Continuous Time Survival in Latent Variable Models

Tihomir Asparouhov[1], Katherine Masyn[2], Bengt Muthen[3]
Muthen & Muthen[1]
University of California, Davis[2]
University of California, Los Angeles[3]

Abstract

We describe a general multivariate, multilevel framework for continuous time survival analysis that includes joint modeling of survival time variables and continuous and categorical observed and latent variables. The proposed framework is implemented in the Mplus software package. The survival time variables are modeled with nonparametric or parametric proportional hazard distributions and include right censoring. The proposed modeling framework includes finite mixtures of Cox regression models with and without class-specific baseline hazards, multilevel Cox regression models, and multilevel frailty models. We illustrate the framework with several simulation studies. Comparison is made with discrete time survival models. We also investigate the effect of ties on the proposed estimation method. Simulation studies are conducted to compare the methods implemented in Mplus with those implemented in SAS.

Keywords: Multilevel Survival Analysis, Latent Variables.

1 Overview

In this article we describe the two-level continuous time survival model implemented in Mplus. The model includes mixture survival models, survival models with random effects (also known as frailty models), time varying covariate models and non-proportional hazard models. Introduction to continuous time survival modeling can be found in Singer & Willett (2003), Hougaard (2000) or Klein & Moeschberger (1997). The model described here is a direct extension of the models described in Larsen (2004, 2005). In Section 2 we describe the proportional hazard model which serves as the basis for modeling time-to-event variables. In Section 3 we describe the general multilevel latent variable mixture model that allows the joint modeling of time-to-event data and other observed and latent variables. We also illustrate the framework with some simple simulation studies. In Section 4 we discuss some aspects of mixture survival models. In Section 5 we compare continuous time and discrete time survival models. In Section 6 we investigate the effect of ties on the proposed estimation method.

2 The Proportional Hazard Model

Let the variable T_0 be a time-to-event variable such as time to death for example. Let I be the time when the individual leaves the target cohort due to death or other types of censoring such as lost to follow up. The survival variable T and the censoring indicator δ are defined by

$$T = \min\{T_0, I\} \tag{1}$$

$$\delta = \begin{cases} 1 & \text{if } T_0 > I \\ 0 & \text{if } T_0 \leq I \end{cases}. \tag{2}$$

Let X be an observed vector of predictor variables. The proportional hazard (PH) model specifies that the hazard function is proportional to the baseline hazard function, i.e.,

$$h(t) = \lambda(t)Exp(\beta X) \tag{3}$$

where $h(t)$ is the hazard function and $\lambda(t)$ is the baseline hazard function at time t.

Two proportional hazard models are described in this article. The first model assumes a completely nonparametric shape for the baseline hazard function $\lambda(t)$. This model is known as the Cox regression model. The second model is based on a parametric model for the baseline hazard function $\lambda(t)$. This model is known as the parametric PH model. The parametric model for the baseline hazard function we describe here is a step function with arbitrary number of steps, however through additional parameter constraints this parametric model can serve as an approximation to any other parametric model, including models such as Exponential, Weibull and Gompertz models. This approximation is based on the fact that any continuous function can be closely approximated by a step function. First we describe the parametric PH model.

2.1 Parametric PH model

For this model the baseline hazard is estimated as a step function. The step function is constant within each of the defined intervals. If all possible values for T are positive, we split the interval $[0, \infty)$ into Q non-overlapping intervals $l_1 = [0, t_1), l_1 = [t_1, t_2), ..., l_Q = [t_{Q-1}, \infty)$, where $t_1, t_2, ..., t_{Q-1}$ are the points of change for the baseline hazard function. These points are fixed constants that are specified prior to the model estimation. Later we discuss strategies for selecting these points in practical applications. The baseline hazard function is now defined by

$$\lambda(t) = \begin{cases} h_1 & \text{if } 0 < t \leq t_1 \\ h_2 & \text{if } t_1 < t \leq t_2 \\ ... & \\ h_Q & \text{if } t_{Q-1} < t < \infty \end{cases} \qquad (4)$$

where h_1,...,h_Q are parameters to be estimated together with the β parameters in equation (3).

There are two ways to approach the estimation of the baseline parameters h_i. The first approach is to treat these parameters as regular parameters. Standard errors can then be computed and, using the delta method, standard errors for the survival rates can also be derived. This approach also allows us to constrain the baseline parameters during the optimization to follow certain parametric shapes or to mandate baseline hazard function to be class invariant in finite mixture models. The approach however should only be used for baseline hazards with relatively few steps, because large number of baseline parameters will be computationally very demanding.

The second estimation approach treats the baseline parameters h_i as nuisance parameters. The profile likelihood is formed by explicitly maximizing the full likelihood over the baseline parameters. The profile likelihood is then treated as regular maximum likelihood (see Murphy and van der Vaart, 2000). Standard errors are not computed for the baseline parameter; however, this approach is computationally feasible even with large number of baseline parameters.

2.2 Cox Regression Model

This model assumes a non-parametric baseline hazard function. One way for estimating such a model is to estimate a PH parametric model with a step function that is constant between every two consecutive event times thus estimating the most detailed hazard function possible. If all event times, including censored observations are $t_1 < t_2 < .... < t_n$ then we estimate the baseline step function as in equation (4) with $Q = n + 1$ parameters. This estimation approach was first developed in Breslow (1974) and is now referred to as the Breslow likelihood approach or the profile likelihood approach. When estimating the Cox regression model the parameters h_i are estimated as nuisance (unrestricted) parameters. Equal event times are treated as one event time.

An alternative estimation approach known as the partial likelihood approach was first developed by Cox (1972). This estimation method eliminates the estimation of the baseline hazard function completely and maximizes only a part of the complete log-likelihood that contains the regression coefficients β. It has been shown however that the partial likelihood estimation and the profile likelihood estimation are equivalent, see Johansen (1983) and Clayton (1988). In the presence of ties for the T variable there are a number of different variations of the partial likelihood approach, see Section 6 below.

2.3 The Likelihood Function

The likelihood function described in this section applies to both the Cox regression model and the parametric PH model. Given the baseline hazard function in (4), the cumulative baseline hazard function $H_0(t)$ at time t represents the total hazard an individual is exposed to up to time t. If $t_k < t < t_{k+1}$ then

$$H_0(t) = \int_0^t \lambda(x)dx = \sum_{i=1}^{k-1} h_i(t_i - t_{i-1}) + h_k(t - t_{k-1}) \quad (5)$$

The survival function is the probability that the survival variable T is greater than t; that is,

$$S(t) = P(T > t) = Exp(-Exp(\beta X)H_0(t)). \qquad (6)$$

The likelihood function of the survival variable T is

$$L(T) = (\lambda(T)Exp(\beta X))^{(1-\delta)}S(T) \qquad (7)$$

where δ is the censoring indicator.

2.4 Weibull PH Model Specification

In this section we will illustrate how continuous baseline hazard function models can be approximated by the baseline hazard step function model described above. Consider for example the Weibull model which assumes that the baseline hazard function is

$$\lambda(t) = \alpha s(\alpha t)^{s-1}, \qquad (8)$$

for some parameters α and s (see Bradburn et al., 2003). The precision of the approximation depends on the number of intervals Q used in the baseline step function. The more intervals are used the better the approximation. In practical applications $Q = 50$ or $Q = 100$ will be sufficient. Denote the largest event time T by M. Let $h = M/Q$. Consider the parametric PH model with baseline hazard step function based on equal size intervals of length h. We estimate the parametric PH model with the constraint maximum likelihood where the hazard parameters h_i are constraint by

$$h_i = \alpha s(\alpha h(i - 0.5))^{s-1}. \qquad (9)$$

The value for t has been substituted with the midpoint for each of the time intervals. The LRT test can be used to test the model constraint equations (9), i.e., to test the assumption of Weibull baseline hazard.

3 The General Latent Variable Model

Let T_{rij} be the $r-$th observed time-to-event variable for individual i in cluster j. Let y_{pij} be the $p-$th observed dependent variable for individual i in cluster j. We only consider two types of variables, categorical and normally distributed continuous variables. However it is possible to incorporate other types of distributions and link function in this model as in the generalized linear models of

McCullagh and Nelder (1989). Suppose that C_{ij} is a latent categorical variable for individual i in cluster j which takes values $1, ..., L$.

To construct the structural model for the dependent variables y_{pij} we proceed as in Muthen (1984). If y_{pij} is an ordered categorical variable, we define an underlying normally distributed latent variable y^*_{pij} such that for a set of threshold parameters τ_{ck}

$$[y_{pij} = k | C_{ij} = c] \Leftrightarrow \tau_{ck} < y^*_{pij} < \tau_{ck+1}. \qquad (10)$$

A linear regression for y^*_{pij} is thus equivalent to a Probit regression for y_{pij}. Alternatively, y^*_{pij} can have a logistic distribution. Linear regression for y^*_{pij} will then translate to a logistic regression for y_{pij}. For continuous variables we define $y^*_{pij} = y_{pij}$.

Let y^*_{ij} be the vector of all dependent variables and let x_{ij} be a vector of all covariates. The structural part of the model is defined by

$$[y^*_{ij} | C_{ij} = c] = \nu_{cj} + \Lambda_{cj}\eta_{ij} + \varepsilon_{ij} \qquad (11)$$

$$[\eta_{ij} | C_{ij} = c] = \mu_{cj} + B_{cj}\eta_{ij} + \Gamma_{cj}x_{ij} + \xi_{ij} \qquad (12)$$

$$P(C_{ij} = c) = \frac{\exp(\alpha_{cj} + \beta_{cj}x_{ij})}{\sum_c \exp(\alpha_{cj} + \beta_{cj}x_{ij})}. \qquad (13)$$

where η_{ij} are normally distributed latent variables, ε_{ij} and ξ_{ij} are zero mean normally distributed residuals. The model for the time-to-event variables T_{rij} is described by the following model for the hazard functions

$$[h_{rij}(t) | C_{ij} = c] = \lambda_{rc}(t) Exp(\iota_{rcj} + \gamma_{rcj}x_{ij} + \kappa_{rcj}\eta_{ij}). \qquad (14)$$

The likelihood for T_{rij} is computed as in Section 2.3. Some parameters in the above model have to be restricted for identification purpose. For example, in equation (14) the intercept parameter ι_{rcj} can be identified only when the baseline hazard function is class invariant. Then ι_{rcj} can be identified in all but one class. If the baseline hazard function is class specific then ι_{rcj} is not identified in any of the classes and it is fixed to 0. When ι_{rcj} is a cluster random effect then its mean is fixed to 0. In addition, for categorical variables y_{pij}, the variance of ε_{pij} is not identified and is typically fixed to 1. Also in the multinomial logit regression (13) we have $\alpha_{Lj} = \beta_{Lj} = 0$.

The multilevel part of the model is introduced as follows. Each of the intercepts, slopes or loading parameters in equations (11-14) can be either a fixed coefficient or a cluster random effect, i.e., a normally distributed cluster specific coefficient. Let η_j be a vector of all such random effects and let x_j be a vector of cluster level covariates. The between level model is then described by the following equation

$$\eta_j = \mu + B\eta_j + \Gamma x_j + \xi_j. \qquad (15)$$

where ξ_j is a normally distributed residual. The above five equations comprise the definition of the basic multilevel survival mixture model.

One example of the above framework is illustrated in Larsen (2005) where the predictor in the Cox regression is a latent factor measured by several binary indicators. In the next three sections we illustrate this framework with several examples.

3.1 Frailty Models

Frailty models are models that introduce association between two or more time-to-event variables or between one time-to-event variable and other types of dependent variables. For example in cancer studies the time from remission to relapse T_1 can be correlated with the time from relapse to death T_2. In modeling the mortality rates of married couples association between the two survival variables can be caused by some shared values that are not explicitly known and included in the model. Clayton (1988) describes one possibility for modeling the association between two survival variables. Let the survival times for individual i be T_{1i} and T_{2i} and the corresponding hazard functions be $h_{1i}(t)$ and $h_{2i}(t)$. The model in Clayton (1988) is described by

$$h_{1i}(t) = \xi_i\lambda_1(t) \qquad (16)$$

$$h_{2i}(t) = \xi_i\lambda_2(t) \qquad (17)$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are non-parametric functions and ξ_i are independent gamma variables with mean 1 and variance γ. This model has only one parameter, namely the γ parameter. The larger this parameter is the stringer the association between the two survival times. If $\gamma = 0$ the two survival times are independent.

The latent variable framework described in the previous section can also model the association between two survival variables. Let η_i be a normally distributed random variables with mean 0 and variance σ. The Cox regression of T_1 and T_2 on η_i with fixed slope 1 gives us the following expressions for the hazard functions

$$h_{1i}(t) = Exp(\eta_i)\lambda_1(t) \qquad (18)$$

$$h_{2i}(t) = Exp(\eta_i)\lambda_2(t). \qquad (19)$$

The difference between this model and the Clayton model (16-17) is only in the prior distribution for the coefficient of proportionality. In model (18-19) the coefficient is distributed as an exponential of normal while in (16-17) it is gamma distributed. This difference in the priors will typically have a marginal effect on the estimation of the association between the two survival variables. For the leukemia cancer data presented in Clayton (1988), the $\hat{\gamma}$ parameter in model (16-17) is 0, while the $\hat{\sigma}$ parameter in model (18-19) is 0.00004, which is not significant with a p-value of 0.45. Therefore both models lead to the same conclusion: there is no association between the two survival variables for this leukemia cancer data.

To evaluate the performance of the proposed estimation method for this frailty model we conduct a simple

Table 1: Estimates for the Variance $\sigma = 0.3$ in the Frailty Model Using Parametric and Non-parametric Approach.

n	100	500	1000
Non-Parametric Bias	-0.07	-0.03	-0.01
Parametric Bias	-0.01	-0.01	-0.01
Non-Parametric MSE	0.032	0.007	0.004
Parametric MSE	0.022	0.005	0.003
Non-Parametric Coverage	0.76	0.91	0.91
Parametric Coverage	0.88	0.96	0.90

simulation study. We generate data according to model (18-19) using the following baseline hazard functions

$$\lambda_1(t) = \begin{cases} 0.1 & \text{if } t < 5 \\ 0.2 & \text{if } 5 \le t < 10 \\ 0.5 & \text{if } 10 \le t < 15 \\ 15 & \text{if } 15 \le t \end{cases} \tag{20}$$

and $\lambda_2(t) = \lambda_1(t) + 0.2$. The parameter σ is chosen to be 0.3. We generate 100 samples of size $n = 100, 500$ and 1000. We analyze the data with the non-parametric approach as well as a parametric approach based on the assumption that the baseline hazard function is constant over the intervals [0,5), [5,10), [10,15) and $[15, \infty)$. Table 1 shows the bias and the mean squared error of the parameter estimates as well as the confidence interval coverage probability, i.e., the probability that the 95% confidence interval contains the true value. It is clear from these results that parametric method outperforms the non-parametric on all three criteria for small sample size. For large sample size the difference between the methods is negligible and both methods perform well.

This example emphasizes the advantages of the parametric approach. We see substantial efficiency gains even in simple examples. Therefore it is important to develop tools for constructing appropriate parametric models and for testing the parametric models against the unrestricted non-parametric model. Graphical methods can be used to determine the shape of the hazard function. For example, estimating a model with a stepwise hazard can indicate a particular shape. Figure 1 shows the estimate of the baseline hazard function $\lambda_1(t)$ for a sample with 1000 observations based on a model assuming constant hazard over consecutive intervals of length 1. In addition the Hausman (1978) test of misspecification can be used to formally test the parametric model against the non-parametric model. Let $\hat{\theta}_p$ and $\hat{\theta}_n$ be the parameter estimates obtained from the parametric and the non-parametric models. Let $\hat{V}_p$ and $\hat{V}_n$ be estimates for the variance of these parameter estimates. Under the null hypothesis that the parametric test is correct the test statistic

$$(\hat{\theta}_n - \hat{\theta}_p)'(V_n - V_p)^{-1}(\hat{\theta}_n - \hat{\theta}_p)$$

has a chi-square distribution with degrees of freedom equal to the number of parameters in the model.

Figure 1: Estimated Baseline Hazard in Frailty Model

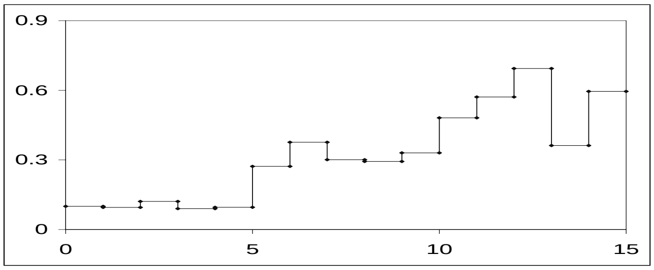

Table 2: Estimates for the With and Between Variance for Two-level Frailty Model.

n	100	500	1000
Bias for σ_w	-0.05	-0.01	-0.01
Bias for σ_b	-0.04	-0.02	0.00
Coverage for σ_w	0.77	0.89	0.89
Coverage for σ_b	0.75	0.86	0.87

3.2 Two-level Frailty Models

The two-level frailty models allow for association between survival times not only on the individual level but also on the cluster level. Suppose as in the previous section that T_1 and T_2 are the time from remission to relapse and the time from relapse to death in cancer patient. Suppose that the patients are grouped by facility of treatment. Denote the survival times for patients i in hospital j are T_{1ij} and T_{2ij} and the corresponding hazard functions by h_{1ij} and h_{2ij}. The two-level frailty model is described by

$$h_{1ij}(t) = Exp(\eta_{wij} + \eta_{bj})\lambda_1(t) \tag{21}$$

$$h_{2ij}(t) = Exp(\eta_{wij} + \eta_{bj})\lambda_2(t). \tag{22}$$

where η_{wij} and η_{bj} are individual and cluster level normally distributed random effects with zero mean and variance σ_w and σ_b. We conduct a simulation study for this model using the same baseline hazard functions as in the previous section and $\sigma_w = 0.3$ and $\sigma_b = 0.2$. We use 100 replications and sample sizes $n = 100, 500$ and 1000. All clusters are of size 10. The bias and the coverage probability are presented in Table 2. Small parameter underestimation occurs for small sample size, however, when the sample size increases the bias is eliminated and the coverage probabilities approach the nominal 95% level. For this model as well improvements in the estimates can be obtained by the parametric approach.

3.3 Time Varying Covariates and Latent Variables

Time varying covariates and latent variables can be incorporated in the above framework. Suppose that a covariate x_{ij} changes over time. Denote the covariate at time t by x_{ijt}. Similarly, suppose that a latent variable η_{ij} changes over time. Denote η_{ij} at time t by η_{ijt}. The proportional hazard model is now given by

$$[h_{ij}(t)|C_{ij}=c] = \lambda_c(t)Exp(\iota_{rcj}+\gamma_{rcj}x_{ijt}+\kappa_{rcj}\eta_{ijt}) \tag{23}$$

This model generalizes the Xu-Zeger (2001) model to multilevel and mixture settings. Suppose that the covariates and the latent variables change at times d_1,...,d_K. The likelihood of the survival variable T is then equivalent to the likelihood of K survival variables T_1,...,T_K defined as follows

$$T_k = \begin{cases} d_k - d_{k-1} & \text{if } d_k < T \\ missing & \text{if } T < d_{k-1} \\ T - d_{k-1} & \text{otherwise} \end{cases} \tag{24}$$

$$\delta_k = \begin{cases} 1 & \text{if } d_k < T \\ missing & \text{if } T < d_{k-1} \\ \delta & \text{otherwise} \end{cases} \tag{25}$$

where δ_k is the censoring indicator of T_k. The proportional hazard model for T_k is

$$[h_{ijk}(t)|C_{ij}=c] = \lambda_{ck}(t)Exp(\iota_{cj}+\gamma_{cj}x_{ijk}+\kappa_{cj}\eta_{ijk}) \tag{26}$$

Thus the model for the survival variable T with time varying covariates is equivalent to a multivariate model for T_1,...,T_K without time varying covariates.

We illustrate the above idea with an example following the discussion in Xu-Zeger (2001). Let Y_{it} be an observed dependent variable for individual i at time $t = 0, ..., 5$. Suppose that Y_{it} follows a linear growth model

$$Y_{it} = \eta_{it} + \varepsilon_{it} \tag{27}$$

$$\eta_{it} = \mu_i + \beta_i t \tag{28}$$

where μ_i and β_i are normally distributed random effects. For t in the interval [k,k+1) the hazard function for T_i is given by

$$h_i(t) = \lambda(t)Exp(\gamma\eta_{ik}). \tag{29}$$

This model that can be very useful in practice. For example Y_{it} can represent a biological marker that predicts the survival variable T_i. The model estimates a linear trend for Y_{it} and allows for a measurement error.

In the model described by Xu-Zeger (2001) the predictor varies continuously over time, while the predictor in model (29) changes stepwise. Xu-Zeger (2001) estimate the model using an advance MCMC algorithm, while model (29) can be estimated simply by the ML algorithm. If the step size in the changes is chosen to be sufficiently small the difference between the two models will be negligible.

Table 3: Average Parameter Estimates and Coverage Probability for Survival Analysis with Time Varying Latent Variable

Para-	True			
meter	Value	n=100	n=200	n=500
γ	0.3	0.31(0.95)	0.30(0.95)	0.30(0.94)
α_1	0.2	0.20(0.93)	0.18(0.95)	0.19(0.97)
α_2	0.1	0.10(0.96)	0.10(0.96)	0.10(0.96)
σ_1	1.0	0.96(0.90)	0.99(0.94)	0.98(0.94)
σ_2	0.2	0.19(0.92)	0.20(0.89)	0.20(0.93)
ρ	0.1	0.11(0.96)	0.10(0.94)	0.10(0.92)
θ_0	1.0	1.01(0.96)	0.99(0.93)	1.00(0.96)

To evaluate the performance of the proposed estimation method we conduct a simulation study. We generate and analyze the data with model (27-29). There are 12 parameters in the model. The parameter γ, the means α_1 and α_2 of the random effects μ_i and β_i, the variances σ_1 and σ_2 of these random effects and their covariance ρ, and the six Y_i residual variance parameters θ_i. The data is generated with the following parameter values $\gamma = 0.3$, $\alpha_1 = 0.2$, $\alpha_2 = 0.1$, $\sigma_1 = 1$, $\sigma_2 = 0.2$, $\rho = 0.1$ and $\theta_i = 1$. The data is right censored at 5. The measurements Y_i are available only if $T > i$. The baseline hazard function used in the generation process is $\lambda(t) = 0.1i$ for t in the interval $[i-1, i)$. We generate 100 samples of size $n = 100, 200$ and 500. Using formulas (24) and (25) we transform the data into the multivariate format. A sample record looks like this

$$1, 1, 1, 0.12, *, -0.34, 0.41, -0.57, -1.32, *, *, 1, 1, 1, 0, *$$

where the data is in this order $T_1, ..., T_5, Y_0, ..., Y_5, \delta_1, ..., \delta_5$. The data is interpreted as follows. Since $T_1 = 1$ and $\delta_1 = 1$ the individual survived during the interval $[0, 1]$. The same occurs in intervals [1,2] and [2,3]. During the interval [3,4] the individual dies at time $T = 3.122$, since $T_4 = 0.122$. Both T_5 and δ_5 are unobserved since death has occurred already. The biological marker variables Y_i are observed at times 0,...,3 and are missing for times 4 and 5 since death has occurred already.

Table 3 contains the average parameter estimates and the coverage probabilities for most parameters. During the data analysis 5 replications encountered convergence problems, 4 of these occurred for sample size 100 and 1 for sample size 200. The results indicate that the estimator performs very well. The bias in the parameter estimates is small and the coverage probabilities close to the nominal 95% value.

4 Aspects of Mixture Survival Modeling

Survival models with normally distributed latent variables have been utilized much more than survival models with categorical latent variables. Larsen (2004) describes one application of mixture survival models however many aspects of these models have not been explored yet. In this section we discuss some of the challenges and unique modeling capabilities presented by the mixture survival models. The approach implemented in Mplus differs from Larsen's (2004) approach in one important aspect. In

Larsen (2004) the baseline hazard function varies across classes only by a single multiplicative factor. Consider a two class model where the baseline hazard function in the two classes are $\lambda_1(t)$ and $\lambda_2(t)$. In Larsen (2004) $\lambda_1(t)$ is estimated as a non-parametric step function while $\lambda_2(t)$ is constrained by the following equation

$$\lambda_2(t) = \alpha\lambda_1(t) \tag{30}$$

where α is parameter that is estimated. In contrast Mplus will estimates both $\lambda_1(t)$ and $\lambda_2(t)$ as unconstrained nonparametric step functions. The advantage of Larsen's approach is that the class effect on the baseline can be explicitly estimated, while in the Mplus approach the effect of the class variable is not obtained directly, simply because the two baseline function are completely unconstrained and no parameter summarizes the difference between the two baseline hazards. Essentially when using the Mplus approach we know that the baselines are different across class but we do not know how. Using a parametric approach can resolve this problem.

The advantage of the Mplus approach is that it does not depend on the proportionality assumption (30). In the following simulation study we explore the consequences of incorrectly assuming (30). We use a Cox regression model for a survival variable on a single predictor X with standard normal distribution and with slope $\beta = 1$. We generate the data according to a two class model where the baseline hazard in the first class $\lambda_1(t)$ is as in (20) while in the second class $\lambda_2(t) = 0.1$. The Cox regression slope β is 1 in both classes. For simplicity, to avoid the complexities of the class measurement model we identify exactly the class variable by a single binary indicator U, $P(U = 1|C = 1) = 1$ and $P(U = 2|C = 2) = 1$. Thus the C variable is essentially observed by its perfect binary indicator U. We generate 100 samples of size 5000. The large sample size eliminates any possible finite sample size effects. When we analyze this data with the Mplus approach the average parameter estimate for β is 1.0023 and the MSE for this parameter is 0.0004. When we analyze the data with Larsen's approach, based on the incorrect proportionality assumption, the average parameter estimates for β is 0.9074 and the MSE is 0.0089. We conclude that incorrectly assuming the proportionality property can result in biased estimates and larger MSE.

In the next simulation we explore whether for situations when the proportionality property holds Larsen's method can lead to a reduction in MSE. Using $\lambda_1(t)$ again as in (20) and $\lambda_2(t) = 2\lambda_1(t)$ with sample size $n = 100$ we obtained MSE for $\hat{\beta}$ of 0.0276 for Larsen's method and 0.0177 for Mplus method. Thus surprisingly, in this example the less restricted Mplus method gives more accurate estimates even when the proportionality assumption holds.

Survival mixture models can be viewed as the joint models for survival variables and latent class variables. The models can be used to explore population heterogeneity while modeling the survival variables. The latent class variable can be viewed as a predictor for the survival variables but also the survival variables can be viewed as class indicators. The model within each class is based on the PH property (3), however the property will not hold in general for the total population combining all the classes. Thus, mixture survival models can also be used to model survival data that does not satisfy the PH property.

5 Comparison Between Continuous and Discrete Time Survival Analysis

Suppose that the survival variable T takes only a finite number of values. Then an alternative modeling approach known as discrete time survival analysis can be used, see for example Muthen and Masyn (2005). The model consists of a number of logistic regressions fitting the incremental probability of survival

$$P(T > k + 1|T > k) = \frac{Exp(\tau_k - \beta X)}{1 + Exp(\tau_k - \beta X)}.$$

The discrete time survival approach can also be used when the variable T is continuously distributed simply by categorizing the T variable into a finite number of values. Let $T^* = [T/h]h$, where h is a small number and $[x]$ denotes the nearest integer to x. T^* is a discrete approximation of T. There are three possible modeling approaches, modeling T with continuous time survival model, modeling T^* with continuous time survival model and modeling T^* with discrete time survival model. Here we will show that all three models lead to similar parameter estimates for sufficiently small h. As h converges to 0, the differences between T and T^* becomes negligible and thus the Cox regression for T and T^* yields similar results. Next we show that the Cox regression for T^* yields similar results as the discrete time survival model for T^*. For the Cox regression model we have

$$P(T > k + 1|T > k) = Exp(-Exp(\beta X)c_1)$$

where c_1 is independent of X and converges to 0 as h converges to 0. Using first order Taylor expansion we get that

$$P(T = k + 1|T > k) = 1 - Exp(-Exp(\beta X)c_1) \approx$$

$$Exp(\beta X)c_1.$$

The same holds for the discrete time survival model

$$P(T = k+1|T > k) = \frac{Exp(\beta X - \tau_k)}{1 + Exp(\beta X - \tau_k)} \approx Exp(\beta X - \tau_k)$$

since τ_k converges to infinity as h converges to 0. The incremental survival probabilities in both model depend on X approximately via the same functional form and thus the parameter estimates will be approximately the same. Thus the discrete time survival can be used as an approximation to the the Cox regression model as long as the

Table 4: MSE of β for continuous data with ties.

n	Breslow Mplus	Breslow SAS	Efron SAS	Exact SAS	Discrete SAS
100	0.026	0.025	0.026	0.026	0.027
1000	0.002	0.002	0.002	0.002	0.002

categorization of the T variable is sufficiently detailed. Note however that if the categorization is very detailed the discrete time survival method estimates many logistic regression equations and thus the estimation becomes numerically very inefficient.

6 Comparing Mplus and SAS Estimation for Survival Data with Ties

In this section we evaluate the performance of the methods implemented in Mplus with those implemented in SAS for survival analysis of data with ties. Four partial likelihood estimation methods are implemented in the SAS PHREG procedure. These methods are known as the Breslow (1974) likelihood, the Efron (1977) likelihood the exact likelihood (Kalbfleisch and Prentice, 1980), and the discrete likelihood (Cox and Oakes, 1984). Two methods are available in Mplus, the Breslow profile likelihood approach and the discrete time survival analysis. Since the Breslow profile likelihood approach implemented in Mplus and the Breslow partial likelihood implemented in SAS are both based on the step baseline hazard function we expect to get the same results with these methods for data with and without ties.

To conduct our simulation study we first generate samples drawn from a Cox regression model for a variable T and a single covariate X. The covariate X has a standard normal distribution and slope $\beta = 1$. The baseline hazard function is as in (20). We introduce right censoring in the data by generating an independent exponential variable C with mean 10. If $C < T$ then censoring occurs at time C. We generate 100 samples of size $n = 100$ and 1000. We introduce ties in the data by two methods. The first method preserves the continuity of the data while the second discretizes the data.

6.1 Continuous Data with Ties

To introduce ties in the data while preserving the continuity of the data we create new samples that consist of two identical copies of the samples generated above. Thus each observation is tied to another observation in the sample. This simulation study is intended to mimic practical situations where the data is continuous but there are incidental ties, i.e., the variable T takes many different values but they are not necessarily unique. Table 4 shows the MSE for $\hat{\beta}$ for five estimation methods. Mplus discrete method is not available for continuous data. All five methods perform equally well in this situation.

6.2 Discrete Data

In this section we introduce ties in the data by categorizing the T values into intervals of length h. We use 3 values for $h = 2, 1$, and 0.5. Because T is always less than 16, the maximum number of values that T can attain is thus $L = 8, 16$, and 32 respectively. There are many ties in this data.

The results are presented in Tables 5 and 6. The Mplus Breslow method gives the same results as the SAS Breslow method. The parameter estimates for these two methods are identical for all replications. The Mplus discrete method generally differs from the SAS discrete method, however, for large sample size (n=1000) the estimates obtained by the two methods are identical. The Efron and the Exact methods outperform the other methods especially for the coarse categorizations $h = 2$ and $h = 1$. With $n = 100$ and $h = 0.5$ the performance of the Breslow method is the same as the Efron and the Exact methods. For $n = 1000$ and $h = 0.5$ the Breslow method is slightly worse. This is because the ties influence is stronger as now 1000 points are tied on the same 32 values. The Breslow and Efron methods tend to underestimate the slope while the discrete methods tend to overestimate the slope. The Exact method has a very small bias in general. The bias decreases as the sample size increases for all methods except the Breslow methods. This leads us to the conclusion that the MSE decreases to 0 as the sample size increases for all methods except the Breslow methods. The Breslow methods actually shows smaller variation in the estimates than the Efron and the Exact methods. Thus the larger MSE for the Breslow methods is due to the larger bias.

The methods implemented in SAS are available only for the Cox regression model while the methods implemented in Mplus are available for the general latent variable models. Thus it is important to compare the performance of the two Mplus methods as they are the only possible alternatives for these models. The results in Table 5 indicate that for large sample size (n=1000) and small number of categories (L=8) the discrete method outperforms the Breslow method. This result is reversed for small sample size $n = 100$. This is due again to the fact that the ties for the $n = 1000$ sample are more influential. We conclude that when the data is truly categorical and the number of categories is less than 20 the discrete method should be preferred, especially for large sample sizes. In this case the discrete survival model is a better alternative even if the model we want to estimate is the Cox regression model. When there are more than 20 categories or the sample size is small the Breslow method performs well when compared to the other methods.

7 Conclusion

The continuous time survival modeling framework described here includes many new models that combine survival modeling with latent variable modeling. The model

Table 5: MSE of β for discrete data.

n	h	L	Breslow Mplus	Discrete Mplus	Breslow SAS	Efron SAS	Exact SAS	Discrete SAS
100	2	8	0.068	0.071	0.067	0.024	0.023	0.057
100	1	16	0.035	0.052	0.035	0.021	0.024	0.041
100	0.5	32	0.024	0.040	0.024	0.022	0.024	0.031
1000	2	8	0.058	0.033	0.058	0.009	0.004	0.031
1000	1	16	0.021	0.016	0.021	0.003	0.003	0.015
1000	0.5	32	0.008	0.007	0.008	0.003	0.002	0.007

Table 6: Bias of β for discrete data.

n	h	L	Breslow Mplus	Discrete Mplus	Breslow SAS	Efron SAS	Exact SAS	Discrete SAS
100	2	8	-0.24	0.19	-0.24	-0.09	-0.02	0.15
100	1	16	-0.14	0.14	-0.14	-0.04	0.00	0.10
100	0.5	32	-0.08	0.10	-0.08	-0.02	0.00	0.05
1000	2	8	-0.24	0.17	-0.24	-0.08	-0.03	0.16
1000	1	16	-0.14	0.11	-0.14	-0.03	-0.01	0.11
1000	0.5	32	-0.08	0.07	-0.08	-0.01	0.00	0.06

estimation is the standard maximum likelihood estimation, which is computationally feasible even for advanced models. The framework is implemented in Mplus and can be utilized in many practical applications.

References

Asparouhov, T. and Muthen, B. (2005), " Multivariate Statistical Modeling with Survey Data," *Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.* http://www.fcsm.gov/events/papers05.html

Bradburn, M. J., Clark T.G., Love S.B., and Altman D.G. (2003), "Survival Analysis Part II: Multivariate Data Analysis - an introduction to concepts and methods," *British Journal of Cancer*, **89**, 431–436.

Breslow, N.E. (1974), "Covariance Analysis of Censored Survival Data," *Biometrics*, **30**, 89–99.

Clayton, D. G. (1988), "The analysis of event history data: A review of progress and outstanding problems," *Statistics in Medicine*, **7**, 819–841.

Cox, D.R. (1972), "Regression Models and Life-Tables (with Discussion)," *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

Cox, D.R., and Oakes, D. (1984), *Analysis of Survival Data*, London: Chapman and Hall.

Hsieh, F. Y. (1995), "A Cautionary Note on the Analysis of Extreme Data with Cox Regression," *The American Statistician*, **49(2)**, 226–228.

Hougaard, P. (2000) *Analysis of Multivariate Survival Data*, Springer, New York.

Hausman, J.A. (1978), "Specification Tests in Econometrics," *Econometrics*, **46**, 1251–1271.

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

Klein J.P. & Moeschberger, M.L. (1997) *Survival analysis: techniques for censored and truncated data,* New York: Springer.

Larsen, K. (2004), "Joint Analysis of Time-to-Event and Multiple Binary Indicators of Latent Classes," *Biometrics*, **60(1)**, 85–92.

Larsen, K. (2005), "The Cox Proportional Hazards Model with a Continuous Latent Variable Measured by Multiple Binary Indicators," *Biometrics*, **61(4)**, 1049–1055.

McCullagh P. & Nelder, J. A. (1989) *Generalized Linear Models*, London, Chapman & Hall.

Murphy, S.A. & van der Vaart, A.W. (2000), "On profile likelihood," *Journal of the American Statistical Association*, **95**, 449–465.

Muthen, B. (1984), "A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators," *Psychometrika*, **49**, 115–132.

Muthen, B. & Masyn, K. (2005), "Discrete-time survival mixture analysis," *Journal of Educational and Behavioral Statistics*, **30**, 27–28.

Muthen, L.K. & Muthen, B.O. (1998-2006), *Mplus User's Guide,* Forth Edition, Los Angeles, CA: Muthen & Muthen.

Johansen (1983), "An extension of Cox regression Model," *International Statistical Review*, **51**, 165–174.

Singer, J. D. & Willett, J. B. (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence,* New York: Oxford University Press.

White, H. (1980), "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity," *Econometrica*, **41**, 733–750.

Xu J, Zeger SL (2001), "Joint analysis of longitudinal data comprising repeated measures and times to events," *Applied Statistics*, **50(3)**, 375–387.