Modelling long term survival with non-proportional hazards

Perperoglou, A.

Chapter 4

A relaxation of the Gamma frailty (Burr) model.

Abstract

Frailty models are used in univariate data to account for individual heterogeneity. In the popular gamma frailty model the marginal hazard has the form of a Burr model. Although the Burr model is very useful and can offer insight on the data, it is far from perfect. The estimation of the covariate effects is linked to the baseline hazard and this makes the model coefficients hard to interpret. At the same time, the frailties are assumed constant over time, while biological reasoning in some cases may indicate that frailties may be time dependent. In this Chapter we present a relaxation of the Burr model which is based on loosening the link between the estimation of the covariate effects and the baseline hazard. This can be achieved by replacing the cumulative baseline hazard in the Burr model by a set of time functions, and the frailty variance by a vector of coefficients directly estimated from the data using a partial likelihood. We illustrate the similarities of the model with the Burr model and a further extension of the latter, a model with an autoregressive stochastic process for the frailty. We compare the models on simulated data sets with constant and time dependent frailties and show how the relaxed Burr models performs on two different real data sets. We show that the relaxed Burr model serves as a good approximation to the Burr model when the frailty is constant, and furthermore it gives better results when the frailty is time dependent.

4.1 Introduction

Frailties are used in time to event data in a variety of contexts and for a variety of purposes. One use is in modelling multivariate survival times, arising from clustered data or repeated measures on the same subject, for instance in a

treatment which is randomly assigned to the left or the right eye of a patient. Then each eye is an individual observation, but a pair of eyes forms a cluster. Moreover, frailties can be used in univariate data to account for unexplained heterogeneity amongst cases, or, in general, as a model modifier, to explain the lack of fit due to missing covariates.

In this Chapter we deal with individual frailty as the means to model and explain individual heterogeneity. In a survival study each patient has their own frailty and those that are more frail tend to fail earlier in time. However, there is no reason to assume that the frailty remains constant over a long time period. An example of increasing frailty can occur in diseases where recurrent unobserved infections can influence the body's ability to deal with the disease. Although biological reasoning will lead to the idea of increasing or decreasing frailty as time passes by, in most frailty models the random effect is assumed to be constant over follow-up. However, this assumption is restrictive, especially in large follow up studies. In such cases it is more reasonable to assume that the random effect may change over time in the same manner as time varying effects of fixed covariates.

While many models have been suggested for modelling time varying effects of fixed covariates ([43], [58], [103],[77]) not much literature is devoted to modelling time dependent frailties. Paik et al. [72] worked on a generalization of a multivariate frailty model by introducing additional frailty terms for different time-intervals, Wintrebert et al. [106] proposed a multivariate frailty model with a power parameter which allows a centre-specific frailty to vary among individuals and extended this model to allow the center-specific frailty to change with time, Yau et al. [109] allowed the frailties to be time varying according to an AR(1) process and proposed ML and REML methods for estimation, while Manda and Meyer [66] presented a similar model, within a Bayesian framework. Other approaches include the frailty models based on Lévy processes, proposed by Gjessing et al. [35] while Yashin and Manton [108] provide a review of hazard models with unobserved and partially observed covariate processes. These last two approaches provide explicit answers and expressions for the survival function, though without the direct extension of standard gamma frailty models which we provide in this work.

For estimating a standard gamma frailty model the *marginal model* is needed which is written in terms of the marginal survival functions via a Laplace transformation. However, when the random effects are assumed to be time dependent and have some autocorrelation structure, the creation of an appropriate marginal likelihood is not always feasible. The exact likelihood

method requires integrating out the random effects which can only be done in a few special cases or by the means of numerical integration. To avoid the problem of making very restrictive assumptions we present an approximation to what would have been a Burr model with autocorrelated frailties. In the Burr model the conditional (given the frailty) baseline hazard appears in the marginal hazard, see (4.1) below. In this Chapter we undo that link by introducing in its place an unspecified function, which can be flexibly modeled as a combination of B-spline basis functions. The Chapter is organized as follows: in Section 4.2 the standard Burr model and an extension for autocorrelated frailties is presented. Section 4.3 contains the relaxed Burr model and illustrates the estimation process along with the some properties of the model. In Section 4.4 we present simulation studies followed by applications to real data sets in Section 4.5. The Chapter closes with a discussion. Some mathematical detail is given in the Appendix.

4.2 Burr model and autocorrelated frailties

A frailty model can be considered a one dimensional expansion of a proportional hazards model. A random term Z is included in a simple Cox model [24] , to account for individual heterogeneity, which is assumed to act multiplicatively on the baseline hazard. The model is defined as

$$\lambda(t|X,Z) = Z\lambda_0(t)\exp(X\beta)$$

where $\lambda_0(t)$ is the baseline hazard, X the vector of covariates and β the coefficients to be estimated. Since the frailties are unobservable, one has to work with the marginal hazard given by

$$\lambda(t|X) = \lambda_0(t)\exp(X\beta)E(Z|T \geq t)$$

A popular choice for the distribution of the random effects is gamma. This leads to the gamma frailty model, or Burr model, where given that $Z \sim \Gamma(1/\xi, 1/\xi)$ and $\mathrm{E}(Z) = 1$ with $\mathrm{var}(Z) = \xi$, the marginal model is derived as:

$$\lambda(t|X) = \frac{\lambda_0(t)\exp(X\beta)}{1+\xi\Lambda_0(t)\exp(X\beta)} \tag{4.1}$$

where $\Lambda_0(t)$ is the cumulative baseline hazard. The most important deviation from a simple Cox model is that the effect of the covariates disappears over time, that is the hazard ratio $\lambda(t|X_1)/\lambda(t|X_2)$ converges to one for any pair of covariate values, while in the Cox model this ratio is constant over time.

Another feature of the model is that the effects of covariates are exchangeable and their estimation is linked to the estimation of the baseline hazard. The latter leads to problems of inference, since the coefficients can be only explained at time $t = 0$. An EM algorithm may be used to estimate the model. For an introduction to frailty models see Aalen [1].

Extension from the simple frailty model

In the gamma frailty model the frailties are assumed constant. In some applications however this assumption may not hold. A frailty that changes over time is a reasonable assumption for cases with long follow up. It is reasonable to assume frailties that arise from a stochastic process over time, $Z(t)$, with mean=1 and covariance function $C(s,t)$ =Cov$(Z(s), Z(t))$. An example of a possible model is an autoregressive frailty model, where the covariance could be $C(s,t) = \sigma^2 e^{-\kappa|s-t|}$, with σ the variance of the random effect and κ to be estimated from the data. The conditional hazard function is then defined as:

$$\lambda(t|X, Z(t)) = \lambda_0(t) \exp(X\beta) Z(t)$$

with the marginal survival function given as:

$$S(t) = E[e^{-\exp(X\beta)\int_0^t \lambda_0(s)Z(s)ds}]$$

Under the condition $\mathrm{E}(Z(t)) < \infty$ (proposition 3, [108]), the marginal hazard is given by $\lambda(t|X) = \lambda_0(t) \exp(X\beta) E[Z(t)|T \geq t]$. In the appendix it will be shown that this can be approximated by :

$$\lambda(t|X) = \frac{\exp(X\beta)\lambda_0(t)}{1 + \exp(X\beta)\int_0^t \lambda_0(s)C(s,t)ds} \tag{4.2}$$

This leads to effects of covariates that converge but do not eventually die out. The behavior of the hazard ratio $\lambda(t|X_1)/\lambda(t|X_2)$ depends very much on the covariance structure and the 'memory' of the frailty process. For example, the ratio might converge to one in the beginning and diverge away from one towards the initial value $\exp((X_1 - X_2)\beta)$ for later t, if the frailty process loses its memory and starts behaving as white noise.

Although this is a very useful model, in practice it is very hard to estimate. As in the Burr model, this is not a linear model and the estimation of the coefficients is still linked to the estimation of the baseline function. Furthermore, to acquire an expression for the survival function is far from trivial. Even in the

very special case where the baseline hazard is parametric, say following a simple Weibull distribution, the integration of the hazard leads to an incomplete gamma function. Unless the baseline is considered constant there is no closed form for this integral, and to assume such a restrictive condition for estimating the model is of little practical use.

4.3 A relaxation of the Burr model

The idea for a relaxed Burr model comes simply by combining models (4.1) and (4.2). The two models are very similar in respect to the interaction of covariates, the non-linear effects of the covariates and their exchangeability. The link of the estimation of the covariate effects to the baseline complicates the interpretation of the regression coefficients while in the second model the addition of the autoregressive hazard makes the estimation almost impossible. However, a more flexible form of the hazard ratios is a desirable feature.

To achieve a generalization of the Burr model towards the second model we propose the following:

$$\lambda(t|X) = \frac{\lambda_0(t)\exp(X\beta)}{1 + F(t|\theta)\exp(X\beta)} \tag{4.3}$$

We will refer to this as the relaxed Burr model. Here $F(t|\theta)$ can be any continuous nonnegative function starting at $F(0|\theta)$=0. We will always use a linear model $F(t|\theta) = f(t)\theta$ where $f(t)$ can be a simple time function multiplied by an unknown but estimable coefficient θ or any set of q basis time functions that, put together, form a vector function $F(t)$. In that case θ becomes a q-vector of coefficients estimable from the data, and the pattern of $F(t|\theta)$ is very important for the way the model behaves. The only restriction imposed on these sets of time functions is $F(0) = 0$. The condition $F(t|\theta) \geq 0$ cannot be guaranteed with this set-up, but does not impose a problem in practice. The researcher has the task to choose an appropriate set of functions, among many choices, such as simple polynomials, fractional polynomials and B-splines. This choice can depend on some a priori preference or just by looking at the likelihood of the models. Throughout this Chapter we used B-splines functions for their simplicity and flexibility.

Estimation

For estimation of the relaxed Burr model the partial likelihood and the Newton-Raphson algorithm can be used. A partial likelihood can be constructed by

profiling over $\lambda_0(t)$ in the usual way. Parameters β and θ can then be obtained by maximizing :

$$L = \prod_{i=1}^{n} \left[\frac{e^{X\beta}(1 + F(t|\theta)e^{X\beta})^{-1}}{\sum_{j \in R(t_i)} e^{X_j\beta}(1 + F(t|\theta)e^{X_j\beta})^{-1}} \right]^{\delta_i}$$

where δ_i is the event indicator. We have written an algorithm that maximizes both β's and θ's simultaneously and give standard errors for all the coefficients. Alternatively, one could estimate the β coefficients first, keeping the θ's fixed, and then use the updated estimates of β's for estimating the θ's. It is our experience that estimating all coefficients at the same time is faster and more stable, as the iterative procedure is more sensitive to the starting values and may lead to convergence problems. Once the coefficients are estimated, an estimator of the baseline hazard can be computed using the standard Breslow estimator.

Properties

A basic feature of a simple Cox model is that of *proportional hazards*. That is, for two cases with different covariate values say, X_1 and and X_2, respectively, their hazard ratio remains constant over time. On the other hand, under a gamma frailty model, the hazard ratio for two different cases converges as described earlier. A relaxed Burr model adds more flexibility in modelling the hazards, provided by the pattern indicated by the time functions.

As an approximation of the Burr model, a relaxed Burr model should be able to model converging hazards. In some cases $F(t|\theta)$ can serve as an approximation of $\xi\Lambda_0(t)$ in (4.1) and give almost identical estimates. On the other hand there can be more flexibility. For instance, assume for simplicity that $F(t|\theta)$ is a straight line based on just one time function. If the estimated $\hat{\theta}$ from the data is smaller than the estimated $\hat{\xi}$ from a Burr model, then the hazards given by a relaxed Burr model will converge, but at a slower rate than the Burr model.

Another property of the relaxed Burr model is its ability to model hazards that converge in a similar way to model (4.2). A model with such hazards is met when $F(t|\theta)$ is an increasing function to begin with, but constant at the later times, thus mimicking the behavior of $\int_0^t \lambda_0(s)C(s,t)ds$ for the autoregressive model mentioned above.

The merit of flexibility can be seen when $F(t|\theta)$ is not a monotonic function. A pattern that can emerge is that of converging hazards up to one point, when

Table 4.1: Average coefficient estimates and standard errors in brackets, under the three different models, proportional hazards (PH), Burr (B) and relaxed Burr (RB), for simulated survival data on n=400 cases with time dependent frailties from a multivariate gamma distribution. Two covariates present, a dichotomous and a polytomous with coefficients $\beta_1 = 0.8$ and $\beta_2 = 0.4$. The relaxed Burr model was fitted with B-splines as time functions on df=3. Each data set was simulated 200 times.

		$\xi = 1$			$\xi = 0.5$		
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\xi}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\xi}$
	PH	0.44 (.11)	0.22 (.03)		0.58 (.10)	0.28 (.03)	
$c = 1.00$	B	0.76 (.17)	0.38 (.06)	0.79	0.81 (.16)	0.39 (.06)	0.41
	RB	0.83 (.22)	0.40 (.09)		0.88 (.25)	0.42 (.14)	
	PH	0.48 (.11)	0.24 (.03)		0.59 (.11)	0.29 (.03)	
$c = 0.75$	B	0.68 (.16)	0.35 (.07)	0.45	0.76 (.18)	0.37 (.06)	0.30
	RB	0.84 (.25)	0.41 (.16)		0.89 (.29)	0.42 (.07)	
	PH	0.50 (.10)	0.25 (.03)		0.62 (.11)	0.30 (.04)	
$c = 0.50$	B	0.62 (.16)	0.31 (.06)	0.26	0.77 (.15)	0.38 (.06)	0.264
	RB	0.77 (.25)	0.38 (.14)		0.90 (.20)	0.44 (.09)	
	PH	0.54 (.11)	0.27 (.03)		0.65 (.10)	0.32 (.03)	
$c = 0.25$	B	0.63 (.16)	0.31 (.05)	0.18	0.74 (.14)	0.36 (.05)	0.16
	RB	0.85 (.31)	0.40 (.12)		0.89 (.17)	0.42 (.17)	

$F(t|\theta)$ increases, and from then on diverging hazards can emerge as $F(t|\theta)$ drops towards zero. At the low points of $F(t|\theta)$ the hazards will diverge toward proportionality, since the model essentially becomes a simple proportional hazards model when $F(t|\theta) = 0$.

4.4 Simulations

To investigate how the model behaves in different situations we simulated several data sets. In this series of simulations we wanted to first reproduce situations where the frailty is constant and the real model is a Burr model, and check how the relaxed Burr model performs. Next we simulated a series of data where the individual frailties are time dependent and arise from a simple auto-

correlation function. For the second more complicated simulation we followed the algorithm proposed in [45]. Two frailty variances ξ were considered, 0.5 and 1, and we took $Z(t) \sim \Gamma(1/\xi, 1/\xi)$ and $\text{Corr}(Z(t), Z(t+u)) = \exp(-\rho|u|)$. Survival times were censored at 10 units, with the censoring rate lying between 1% and 4%. The values of ρ were chosen so that the frailty correlation, say c, at 5 units apart ($u = 5$) was either 0.25, 0.50, 0.75 or 1. We simulated two independent covariates, one dichotomous taking values 1 or -1 with probability 0.5 and a second categorical covariate with values -2,-1, 1 and 2 with probability 0.25. The true value of the coefficients was 0.8 for the dichotomous covariate and 0.4 for the polytomous.The baseline hazard was taken to be constant at the value of 1.

Note that when $\text{Corr}(Z(t), Z(t+u)) = 1$ the frailties are constant in time and that the data are simulated from model (4.1) . On the other extreme, if $\text{Corr}(Z(t), Z(t+u)) = 0$ the frailties are uncorrelated and regarded from the models as noise which is absorbed in the baseline hazard. The number of individuals n was 400, and each of these data sets, under the different conditions, were simulated 200 times. Table 4.1 presents the average estimated coefficients from the three models, along with standard deviations obtained as the variation between the simulations which serves as an estimate of the standard error for the individual estimates, and the average estimate of the variance under the Burr model and its standard deviation.

The time functions used in the relaxed Burr model were B-splines of second degree with 3 interior knots positioned at the 1st, 2nd and 3rd quantile of survival time. This is a rather flexible setting for the relaxed frailty model, and one that requires more computing time and iterations. However, the results are very encouraging. When the variance of the random effect is 1, the relaxed Burr model gave estimated coefficients closer to the real ones, in all cases, even when the data actually arise from a model with constant frailties. As has been discussed by Barker and Henderson [12], the E-M approach used for estimation leads to finite sample underestimation of the frailty variance, and as a result the corresponding regression parameters are also underestimated. Since the relaxed Burr model does not actually estimate a frailty variance the small sample bias of the Burr model is not present, and the regression coefficients are consequently not affected. When the variance of the random effect is smaller (0.5) the proposed model tends to slightly overestimate the effect of the covariates. Although this overestimation is quite small some care should be taken when fitting the relaxed model to data when little evidence of overdispersion is present.

Table 4.2: Coefficient estimates, standard errors and full log likelihood for the lung cancer data, under the three different models, proportional hazards (PH), Burr (B) and relaxed Burr (RB). Estimated variance of frailty 0.562

	$\hat{\beta}_{age}$	$\hat{\beta}_{sex}$	$\hat{\beta}_{act}$	
PH	0.010 (.008)	-0.683 (.180)	0.346 (.082)	
B	0.012 (.011)	-0.916 (.239)	0.486 (.110)	
RB	0.013 (.007)	-1.060 (.292)	0.567 (.145)	
	$\hat{\beta}_{anor}$	$\hat{\beta}_{hoar}$	$\hat{\beta}_{met}$	log-lik
PH	0.314 (.145)	0.682 (.210)	0.418 (.220)	-1284.06
B	0.577 (.198)	0.850 (.296)	0.508 (.304)	-1282.09
RB	0.664 (.271)	1.071 (.407)	0.532 (.363)	-1280.42

In general, when fitting a Burr model the estimates of the coefficients only make sense at the start of the time period when $t = 0$. That holds of course also for the relaxed Burr model. Having the right β's does not guarantee that the model is right, since the behavior of the hazard and survival functions depends not only on β but also at the baseline hazard and the variance, in the case of the Burr model, and the time functions and their coefficients in the case of the relaxed Burr model. As a consequence the results of the simulations cannot be used as the only argument that a relaxed Burr model provides a good fit to the data, but they have to be viewed in parallel with plots of survival and hazard functions. Nevertheless, getting the right estimates for β's is a good starting point and an important indication that the model is on the right track.

4.5 Applications

Muers et al. (1996) [70] studied 272 patients suffering from lung cancer. The endpoint of the study was survival, the median survival time was 6 months (follow up period up to 34 months) and 83% of the cases died within the follow up period. There is information on six covariates; the age of the patient (38-72 years), gender (0 for females, 1 males), an activity score ranging from 0 to 4 with low values being good, anorexia (0 is absent, 1 present), hoarseness (0 is absent, 1 present) and metastasis (0 is absent, 1 present). Table 4.2 presents the coefficient estimates of the different models along with their full likelihood. For the relaxed Burr model we chose cubic B-splines with no interior knot as time

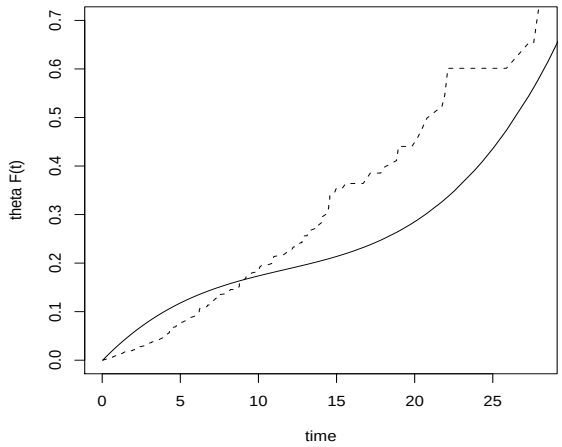

Figure 4.1: The behavior of $\hat{\xi}\hat{\Lambda}_0(t)$ from the Burr model (dotted line) and $\hat{F}(t|\theta)$ from the relaxed Burr model (solid line), versus time (in months), lung data

functions. Using this approach we allow the time functions to be flexible and thus let the model reveal the appropriate functional form of the time varying pattern.

The Burr model was fitted using `R 2.0.1` for Windows [81] and `survival` package [64]. The model has 6 regression parameters to be estimated plus one extra for the estimation of the frailty variance, which in this application was 0.56. The relaxed Burr adds three parameters to estimate $F(t|\theta)$. For this small variance of the random effect the relaxed Burr model gave slightly bigger estimates for the coefficients, as happened in the simulation studies. However, the standard errors are also bigger. The full log-likelihood was computed as the sum of log survival and log hazard. There was a difference in the full likelihood under the three approaches, suggesting that the relaxed Burr model provides a better fit to the data.

In Figure 4.1 a plot of $\hat{F}(t|\theta)$ is presented based on the obtained θ estimates $0.407, -0.184, 1.391$, along with a plot of $\hat{\xi}\hat{\Lambda}_0(t)$ from the Burr model. It can be seen that $\hat{F}(t|\theta)$ increases at a modest rate for the first 20 months and

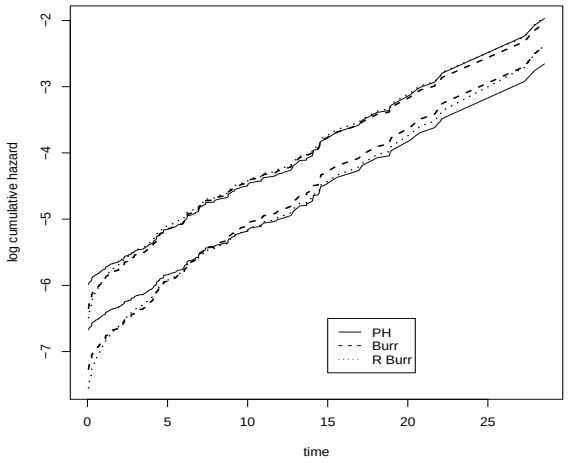

Figure 4.2: Log cumulative hazard functions of male (lower lines) versus female (upper lines) patients with lung cancer under the three different models

then this rate increases. As a result, a plot of the cumulative hazards of male versus female patients (Figure 4.2), with all other covariates at their mean values, shows that the cumulative hazards given by the relaxed Burr model are converging at a slower rate than the Burr model. However, they come very close to the latter at the last few months of the time period. The cumulative hazards of the proportional hazards model are added as a reference in the graph.

For comparing the survival functions under the different models, four covariate patterns emerging from the data were considered. The first group of patients contains fairly young patients (aged 38) with all the other covariates being 0. The second group has a pattern of covariates age=61, sex=0, active index =1, anorexia=0, hoarseness=0 and metastasis=0, for the third group the corresponding values were: 65, 1, 2, 0, 0 and 0, and the highest risk group had covariate values: 70, 1, 3, 1, 0 and 0. Figure 4.3 presents the survival curves for the different covariate patterns of the patients, moving from cases with good prognosis to those with worse. The differences in terms of survival can hardly be seen. The two models show very similar results, with only the estimation of the survival function under the Burr model for the worst prognosis patients

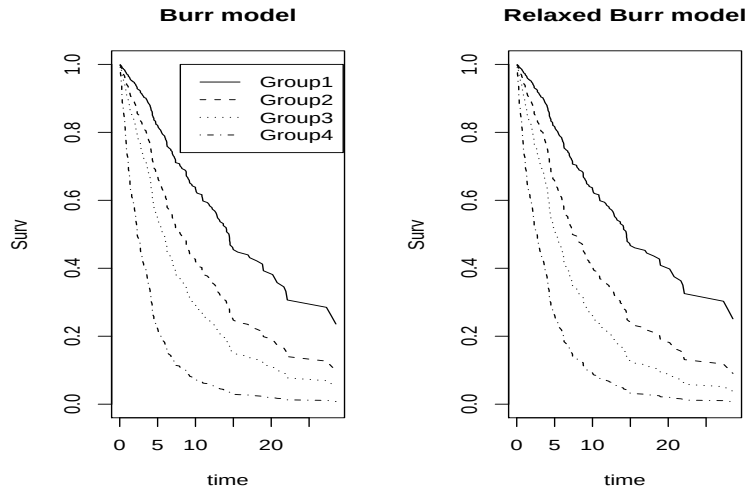

Figure 4.3: Survival of patients with lung cancer moving from cases with good prognosis (group1) to worse prognosis (group4) under the Burr and the relaxed Burr model.

being a bit less optimistic.

The second application comes from register data of adults suffering from acute myeloid leukemia (AML) in Great Britain. There are 1043 cases recorded between 1982 and 1998, originally analyzed in [44]. Median survival time was 6 months and 2.5% of the patients survived for more than 10 years. We have complete information on five covariates, the age of the patients (in years from 14 to 92), gender (0 female, 1 male) , white blood cell (WBC) count at diagnosis (truncated at 500 units with 1 unit $=50\times10^9$/L) and deprivation, an index of poverty (from -6 to 9.5 for the enumeration district of residence) for the subject's residence area. The original analysis showed clear evidence of a large random effect, with a variance estimate from the gamma frailty model of 0.97. We also fitted a simple proportional hazards model to the data and a relaxed Burr model with cubic B-splines with no interior knot as time functions once again. The coefficient estimates are given in Table 4.3 along with the full log likelihoods. The best model according to the likelihood is the relaxed Burr model. The estimated covariate effects are very close for the Burr and relaxed

Table 4.3: Coefficient estimate, standard errors and full log likelihood for the acute myeloid leukemia data, under the three different models, proportional hazards (PH), Burr (B) and relaxed Burr (RB). Estimated variance of frailty 0.976

	$\hat{\beta}_{age}$	$\hat{\beta}_{sex}$	$\hat{\beta}_{wbc}$	$\hat{\beta}_{dep}$	log-lik
PH	0.029 (0.002)	0.052 (0.067)	0.003 (0.001)	0.029 (0.009)	-6204.63
B	0.049 (0.003)	0.058 (0.053)	0.006 (0.001)	0.058 (0.015)	-6182.12
RB	0.053 (0.004)	0.057 (0.056)	0.006 (0.001)	0.056 (0.016)	-6177.37

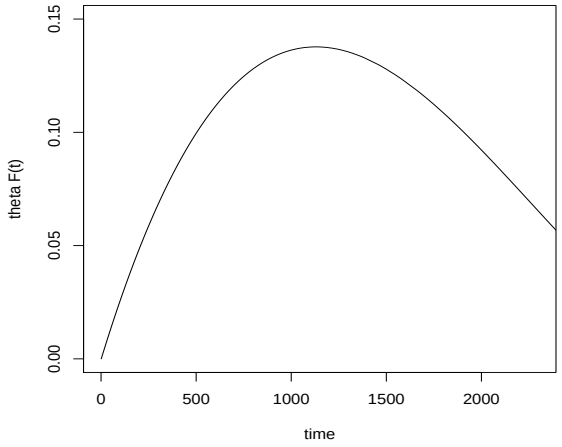

Figure 4.4: $\hat{F}(t|\theta)$ versus time, AML data

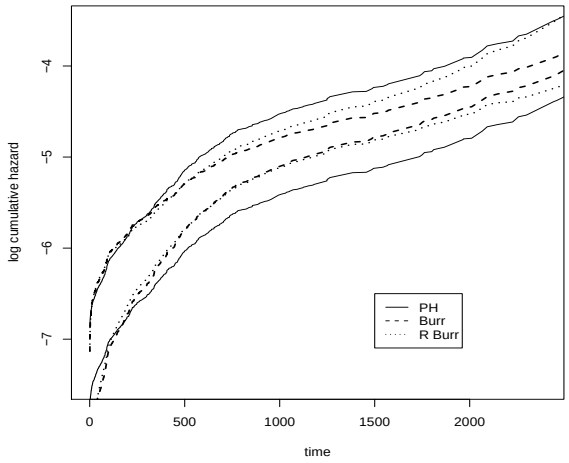

Figure 4.5: Log cumulative hazard functions of younger (lower lines) versus older (upper lines) patients for AML data under the three different models

Burr model, while, as expected, the Cox model underestimates the effects.

The estimated θ coefficients were 0.455, -0.426 and 0.299 for the three different time splines. A plot of $\hat{F}(t|\theta)$ is given in Figure 4.4. The plot reveals a flexible behavior of the time functions, as the line reaches a maximum after approximately 1000 days and from then on it drops towards zero. The hazards estimated for patients aged 40 years old versus 70 years old, with all other characteristics equal to their mean value, show an interesting pattern (Figure 4.5) . The relaxed Burr model gives very similar hazards to the Burr model up to 1000 days and from that point on it starts to divert towards the proportional hazards model. We should point out that the graphs present only what happens up to the sixth year of the follow up (2300 days), since there are only a few cases left during the last years of the study and thus inference for such a small number of events cannot be very reliable.

A plot of survival curves for different covariate patients (4.6), moving from patients with good prognosis (age=19, sex=0, wbc=0 and dep=0) to cases with worse prognosis (age=75,sex=0, wbc=30, dep=2), illustrates how the two models behave. As before the differences are hardly visible. In most covariate

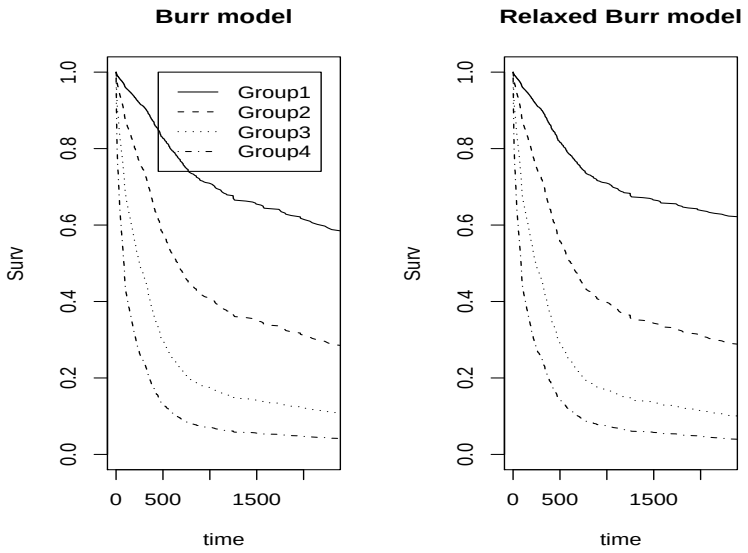

Figure 4.6: Survival of patients with AML moving from cases with good prognosis (group1) to worse prognosis (group4) under the Burr and the relaxed Burr model.

patterns the relaxed Burr and the Burr model give almost identical estimates for survival. A few differences among the two models can be seen towards the end of the follow up for the best prognosis cases.

4.6 Discussion

Frailty models are very useful for modelling individual heterogeneity and can offer an insight on some aspects of the data. However, they are far from perfect. The choice of the frailty distribution is mainly due to mathematical convenience. Based on that fact, and assuming that mathematical convenience is an important factor for the choice of the model it can be seen that a relaxed Burr model will be preferred to a Burr model. The new model described in this work is based on a very simple approximation of the Burr model but it was shown to approximate the latter in a very reliable way. The estimates of the

covariate effects in the two applications were almost identical with regard to the coefficient estimates and their standard errors, and it was also very hard to tell the difference among the different survival curves. However, the differences can be seen when comparing the hazards.

The relaxed Burr model can stand as an approximation somewhere between a Burr and Cox proportional hazards model. It is the flexibility that this model provides that can give it a place as an alternative modelling strategy. The fact that a single model can handle both converging and non-converging hazards is an important merit of the method. In this work we have illustrated how, depending on the pattern of the time functions, hazards of different patients can initially converge and then diverge towards proportionality. Although the coefficient estimates in our examples were very similar, the relaxed Burr model provided additional information to the nature of the frailty term. In the AML data, the estimates given by the Burr and relaxed Burr model where very close, however when plotting the hazard functions (4.5) the differences are clearly visible.

However, there is a limitation that the model does not provide an actual estimate of the frailty variance. Nonetheless, by plotting $F(t|\theta)$ against time, the model provides an insight to the unobserved heterogeneity hidden in the data. In this work we fitted the Burr model and then compared the estimated $\xi\Lambda_0(t)$ from the Burr model with $F(t|\theta)$ from the relaxed Burr model. In practice, one could plot $F(t|\theta)$ against $\hat{\Lambda}_0(t)$ estimated from the relaxed Burr model and get an impression of the frailty variance.

In a series of simulated data with time dependent correlated frailties we have shown that a relaxed Burr model performs better in finding the real effects of the covariates when the variance of the random effect is large. The model was able to estimate the true effects with small bias, regardless of the autocorrelation pattern of the random effects. On the contrary, as the correlation of the random effects dropped, the Burr model tends to underestimate the covariate effects. With small frailty variance the relaxed Burr model has more difficulty in estimating covariate effects, as one should expect.

We have shown that the Burr model does not perform satisfactorily when the data were simulated to include time varying frailties, in which case, a relaxed Burr model is preferred. However, the simple Burr model may not fit well when the frailty distribution, in this case, the gamma distribution, is not the appropriate one to fit the data. In such a situation a different frailty distribution may be chosen that can be more appropriate. Our relaxed Burr model makes no distributional assumptions about the frailty term, nonetheless

it arises as a relaxation of a model that is based on a gamma frailty distribution. It is a matter of ongoing research and simulation studies, to address how the model behaves when the frailty arises from other distributions.

A natural extension would be to allow for hazards diverging out of the limits imposed by a proportional hazards model. This can be achieved by allowing the coefficients of the covariate effects to differ on the numerator and the denominator in equation (4.3). An interesting point raised by a referee is whether a relaxed Burr model could be used for multivariate data. Stochastic frailty can be used to model dependent multivariate data as in [108]. For example in twin studies, the hazards of each pair are correlated, and an event in one sibling may influence the hazard of the other. In such a case the frailty will be shared among siblings and may change in case of an event in one of them. Approximations from model (4.2) and the relaxation from (4.3) could, in principle, be used to extend and relax the correlated gamma frailty model for this situation. However, we made no attempt to do so.

4.7 Appendix

To demonstrate the plausibility of equation (4.2) we have that to show that

$$E[Z(t)|T \geq t] = 1/(1+\exp(X\beta)\int_0^t C(s,t)ds)$$

First define:

$$Q(t) = \int_0^t Z(s)\lambda_0(s)ds$$

Then, the covariance of $Q(t)$ and $Z(t)$ is given as

$$cov(Q(t),Z(t)) = \int_0^t C(s,t)\lambda_0(s)ds$$

Now $Q(t)$ can be written as a linear model

$$Q(t) = \alpha Z(t) + R(t)$$

with

$$\alpha = cov(Q(t),Z(t))/var(Z(t))$$

Then $R(t)$ is uncorrelated with $Z(t)$. If $R(t)$ is independent of $Z(t)$, it is easy to show that

$$\begin{aligned} E[Z(t)|T \geq t] &= E[e^{-\exp(X\beta)Q(t)}Z(t)]/E[e^{-\exp(X\beta)Q(t)}] \\ &= E[e^{-\exp(X\beta)\alpha Z(t)}Z(t)]/E[e^{-\exp(X\beta)\alpha Z(t)}] \end{aligned}$$

If Z(t) has a gamma-distribution, standard calculations show that the latter equation leads to equation (4.2). So, we have shown the plausibility of equation (4.2), if $Z(t)$ has a gamma-distribution and the uncorrelated residual $R(t)$ is independent of $Z(t)$. It can also be shown that equation (4.2) holds if the variances of $Q(t)$ and $Z(t)$ are small, irrespective of their actual distribution.