7578

of section 6 in mind. Further research will be done either to generalize the results or to differentiate them where necessary.

The experiments showed that the application of the generalized estimating equations approach in combination with the method of Buckley and James is a feasible way for handling censored data of multivariate failure times. If any correlation between the failure times within the blocks can be assumed, the equicorrelation working matrix has definitely to be prefered in comparison to the simple independence matrix. But because of the actually more complex correlation structure between censored spells on the one hand and uncensored spells on the other hand, also the GEE/BJ(equ) estimator leads to uncontrolled – although moderate – biases of the regression parameter estimations and in most cases tends to an overestimation of the estimation variance.

The main intention of the paper was to show that we need not necessarily mistrust the fully parametric ML estimator when in an application we do not know the real frailty distribution or even have reason to assume that we have misspecified it. Despite the extreme misspecifications which are reported in tables 2 and 3 the ML regression parameter estimates mostly are less biased than the GEE/BJ(equ) estimates when the hazard is specified correctly or when both the hazard and the frailty are specified incorrectly. On the other hand there is a distinct bias of the ML(W) estimator in table 1 in the case of correctly specified frailty but incorrectly specified hazard.

In no case we found a strong or systematic difference between the standard deviations of the ML estimates and the means of their estimated standard deviations. In terms of root mean squared errors, the ML estimators mostly are at least as efficient as the GEE/BJ(equ) estimator.

Further on, the simulation studies let us suppose that it would often be more important to choose the correct hazard distribution than the correct frailty (see also Haider and Davies, 1996). As in the application we really do not know anything about the hazard distribution a priori, we encounter the problem by specifying three hazard distributions which are typical for failure time analysis and by comparing the outcomes. In tables 5 and 7 we have the nice results that if there are significant parameter estimations in all the three ML cases, also the GEE/BJ(equ) estimation is significant. So younger people and households with lower income definitely more often buy yogurt than others, and brand B is definitely prefered by households with children. If any of the ML estimates is not significant, the GEE/BJ(equ) estimate is not significant, too.

On the other hand, in table 6, all the ML estimates are significant but the GEE/BJ(equ) estimates are not. In the case of the income variable the reason clearly is the overestimated standard error of the GEE/BJ method. The GEE/BJ parameter estimate 0.364 combined with the trusty ML standard errors of about 0.13 would also result in significance. The estimation of the age and the children effect on brand A remains unclear. Perhaps the ML estimation is in these cases sensitive to the frailty misspecification and we should believe the GEE/BJ estimator – or the GEE/BJ estimator is biased and we could believe the ML estimators.

References

Andersen, Per Kragh, **Borgan,** Ørnulf, **Gill,** Richard D. **and Keiding,** Niels (1993). *Statistical Models Based on Counting Processes.* New York: Springer.

Bock, R. Darrell **and Lieberman,** Marcus (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35 (2),* 179-197.

Buckley, Jonathan **and James,** Ian (1979). Linear regression with censored data. *Biometrika, 66,* 429-436.

Butler, J. S. **and Moffit,** Robert (1982). Notes and comments: A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica, 50 (3),* 761-764.

Clayton, David G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics, 47,* 467-485.

Crouchley, Robert **and Pickles,** Andrew R. (1995). Multivariate survival models for repeated and correlated events. *Journal of Statistical Planning and Inference, 47,* 95-110.

Currie, Iain D. (1996). A note on Buckley-James estimators for censored data. *Biometrika, 83,* 912-915.

Davies, Richard B. (1993). Nonparametric control for residual heterogeneity in modelling recurrent behaviour. *Computational Statistics & Data Analysis, 16,* 143-160.

Dennis, J. E. jr. **and Schnabel,** R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Englewood Cliffs, New Jersey: Prentice-Hall.

Fahrmeir, Ludwig, **Hamerle,** Alfred **and Tutz,** Gerhard (eds.) (1996). *Multivariate statistische Verfahren.* (2^{nd} ed.) Berlin: de Gruyter.

Fieger, Andreas, **Heumann,** Christian **and Kastner,** Christian (1996). *MAREG and WinMAREG.* Discussion Paper 45, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München.

Follmann, Dean A. **and Goldberg,** Matthew S. (1988). Distinguishing heterogeneity from decreasing hazard rates. *Technometrics, 30,* 389-396.

Haider, Najeeb **and Davies,** Richard B. (1996). Modelling duration data: Sensitivity to model specification. In: Antonio Forcina, Giovanni M. Marchetti, Reinhold Hatzinger and Gianfranco Galmacci (eds.): *Statistical Modelling.* Proceedings of the 11^{th} International workshop on Statistical Modelling, Orvieto, pp. 422-425.

Hornsteiner, Ulrich **and Hamerle,** Alfred (1996). *A combined GEE/Buckley-James method for estimating an Accelerated Failure Time Model of multivariate failure times.* Discussion paper 47, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München.

Jones, John Philip (1995). *When Ads Work. New Proof that Advertising Triggers Sales.* New York: Lexington Books.

Kalbfleisch, John D. **and Prentice,** Ross L. (1980). *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons.

Kaplan, E.L. **and Meier,** Paul (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53,* 457-481.

Kiefer, Nicholas M. (1983). An integral occuring in duration models with heterogeneity. *Economics Letters, 11,* 251-256.

Lancaster, Tony **and Nickell,** Stephen (1980). The analysis of re-employment probabilities for the unemployed. *Journal of the Royal Statistical Society A, 143,* 141-165.

Lawless, Jerald F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: John Wiley & Sons.

Liang, Kung-Yee **and Zeger,** Scott L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13-22.

Meyer, Bruce D. (1990). Unemployment insurance and unemployment spells. *Econometrica, 58,* 757-782.

Michels, Paul **and Brüne,** Klaus (1996). Zuverlässiger Index. *Werben und Verkaufen, 3/96,* 107.

Miller, Michael E., **Davis,** Charles S. **and Landis,** J. Richard (1993). The analysis of longitudinal polytomous data: Generalized estimating equations and connections with weighted least squares. *Biometrics, 49,* 1033-1044.

Miller, Rupert G. jr. (1981). *Survival Analysis.* New York: John Wiley & Sons.

Nielsen, Gert G., **Gill,** Richard D., **Andersen,** Per Kragh **and Sørensen,** Thorkild I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics, 19,* 25-43.

Pickles, Andrew R. **and Crouchley,** Robert (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine, 14,* 1447-1461.

dos Santos, Dirley M., **Davies,** Richard B. **and Francis,** Brian (1995). Nonparametric hazard versus nonparametric frailty distribution in modelling recurrence of breast cancer. *Journal of Statistical Planning and Inference, 47,* 111-127.

SAS Institute Inc. (1989). *SAS/IML Software: Usage and Reference, version 6.* Cary, NC: SAS Institute.

SAS Institute Inc. (1990). *SAS Language: Reference, version 6.* Cary, NC: SAS Institute.

Scheike, Thomas H. **and Jensen,** Tina Kold (1997). A discrete survival model with random effects: An application to time to pregnancy. To appear in Biometrics.

Schneider, Hilmar (1997). Using the MAXLIK library of GAUSS for the estimation of duration models with time-varying covariates and unobserved heterogeneity. Manuscript of a lecture at SOFTSTAT, Heidelberg, March, 1997.

Spieß, Martin (1995). *Parameterschätzung in Regressionsmodellen mit abhängigen diskreten endogenen Variablen.* Konstanz: Hartung-Gorre.

Spieß, Martin **and Hamerle,** Alfred (1995). *A comparison of different methods for the estimation of regression models with correlated binary responses.* Regensburger Beiträge zur Statistik und Ökonometrie 37.

Spieß, Martin, **Nagl,** Willi **and Hamerle,** Alfred (1997). *Probit models: Consistent ML regression parameter estimation despite misspecification of the correlation structure.* Discussion paper 67, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München.

Stroud, A. H. **and Secrest,** Don (1966). *Gaussian Quadrature Formulas.* Englewood Cliffs, New Jersey: Prentice-Hall.