Log-skew-normal accelerated failure time models

Andrea Callegaro
Department of Statistical Sciences
University of Padua
Italy

Abstract: Accelerated Failure Time models (AFT) are useful alternatives to the Cox model in survival analysis. Recently, AFT models for multivariate data have been considered with flexible distributions of the error term. In this paper, we focus on AFT models with flexible distributions of random effects. In particular, we consider multivariate skew-normally distributed random effects. When the sample size is large, flexible distribution of the random effects provides a better description of the dependence structure on the data. The performance of the model is evaluated by simulations. Further, the proposed log-skew-normal AFT model is illustrated with data on multiple myeloma patients with autologous transplantation from the European Bone Marrow Transplantation Registry.

Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI SCIENZE
STATISTICHE

Final version (2012-03-01)

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168

fax: +39 049 8274170

http://www.stat.unipd.it

Corresponding author:
Andrea Callegaro
tel: +39 049 827 4168
andrea.callegaro@unipd.it
http://www.stat.unipd.it/~name

Log-skew-normal accelerated failure time models

Andrea Callegaro
Department of Statistical Sciences
University of Padua
Italy

Abstract: Accelerated Failure Time models (AFT) are useful alternatives to the Cox model in survival analysis. Recently, AFT models for multivariate data have been considered with flexible distributions of the error term. In this paper, we focus on AFT models with flexible distributions of random effects. In particular, we consider multivariate skew-normally distributed random effects. When the sample size is large, flexible distribution of the random effects provides a better description of the dependence structure on the data. The performance of the model is evaluated by simulations. Further, the proposed log-skew-normal AFT model is illustrated with data on multiple myeloma patients with autologous transplantation from the European Bone Marrow Transplantation Registry.

The Cox proportional hazards (PH) model (Cox, 1972) is now the most widely used method for the analysis of survival data in the presence of covariates. On the other hand, the Accelerated Failure Time model (AFT) has seldom been utilized in the analysis of censored survival data. However, the AFT model is a useful alternative to the Cox regression model in survival analysis for several reasons. First, results have an intuitive physical interpretation (Wei, 1992). Second, fixed effect parameters are robust toward neglected covariates (Hougaard, 1999). Finally, AFT models should lead to more efficient parameter estimates than the Cox model under certain circumstances (Cox, 1984).

AFT models can be rewritten specifying a direct relation between the logarithm of the survival time and the explanatory variables, just as a multiple linear regression model does. A parametric AFT model assumes that the error term of this linear regression follows a distribution of a specific type (e.g., normal, logistic or Gumbel). In contrast, semi-parametric procedures for the AFT model leave the density of the error term unspecified and provide only the estimate of the regression parameters (Kalbfleish and Prentice, 2002). Evidently, parametric AFT models can fit badly in the case of an incorrect specification of the distribution. Estimates of the regression parameters (with the exception of the intercept and the scale parameter) are consistent, but some loss in efficiency is due to an incorrect assumption about the error. In order to achieve more robustness, it is possible to considered a more flexible parametric model, such as the generalized gamma, or the generalized F.

In the first part of this paper we consider parametric AFT models with flexible

distribution of the error term given by the skew-normal (SN) distribution (Azzalini, 1985). The SN distribution is an extension of the normal distribution allowing for non-zero skewness by means of a shape parameter. In recent years, this distribution, in its univariate and multivariate version, has received considerable attention both in theoretical studies, for its numerous stochastic properties, and in applied studies, for the additional flexibility that it provides (Azzalini, 2005).

In the second part of the paper, multivariate survival data are considered. AFT models have been widely used in univariate analysis (Kalbfleish and Prentice, 2002), while multivariate life-time models have received less attention. There has been a strong emphasis on semiparametric Cox PH models with random effects (frailties) (Clayton, 1978; Yashin et al., 1995; McGilchrist & Aisbett, 1991), but there has been little discussion of parametric families. Anderson et al. (1995) proposed a model for bivariate survival data with a parametric or non-parametric frailty distribution. Klein et al. (1999) derived a marginal-likelihood approach from a multivariate normal regression model, and Pan (2001) considered AFT models with gamma frailty. More recently, in order to improve the fit, Komárek et al. (2005), Komárek & Lesaffre (2006) and Lambert et al. (2004) have focused on multivariate AFT models with flexible distributions of the error term. In this paper, instead, we focus on flexible distributions of the random effects. We assume that the random effects are multivariate skew-normal distributed (Azzalini & Dalla Valle, 1996; Azzalini & Capitanio, 1999; Azzalini, 2005) and we show that flexible random effects models are of crucial importance to describe the dependence structure present in the data.

The paper is organized as follows. Sections 2.1 and 2.2 introduce the skew-normal distribution and the AFT models, respectively. In Section 2.3 we derive a univariate model with skew-normal error terms. In Section 2.4 a multivariate AFT model with skew-normal random effects is proposed. The marginal likelihood is derived for the case when the error term is normal. An approximated likelihood approach is proposed for maximum likelihood estimation. Section 3.1 presents the simulation results. In Section 3.2 the proposed new method is illustrated by an application to a real data set of multiple myeloma patients with autologous transplantation from the European Bone Marrow Transplantation Registry. Finally, Section 4 presents a discussion of the proposed models.

1   Methods

1.1   Skew-normal distribution

The skew-normal (SN) distribution is an extension of the normal distribution to allow for non-zero skewness (Azzalini, 1985). In the one-dimensional case, the density with location ξ, scale ω and shape parameter α ($\xi, \alpha \in R, \omega \in R^+$) is given by

$$f(x, \xi, \omega^2, \alpha) = 2\omega^{-1}\phi(\varpi)\Phi(\alpha\varpi), \tag{1}$$

where $\varpi = (x - \xi)/\omega$, ϕ and Φ denote the N(0,1) density and distribution function, respectively. When $\alpha = 0$, the density corresponds to the normal density. The distribution is right skewed if $\alpha > 0$ and it is left skewed if $\alpha < 0$. An alternative parametrization of the SN distribution can be derived in terms of the mean μ,

the standard deviation σ and the Pearson's index of skewness γ_1 of X. We denote $(\mu, \sigma^2, \gamma_1)$ as the "centred parameters" (CP) while the parameters (ξ, Ω, α) are called direct (DP) because they can be read directly from (1). We will adopt the notation $X \sim SN_{DP}(\xi, \omega^2, \alpha)$ if the density is expressed in terms of the direct parameters, and $X \sim SN_{CP}(\mu, \sigma^2, \gamma_1)$ if the density is expressed in terms of the centred parameters.

The multivariate SN distribution with location ξ, scale Ω, and shape parameter α is given by Azzalini & Capitanio (1999):

$$f(x, \xi, \Omega, \alpha) = 2\phi_d(x - \xi; \Omega)\Phi(\alpha^T \omega^{-1}(x - \xi)), \tag{2}$$

where $x \in R^d$, α is the shape parameter $(\alpha \in R^d)$, and ω is the diagonal matrix formed by the standard deviations of Ω. If a d-dimensional variable X has density (2), we say that its distribution is multivariate SN and write $Y \sim SN_{DP}(\xi, \Omega, \alpha)$. The distribution can be expressed in terms of the centered parametrization $Y \sim SN_{CP}(\mu, \Sigma, \gamma_1)$ where μ is the mean, Σ is the variance-covariance matrix and γ_1 is the multivariate index of skewness of X. For statistical inference it is known that centred parametrization provides a more convenient parametrization than the direct parametrization which is commonly employed for writing the density (Arellano et al., 2008).

1.2   Accelerated Failure Time (AFT) models

The AFT model can be written as being log-linear with respect to time, giving

$$\ln(T_j) = X_j^T \beta + \omega_e e_j, \tag{3}$$

where T_j and X_j denote the event time and the vector of covariates measured for the j-th individual, respectively. ω_e denotes the scale parameter and e_j the random error term for the j-th individual. The interpretation of model (3) is straightforward: the survival times are multiplied by a constant effect, and the exponentiated coefficients, $\exp(\beta_k)$, represents time ratios. The interpretation of the parameters differs crucially from the interpretation of the parameters estimated by the proportional hazard model, where $\exp(\beta_k)$ represents hazard ratios. Only in the particular case that the survival times follow a Weibull distribution, it can be shown that the AFT and proportional hazard models are the same. The error terms are usually assumed to follow a specific distributional form (normal, logistic, or Gumbel). Alternative methods include semiparametric AFT models, in which the distribution of the error term is estimated nonparametrically (Kalbfleish and Prentice, 2002). The log-likelihood function may be written as

$$\ell(\beta, \omega_e) = \sum_{j=1}^{n} d_j(-\ln \omega_e + \ln f_e(\varpi_j)) + (1 - d_j) \ln S_e(\varpi_j), \tag{4}$$

where d_j is the indicator of whether the j-th failure time is observed $(d_j = 1)$ or is right censored $(d_j = 0)$, $\varpi_j = (\ln t_j - X_j\beta)/\omega_e$, and $S_e(x) = \int_x^\infty f_e(u)\mathrm{d}u$ is the survival function of the error term.

For multivariate data, the accelerated failure time models have been extended to include random effects

$$\ln(T_{ij}) = X_{ij}^T \beta + Z_{ij}^T b_i + \omega_e e_{ij}, \tag{5}$$

with T_{ij} the event time for subject j from cluster i, X_{ij} the vector of covariates for subject j from cluster i, β the vector containing the covariate effects, b_i the random effect for cluster i, Z_{ij} the known design vector for the random effects, ω_e the scale parameter, and finally e_{ij} the random error term for subject j from cluster i. Different distributional assumptions have been proposed for the random effects and for the error term (Klein et al., 1999; Pan, 2001; Chang, 2004; Komárek et al., 2005; Komárek & Lesaffre, 2006; Lambert et al., 2004).

1.3 Univariate log-skew-normal AFT

Let us consider the univariate AFT model (3) where the error term e_j follows a skew-normal distribution with null location, unit scale, and shape parameter α_e, $e_j \sim SN_{DP}(0, 1, \alpha_e)$, then

$$Y_j = \ln T_j \sim SN_{DP}(X_j^T \beta, \omega_e^2, \alpha_e). \tag{6}$$

The density and the survival function of the error term are $f_e(\varpi) = 2\phi(\varpi)\Phi(\alpha_e \varpi)$ and $S_e(\varpi) = 1 - \Phi(\varpi) + 2O_T(\varpi, \alpha_e)$, respectively, where O_T is the Owen's T function and $\varpi = (\ln t - X^T \beta)/\omega_e$. The corresponding hazard rate is given by $\lambda(t, \beta, \omega_e, \alpha_e) = f_e(\varpi)/(t\omega_e S_e(\varpi))$. Figure 1 shows the hazard function of the log-skew-normal distribution with different values of skewness parameter. As expected, the additional skewness parameter changes the shape of the hazard function considerably.

The log-likelihood is given by equation (4). Maximum likelihood methods can be employed to estimate the parameters of the distribution. A peculiar aspect of the skew-normal log-likelihood function is that, for small/moderate sample size, it can be monotone in α_e. To solve this problem we propose a penalize likelihood approach (see Appendix A for details). The performance of the penalized approach will be evaluated by simulations.

An alternative univariate model can be obtained assuming a null mean error term $e_j \sim SN_{CP}(0, 1, \gamma_{1e})$. An advantage of this null mean model with respect to the null location model of equation (6) is that the parameters β and ω_e have the same interpretation as the parameters estimated by the classical log-normal AFT model.

1.4 Multivariate log-skew-normal AFT

Let us consider the multivariate AFT model (5) where the random effect b_i is multivariate skew-normal distributed $b_i \sim SN_{CP}(0, \Sigma_b, \gamma_{1,b})$. The log-likelihood of the i-th cluster conditional on the random effects b_i can be expressed by

$$\ell_i^C(\beta, \omega_e | b_i) = \sum_{j=1}^{n_i} d_{ij}[-\log(\omega_e t_{ij}) + \log f_e(\varpi_{ij})] + (1 - d_{ij})\log S_e(\varpi_{ij}),$$

where $\varpi_{ij} = \frac{\log(t_{ij}) - X_{ij}\beta - Z_{ij}^T b_i}{\omega_e}$ and $S_e(x) = \int_x^\infty f_e(u)\mathrm{d}u$. The marginal likelihood is then

$$L(\theta) = \prod_{i=1}^{G} \int \exp\{\ell_i^C(\beta, \omega_e | b_i) + \log f_b(b_i)\}\mathrm{d}b_i. \qquad (7)$$

Note that different distributions of the error term (extreme value, normal, skew-normal, logistic) can be combined with skew normal random effects. For example, if the error term follows a standard type I extreme value (Gumbel) distribution, then $f_e(x) = \exp(x - \exp(x))$ and $S_e(x) = \exp(-\exp(x))$. If the error term follows a logistic distribution, then $f_e(x) = \exp(x)(1 + \exp(x))^{-2}$ and $S_e(x) = (1 + \exp(x))^{-1}$.

If the error term is normally distributed: $e_{ij} \sim N(0,1)$, then $f_e(x) = \phi(x)$ and $S_e(x) = 1 - \Phi(x)$. In this case Y follows a skew normal distribution: $Y \sim SN_{DP,n}(\xi, \Omega, \alpha)$. The contribution to the likelihood of a group where all failure times are observed ($n = n_D$) is the joint density function of equation (2). If there are any censored observations among the n individuals, then rewrite $Y = (Y_C, Y_D)$, where Y_D is the vector of the (log-transformed) failure times and Y_D is the vector of the censored failure times. The contribution to the marginal likelihood is the product of the (marginal) density of Y_D times the conditional survival function of Y_C given Y_D evaluated at y. In this case, the contribution to the likelihood for this group is given by

$$L(\theta) = \phi_{n_D}(y_D - \xi_D, \Omega_D) \int_{\varpi_{C \cdot D}^*}^\infty \phi_{n_C+1}(u, \bar{\Omega}_{C \cdot D}^*)\mathrm{d}u.$$

All of the parameters $(\xi, \Omega, \alpha, \xi_D, \Omega_D, \varpi_{C \cdot D}^*, \bar{\Omega}_{C \cdot D}^*)$ are derived in Appendix B . The derived marginal likelihood is an extension of the marginal likelihood proposed by Klein et al. (1999) for normal random effects and normal error terms. In appendix B the marginal likelihood is also derived in the case that random effects are Gaussian and the error terms are skew normally distributed.

If the error term is not normal, then it is more complex to derive the marginal likelihood (7). It is, however, possible to approximate the likelihood by numerical methods or Monte Carlo simulation. In this paper, we use a Gaussian quadrature method with 100 nodes. The parameters are estimated maximizing the approximated likelihood. In order to speed up the estimation we propose a two-step procedure. In the first step of the procedure, all of the parameters, with the exception of the skewness parameters, are estimated assuming normal distributed random effects. In the second step, the approximated marginal likelihood is maximized with respect to the skewness parameters, with all the other parameters fixed at the values estimated at the first step. Standard software can be used to estimate the parameters at the first step. In particular, penalized likelihood approaches (Therneau et al., 2003) implemented in the R package are particularly fast and efficient. The performance of the two-step (fast) procedure will be evaluated by simulations. Standard errors for the estimates will be computed by inverting the Hessian matrix of the approximated likelihood.

For prediction of the random effects, it is possible to use the conditional expec-

tations of the random effects

$$E\{b_i\} = \frac{\int b_i \exp[\ell_i^C + \log f_b(b_i)]\mathrm{d}b_i}{\int \exp[\ell_i^C + \log f_b(b_i)]\mathrm{d}b_i}.$$

The integral can be approximated by numerical methods or Monte Carlo simulations. Prediction is beyond the scope of this paper and will not be further considered.

1.4.1 Shared AFT model

In this section we study the relation between the skewness of the random effects and the dependence structure of the data. To simplify the exposition, we consider the bivariate case ($n_j = 2$) with shared random effects:

$$\ln(T_{ij}) = Y_{ij} = X_{ij}^T\beta + b_i + \omega_e e_{ij}, \tag{8}$$

where $b_i \sim SN_{CP}(0, \sigma_b^2, \gamma_{1,b})$ is the null mean random effect with standard deviation σ_b and the Pearson index of skewness is given by $\gamma_{1,b}$, $i = 1, ..., G$ and $j = 1, 2$.

If the error term is normal, $e_{ij} \sim N(0, 1)$, then parameters are easily interpretable. In fact, $E\{Y_{ij}\} = X_{ij}^T\beta$; $\mathrm{var}\{Y_{ij}\} = \sigma_b^2 + \omega_e^2$, and $\mathrm{cov}\{Y_{ik}, Y_{il}\} = \sigma_b^2$, $k \neq l$. The correlation between the log-survival times of any two members of a group is $\rho = \sigma_b^2/(\sigma_b^2 + \omega_e^2)$.

A classical measure of local dependence is given by the cross-ratio function (Clayton, 1978). The cross-ratio function can be expressed as the ratio of two hazards, i.e., the ratio of the hazard of subject j given that subject l "died" at t_l and the hazard of j given that $T_l > t_l$.

$$CR(t_1, t_2) = \frac{\lambda(t_1|T_2 = t_2)}{\lambda(t_1|T_2 > t_2)} = \frac{S(t_1, t_2)D_1D_2S(t_1, t_2)}{D_1S(t_1, t_2)D_2S(t_1, t_2)}, \tag{9}$$

where $D_j = \frac{\partial}{\partial t_j}$ and, in an obvious notation,

$$S(t_1, t_2) = \int_{-\infty}^{\infty} S_e(\varpi_1)S_e(\varpi_2)f_b(u)\mathrm{d}u.$$

denotes the marginal bivariate survival function, with $\varpi_j = (y_j - X_j^T\beta - u)/\omega_e$. In our case, the marginal survival function does not have a closed form, and we use Gaussian quadrature to approximate it.

Figure 2 shows the marginal survival function and the cross-ratio function $CR(t, t)$ for the bivariate log-skew-normal AFT model (8) with three different values of the skewness parameter $\gamma_{1,b} = -0.9$, 0, and 0.9, respectively. Figure 2(a) shows the marginal survival function of the log-normal AFT model with normal error terms, $\beta = 0$, $\sigma_b = 0.5$, and $\omega_e = 1$. Figure 2(b) shows the cross-ration function of the model with normal error terms. If $\gamma_{1,b}$ is null, this model corresponds to the multivariate log-normal AFT model proposed by Klein et al. (1999). The dependence induced by this model is a function that decreases over time. When $\gamma_{1,b}$ is negative (positive), the dependence decreases over time at a smaller (higher) rate with respect to the log-normal model. Figure 2(c) shows the cross-ratio function of the

model with logistic error terms. The dependence induced by this model is qualitatively similar to the dependence structure induced by the model with normal error terms. Finally, Figure 2(d) shows the cross-ratio function of the model with extreme value distributed error terms. Interestingly, in this case the skew parameter has a strong effect on the dependence structure. When the skew parameter is negative, the cross-ratio function increases at the beginning of the follow-up, and then it decreases slowly over time.

2   Results

2.1   Simulation study

Two numerical studies, based upon 500 replications of simulated data, are presented to evaluate the performance of the proposed estimation procedures.

First, we consider the univariate case. The log-transformed survival times Y_j were generated using a skew-normal distribution $SN_{CP}(\mu_j, \omega_e^2, \gamma_{1,e})$ with $\mu_j = \beta_0 + \beta_1 x_j$, $\beta_0 = 5$, $\beta_1 = 0.3$ and $x_j \sim N(0,1)$. The corresponding censoring times C_j were generated from a uniform distribution between infinity and a lower limit determined in order to achieve approximately the right censoring rate around 30%. Data were simulated with $\omega_e = 1$ and skewness parameters $\gamma_{1,e}$ equal to -0.9, 0 or 0.9. We considered a sample size of $n = 200$. Table 1 shows the results of the univariate log-skew-normal AFT method with penalized likelihood (as described in Appendix A). The penalty avoids infinity estimates of the shape parameter, with a very small bias. As a comparison, we also fitted the classical log-normal AFT model using `survreg` in the survival package in R. As expected, when the shape parameter differs from zero, the proposed model outperforms the log-normal AFT model.

Secondly, we investigate the performance of the multivariate skew-normal AFT model. The log-transformed survival times Y_{ij}, conditioned on the random intercept b_i, were generated using a normal distribution $N(\mu_{ij} + b_i, \omega_e^2)$ with $\mu_{ij} = \beta_0 + \beta_1 x_{ij}$, $\beta_0 = 5$, $\beta_1 = 0.3$, $x_{ij} \sim N(0,1)$. The random intercept was skew-normally distributed $b_i \sim SN(0, \sigma_b^2, \gamma_{1,b})$ with $\sigma_b = 0.5$ and skew parameters $\gamma_{1,b} = -0.9$, 0, or 0.9. We considered $G = 200$ clusters with $n_i = 10$ subjects in each cluster. The corresponding censoring times C_{ij} were generated from a uniform distribution between infinity and a lower limit determined in order to achieve approximately the right censoring rate around 30%. Table 2 shows the parameters estimated by the two-step (fast) version of the maximum likelihood estimation approach. In the first step we estimated $(\tilde{\beta}, \tilde{\omega}_e, \tilde{\sigma}_b)$ using `survreg` in the R survival package with Gaussian frailties and Gaussian errors. In the second step, $\gamma_{1,b}$ was estimated by maximizing the approximated marginal likelihood (7) with fixed $(\tilde{\beta}, \tilde{\omega}_e, \tilde{\sigma}_b)$. Integrals were computed by means of Gaussian quadrature with 100 nodes. First, note that the parameters estimated by the penalized likelihood (Therneau et al., 2003) approach are slightly biased. In particular, ω_e is underestimated, irrespectively of the skew parameter. A different method (software) could be used in the first step. Second, note that the estimated skew-parameter is biased toward zero however, even in the presence of this bias, the power to detect skewed random effects at the level of 5% is about 80%. Finally, note that with this sample size the approximated log-likelihood

is not monotone in $\gamma_{1,b}$ (data not shown).

2.2 Application to multiple myeloma patients with autologous transplantation

We applied the proposed skew-normal AFT model to a real data set from the European Group for Blood and Marrow Transplantation (EBMT) registry. Multiple Myeloma (MM) patients with autologous transplantation were considered. The original MM autologous registry includes about 52,000 cases. As an illustration, in this analysis we considered only a subset of 3,081 patients i) with common myelomas (IgA, IgG), ii) transplanted after 1998, and iii) without inconsistent records or missing values in a list of relevant prognostic factors. Since missing data may be centre/patient-related, with potential patterns of selection bias, this application is not meant to reach conclusions relevant to clinical knowledge or practice.

Patients are clustered into $G = 190$ centres with a mean of 16 individuals per cluster (median 7, minimum=1, maximum=134). We studied Progression Free Survival (PFS), where the failure time is defined as the time from autologous transplantation to any of the three following events: a) disease progression, b) disease relapse/recurrence (for patients transplanted in Complete Remission (CR) or achieving CR later on), and c) death without relapse/progression. The number of observed events is 1,548 (51%), and the median PFS time is 14.85 months (minimum=0.03, maximum=119.18). We considered the following risk factors: year of the transplant (from 1999 to 2009), gender (female 42% versus male 58%), stage at diagnosis (I 16%, II 26%, and III 58%), disease status at conditioning (no complete remission 89% versus complete remission 11%), serum beta(2)-microglobulin level at diagnosis ($31\% > 4$ versus $69\% \leq 4$), the interval between diagnosis and the first transplant in years (median 0.6, minimum=0.003, maximum=17.8), and age at transplant (median 58.4, minimum=20, maximum=78.9).

Initially, different models were fit without random effects, in order to identify the distribution of the error term. The failure times were modelled using Weibull, log-logistic, log-normal and log-skew-normal distributions. No skewness was detected by the log-skew-normal AFT model, and the log-likelihood value of the Weibull, log-normal, and log-logistic AFT model were -3204.4, -3191.9, and -3163.7, respectively. Based on the AIC criterion, we selected the log-logistic model. To check the model adequacy, we stratified the linear predictor values $X_j^T \hat{\beta}$ into seven categories. The estimated survivor functions were computed for each patient and averaged within each stratum. Figure 3(a) shows the fitted survival functions and the Kaplan-Meier estimates for the first, fourth, and seventh strata of the skew-normal AFT model. The data are fit reasonably well by the considered model, especially before 30 months of follow-up, where most of the observed events are concentrated.

Secondly, we performed a multivariate analysis where centre effects were modelled through a shared random effect. We fitted a shared AFT model with skew-normal random effects and logistic error terms (see the third column of Table 3). The estimated skew parameter is -0.78 (se= 0.21). Both Wald type and likelihood ratio tests indicates that the centre effect is negatively skewed. Further, we fitted the penalized method of Therneau et al. (2003) with logistic error term and log-normal

(penalized log-likelihood value= -7442.25) or gamma distributed frailties (penalized log-likelihood value= -7418.25) (see the first two columns of Table 3). The great difference in the values of the two penalized likelihoods confirms what was detected by the proposed model. Further, it is interesting to note that the frailty distribution has little or no effect on the estimates of the fixed effects (see Table 3).

Finally, we extended the model by including the country effect. Centres are clustered into 26 countries, with a mean of 118 centres per country (median 30, minimum=1, maximum=690). The log-transformed PFS time was modelled by

$$Y_{kij} = X_{kij}^T \beta + a_{ki} + c_k + \omega_e e_{kij}, \tag{10}$$

where the centre random effect $a_{ki} \sim SN_{CP}(0, \sigma_a, \gamma_{1,a})$ was independent of the country random effect $c_k \sim SN_{CP}(0, \sigma_c, \gamma_{1,c})$, $k = 1, ..., 26$. The error term e_{kij} was assumed to be logistic distributed. Note that model (10) is slightly different from the general model discussed in Section 2.4. In fact, the sum of two independent skew-normal random variables is not generally skew-normal. The marginal likelihood was approximated using Gaussian quadrature, with 100 nodes for each random effect. The estimated parameters of the random effects are $(\hat{\sigma}_c, \hat{\gamma}_{1,c}) = (0.27, -0.99)$ and $(\hat{\sigma}_a, \hat{\gamma}_{1,a}) = (0.19, -0.60)$. The other estimated parameters are substantially unchanged with respect to the model without country effects. Both random effects are significant and left skewed. Figure 3(b) shows the approximated profile log-likelihood of the shape parameter $\alpha_{1,c}$ (which is just a reparametrization of $\gamma_{1,c}$). It is evident from the profile log-likelihood that the shape parameter is significantly different from zero, so the dependence due to the country effect is slowly decreasing over time, as shown in Figure 2(c).

3   Discussion

In this paper, we have described accelerated failure time models with skew-normal distributed (Azzalini, 1985) random effects and/or error terms.

For univariate data, the model is an extension of the lognormal AFT model, with an additional parameter to allow for non-zero skewness of the error term. In order to prevent bias of the shape parameter, we proposed a penalized likelihood approach. Simulation results show that, in the presence of skewed error terms, the proposed model outperforms the classical log-normal AFT model.

For multivariate data, we considered log-linear regression models with multivariate skew-normal (Azzalini & Dalla Valle, 1996) random effects and general distribution of the error terms. We derived the marginal likelihood for the model with normal error terms and we studied the effect of the skewness on bivariate survival times with shared random effects. We showed that the skewness parameter qualitatively changes the cross-ratio function. It follows that the proposed model is useful to better understand the dependence structure between multivariate failure times. To estimate the parameters, we proposed to maximize the marginal likelihood approximated through Gaussian quadrature. In order to speed up the estimation procedure we developed a two-step procedure where at the first step parameters are estimated assuming Gaussian random effects. Simulation results showed good

performance of this approach. As illustration, we analyzed a real data of multiple myeloma patients with autologous transplantation. We fitted a hierarchical AFT model with logistic error terms and skew normal centres nested within skew normal countries. Both random effects were negatively skewed, confirming the usefulness of the proposed model.

Recent literature (Komárek et al., 2005; Komárek & Lesaffre, 2006; Lambert et al., 2004) has emphasized the fact that often the choice of the distribution of the random effect is not important. This is true for small/moderate sample sizes. On the other hand, for large samples, AFT models with flexible distribution of the random effects are crucial to arrive at a good description of the dependence structure present in the data.

The R code used in this paper is available upon request to the author.

Appendix

Appendix A: bias correction for the shape parameter α_e

In order to solve the problem of the monotone likelihood in α_e, we propose a penalized likelihood approach

$$\ell(\theta)^* = \ell(\theta) + \frac{1}{2}\ln[-\frac{\partial^2 \ell(\theta)}{\partial \alpha_e^2}],$$

where $\ell(\theta)$ is given by Equation (4) with $f_e(\varpi) = 2\phi(\varpi)\Phi(\alpha_e\varpi)$ and $S_e(\varpi) = 1 - \Phi(\varpi) + 2O_T(\varpi, \alpha_e)$. This approach is similar in spirit to the Firth's method for bias prevention (Firth, 1993) in generalized linear models. Sartori (2006) proposed to use Firth's correction for the shape parameter of the scalar skew-normal distribution.

Appendix B: marginal likelihood derivation when the error term is normal

Let $Y_i = \ln(T_i)$ be a $(n_i \times 1)$ vector of the log-transformed times of the i-th cluster, $i = 1, ..., m$

$$Y_i = X_i\beta + Z_i b_i + e_i,$$

where $b_i \sim SN_{q,DP}(0, \Omega_b, \alpha_b)$ and $e_i \sim N_n(0, \Omega_{ei})$. For the skew-normal distribution we use the parametrization proposed by Azzalini & Capitanio (1999). It follows that

$$Y_i|b_i \sim N_{n_i,DP}(X_i\beta + Z_i b_i, \Omega_{ei}),$$
$$b_i \sim SN_{q,DP}(0, \Omega_b, \alpha_b).$$

The sum of a multivariate SN variate and an independent multivariate normal variate is $Y_i \sim SN_n(\xi_i, \Omega_i, \alpha_i)$, where $\xi_i = X_i\beta$, $\Omega_i = Z_i\Omega_b Z_i^T + \Omega_{ei}$ and

$$\alpha_i = \frac{\omega_i\Omega_i^{-1}B_i^T\alpha_b}{(1 + \alpha_b^T(\bar{\Omega}_b - B_i\Omega_i^{-1}B_i^T)\alpha_b)^{1/2}}$$
$$B_i = \omega_b^{-1}\Omega_b Z_i^T.$$

(see Appendix C). In Appendix C, α_i is also derived for the case when $b_i \sim N_q(0, \Omega_b)$ and $e_i \sim N_n(0, \Omega_{ei}, \alpha_{ei})$.

In the following, for ease of notation, the subscript i is omitted. Let $d = \sum_{k=1}^{n} d_k$ be the number of observed failure times in the group. If there are censored observations among the n individuals, then rewrite $Y = (Y_D, Y_C)$, where Y_D is the d-dimensional vector of the log death times and Y_C is the $n - d$-dimensional vector of censored observations. The contribution to the likelihood for the group of n observations is the product of the density function for Y_D times the conditional survival function of Y_C, given Y_D evaluated at $y_D = \ln[t_D]$.

First of all, we consider the contribution of the n observed failure times. Their marginal distribution is $Y_D \sim SN_{n_D, DP}(\xi_D, \Omega_{DD}, \alpha_{D(C)})$, so their contribution to the likelihood is

$$L_D = 2\phi_d(y_D - \xi_D; \Omega_{DD})\Phi(\tau_{C \cdot D}),$$

where $\tau_{C \cdot D} = \alpha_{D(C)}^T \omega_D^{-1}(y_D - \xi_D)$, with $\bar{\Omega}_{CC \cdot D} = \bar{\Omega}_{CC} - \bar{\Omega}_{CD}\bar{\Omega}_{DD}^{-1}\bar{\Omega}_{DC}$ and $\alpha_{D(C)} = (\alpha_D + \bar{\Omega}_{DD}^{-1}\bar{\Omega}_{DC}\alpha_C)(1 + \alpha_C^T\bar{\Omega}_{CC \cdot D}\alpha_C)^{-1/2}$, which is the shape parameter of the marginal distribution of Y_D.

Now we consider the contribution to the likelihood of the censored observations. The conditional distribution of $Y_C|Y_D \sim SN_{n-d, DP}(\xi_{C \cdot D}, \Omega_{CC \cdot D}, \alpha_{C \cdot D}, \tau_{C \cdot D})$ is an extension of the classical skew normal distribution with an additional parameter $(\tau_{C \cdot D})$. The multivariate survival function of this conditional distribution is given by Capitanio et al. (2003)

$$L_C = P(Y_C \geq y_C | Y_D = y_D) = \frac{1}{\Phi(\tau_{C \cdot D})} \int_{\varpi_{C \cdot D}^*}^{\infty} \phi_{n-d+1}(u, \bar{\Omega}_{CC \cdot D}^*)\mathrm{d}u,$$

where

$$\bar{\Omega}_{CC \cdot D}^* = \begin{pmatrix} 1 & -\tilde{\delta}_{C \cdot D}^T \\ -\tilde{\delta}_{C \cdot D} & \tilde{\Omega}_{CC \cdot D} \end{pmatrix},$$

$\tilde{\Omega}_{CC \cdot D} = \tilde{\omega}_{C \cdot D}^{-1}\bar{\Omega}_{CC \cdot D}\tilde{\omega}_{C \cdot D}^{-1}$ is the correlation matrix associated to $\bar{\Omega}_{CC \cdot D}$ and

$$\tilde{\delta}_{C \cdot D} = \tilde{\Omega}_{CC \cdot D}\alpha_{C \cdot D}/(1 + \alpha_{C \cdot D}^T\tilde{\Omega}_{CC \cdot D}\alpha_{C \cdot D})^{1/2}$$
$$\varpi_{C \cdot D}^* = (\tau_{C \cdot D}, (y_C - \xi_{C \cdot D})\tilde{\omega}_{C \cdot D}^{-1})^T.$$

The log-likelihood is given by the sum of the contributions of the m independent clusters:

$$\ell = \sum_{i=1}^{m} \log L_{D,i} + \log L_{C,i}.$$

If random effects are assumed to have zero means, then the location parameter of Y_i becomes $\xi_i = X_i^T\beta - \sqrt{2/\pi}\delta$, with $\delta = \bar{\Omega}\alpha(1 + \alpha^T\bar{\Omega}\alpha)^{-1/2}$.

Appendix C: derivation of the multivariate α

The sum of a multivariate SN variate and a multivariate normal variate is SN. This can be easily seen using the moment generating function. The cumulant generating function of the SN $b \sim SN_{q,DP}(0, \Omega_b, \alpha_b)$ is

$$K_b(t_i) = \frac{1}{2}t^T\Omega_b t + \zeta_0(\delta_b^T\omega_b t),$$

where

$$\delta_b = \bar{\Omega}_b \alpha_b (1 + \alpha_b^T \bar{\Omega}_b \alpha_b)^{-1/2}.$$

The cumulant generating function of $e \sim N_n(0, \Omega_e)$ is

$$K_e(t_i) = \frac{1}{2} t^T \Omega_e t.$$

The cumulant generating function of $Y = Zb + e$, is given by

$$K_Y(t_i) = \frac{1}{2} t^T \Omega t + \zeta_0(\delta_b^T \omega_b Z^T t),$$

where $\Omega = Z\Omega_b Z^T + \Omega_e$. It follows that the distribution is still SN with $\delta^T \omega = \delta_b^T \omega_b Z^T$, so $\delta^T = \delta_b^T \omega_b Z^T \omega^{-1}$. Further, using the following relations $\alpha = (1 - \delta^T \bar{\Omega} \delta)^{-1/2} \bar{\Omega}^{-1} \delta$ and $\delta_b = (1 + \alpha_b^T \bar{\Omega}_b \alpha_b)^{-1/2} \bar{\Omega}_b \alpha_b$, then

$$\alpha = \frac{\omega \Omega^{-1} B^T \alpha_b}{(1 + \alpha_b^T (\bar{\Omega}_b - B\Omega^{-1}B^T)\alpha_b)^{1/2}}$$
$$B = \omega_b^{-1} \Omega_b Z^T.$$

If we assume that the residual term is skewed $b \sim N_q(0, \Omega_b)$ and $e \sim SN_{n,DP}(0, \Omega_e, \alpha_e)$, then $Y \sim SN_{n,DP}(\xi, \Omega, \alpha)$, with

$$\alpha = \frac{\omega \Omega^{-1} B^T \alpha_e}{(1 + \alpha_e^T (\bar{\Omega}_e - B\Omega^{-1}B^T)\alpha_e)^{1/2}}$$
$$B = \omega_e^{-1} \Omega_e.$$

References

Anderson, J. E., Louis, T. A. (1995). Survival analysis using a scale change random effects model *Journal of the American Statistical Association*, 90:669 – 679.

Arellano-Valle, R. B., Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution *Journal of Multivariate Analysis*, 99:1362–1382.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32:159–188.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, ser. B*, 61(3):579–602. Full version of the paper at arXiv.org:0911.2093.

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83:715–726.

Balakrishnan, N., Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, 25:2797–2816.

Capitanio, A., Azzalini, A., Stanghellini E. (2003). Graphical models for skew normal variates. *Scandinavian Journal of Statistics*, 30:129–144.

Chang, S. (2004). Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Analysis*, 10:175–190.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiologic studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, B 34:187–220.

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

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Ser. B-Stat. Methodol.*, 39:1–38

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80: 27–38.

Hougaard, P. (1999) Fundamentals of survival data. *Biometrics*, 55:13–22.

Kalbfleish,J. D., and Prentice, R. L. (2002) *The Statistical Analysis of Failure Time Data* 2nd ed. Wiley, New York. 2002, MR1924807

Klein, J. P., Pelz, C., and Zhang, M. (1999) Modeling random effects for censored data by a multivariate normal regression model. *Biometrics*, 55: 497–506.

Komárek, A. and Lesaffre, E. (2006). Bayesian semi-parametric accelerated failure time model for paired doubly interval-censored data. *Statistical Modelling*, 6: 3–22.

Komárek, A., Lesaffre, E. and Hilton, J.F. (2005). Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics*, 14:726–745.

Lambert, P., Collett, D., Kimber, A., and Johnson, R. (2004). Parametric accelerated failure time models with random eects and an application to kidney transplant survival. *Statistics in Medicine*, 23: 3177–3192.

McGilchrist, C. A., and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47:461–466.

Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis*, 7:55–64.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates: skew normal and skew-t distributions. *Journal of Statisitical Planning and Inference*, 136:4259–4275.

Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 156–175. DOI: 10.1198/1061860031365

Wei, L. J. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11:1871–1879.

Yashin, A. I., Vaupel, J. W., and Iachine, I. A. (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical population studies*, 5:145–159.
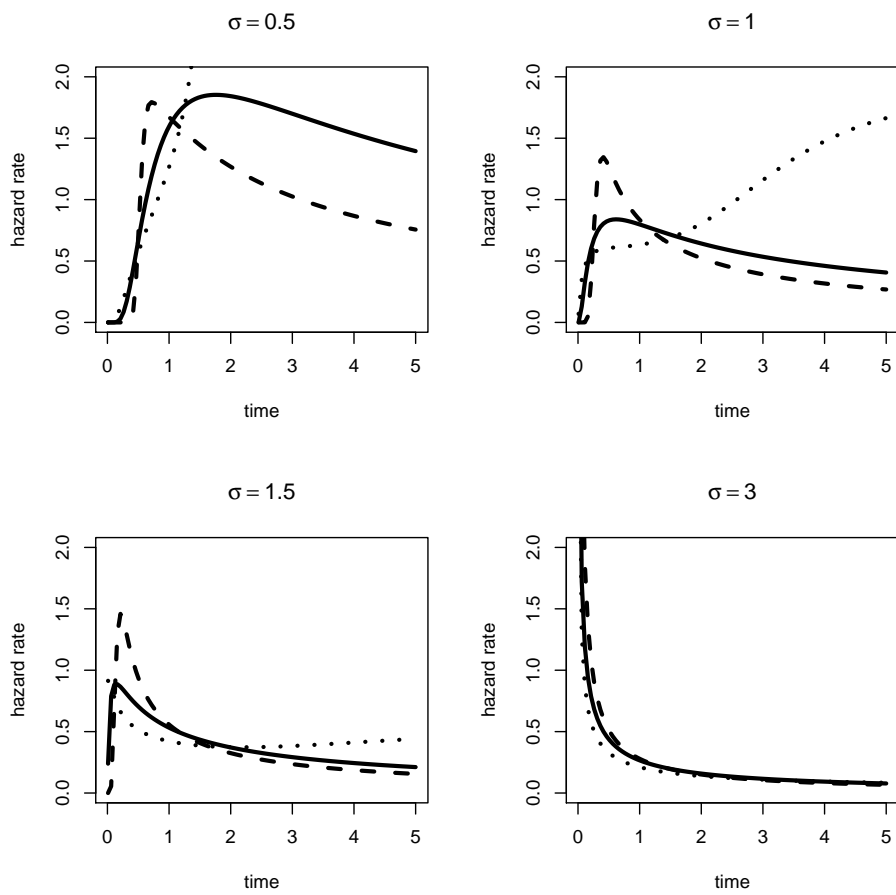
Figure 1: Log-skew-normal hazard function with mean $\mu = 0$, and different values of standard deviation σ and skewness γ_1. Straight, dotted and dashed lines denote the hazard function with $\gamma_1 = 0$, -0.9, and 0.9, respectively.
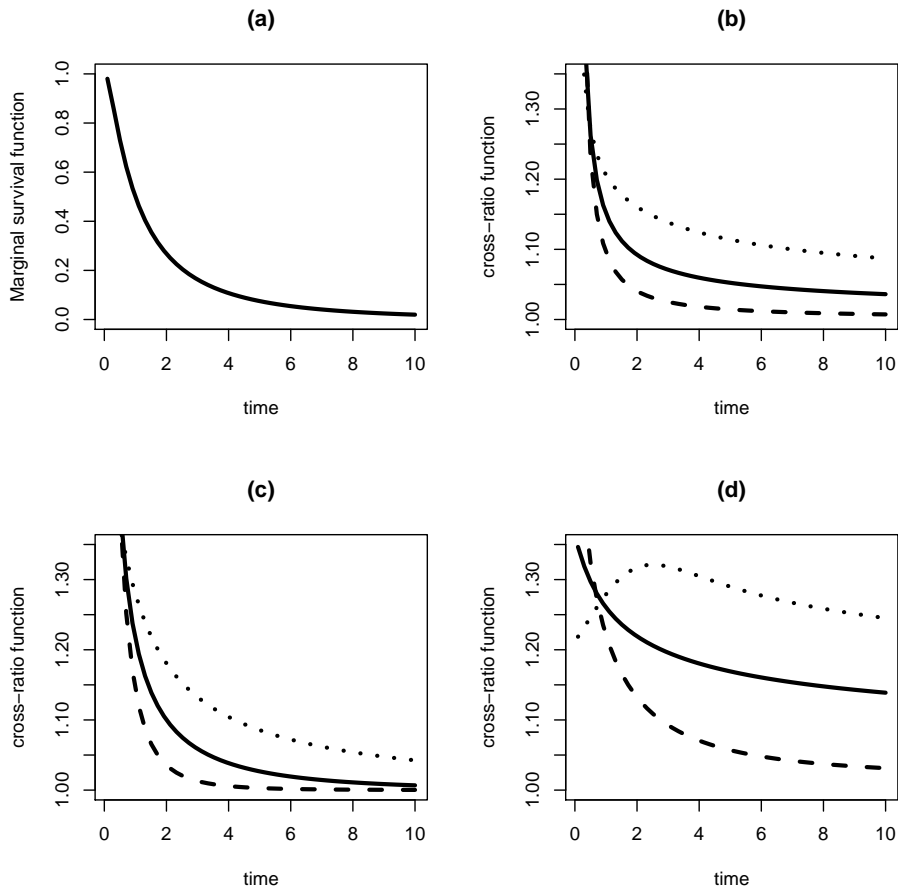
Figure 2: Cross-ratio function of bivariate AFT models with skew-normal random effect and different distributions of the error term. Figure (a): marginal survival function with normal error term and $\beta_0 = (0,0)^T$, $\sigma_b = 0.5$, $\omega_e = 1$. Figures (b), (c), and (d) show the cross-ratio functions $CR(t,t)$ with error term following the normal, the logistic, and the extreme value distribution, respectively. Straight, dotted, and dashed lines denote the cross-ratio function with $\gamma_{1,b} = 0$, -0.9, and 0.9, respectively.
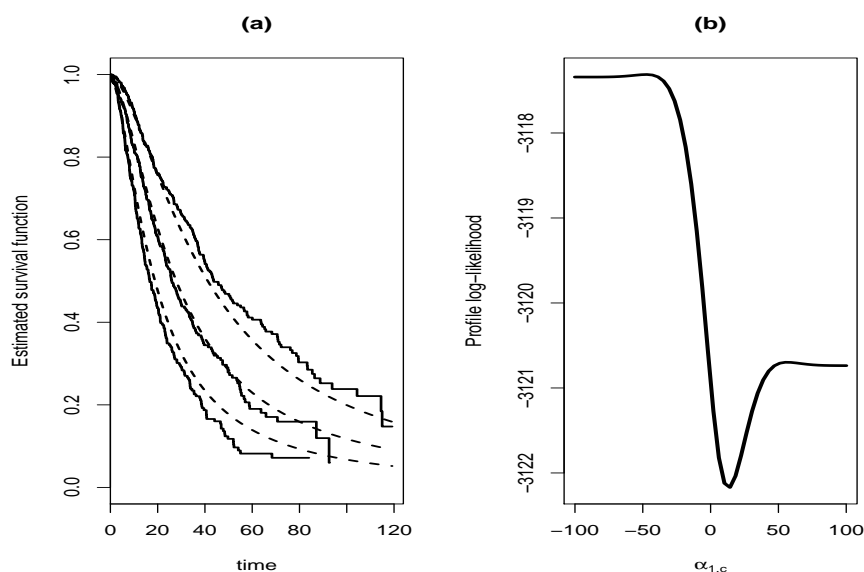
Figure 3: Multiple myeloma data analysis. Figure (a): Kaplan-Meier estimates and fitted log-logistic survival functions (dashed lines) for patients with low, average, and good prognosis, respectively. Figure (b): hierarchical model with skew-normal centres nested within skew-normal countries. Profile log-likelihood of the shape parameter $\alpha_{1,c}$.

Table 1: Simulation results of data generated from the univariate AFT model with skew normal error terms $e_{ij} \sim SN_{CP}(0, 1, \gamma_{1,e})$. Results are based on 500 simulations with $n = 200$ individuals.

Param.	true	log-normal			log-skew-normal		
		mean	Bias ($\times 100$)	MSE ($\times 100$)	mean	Bias ($\times 100$)	MSE ($\times 100$)
β_0	5.00	5.07	6.62	0.97	5.00	-0.24	0.52
β_1	-0.30	-0.32	-1.77	0.69	-0.31	-0.92	0.52
ω_e	1.00	1.10	9.88	1.61	1.00	-0.10	0.39
$\gamma_{1,e}$	-0.90	-	-	-	-0.84	5.98	1.95
β_0	5.00	5.00	-0.09	0.56	5.00	0.14	0.60
β_1	-0.30	-0.30	0.20	0.54	-0.30	0.24	0.55
ω_e	1.00	0.99	-0.65	0.36	1.00	-0.02	0.41
$\gamma_{1,e}$	-0.90	-	-	-	0.04	3.92	5.66
β_0	5.00	4.97	-2.68	0.61	5.01	0.75	0.58
β_1	-0.30	-0.31	-0.51	0.52	-0.30	-0.32	0.28
ω_e	1.00	0.92	-7.63	0.93	1.00	0.25	0.46
$\gamma_{1,e}$	-0.90	-	-	-	0.90	-0.45	0.30

Table 2: Simulation results of data generated from the multivariate AFT model with random effects $b_{ij} \sim SN_{CP}(0, \sigma_b^2, \gamma_{1,b})$ and normal error terms $e_{ij} \sim N(0,1)$. Results are based on 500 simulations with $G = 200$ clusters and $n_i = 10$ individuals per cluster.

Param.	true	mean	SE	Emp. SE	Bias ($\times 100$)	MSE ($\times 100$)
β_0	5.00	4.99	0.04	0.04	-1.32	0.21
β_1	-0.30	-0.30	0.03	0.02	0.36	0.07
ω_e	1.00	0.95	0.02	0.02	-5.03	0.29
σ_b	0.50	0.51	0.04	0.04	1.16	0.19
$\gamma_{1,b}$	-0.90	-0.78	0.22	0.23	12.11	6.56
β_0	5.00	4.98	0.04	0.04	-1.71	0.21
β_1	-0.30	-0.30	0.03	0.02	0.24	0.07
ω_e	1.00	0.95	0.02	0.02	-4.95	0.29
σ_b	0.50	0.50	0.04	0.04	0.31	0.14
$\gamma_{1,b}$	0	0.01	0.30	0.30	0.87	8.90
β_0	5.00	4.99	0.05	0.04	-1.37	0.24
β_1	-0.30	-0.30	0.02	0.02	0.04	0.06
ω_e	1.00	0.95	0.02	0.02	-4.55	0.25
σ_b	0.50	0.49	0.04	0.04	-0.83	0.16
$\gamma_{1,b}$	0.90	0.81	0.21	0.20	-9.25	5.05

Table 3: Model fits to the multiple myeloma data with logistic error term.

Variable	$\hat{\beta}$ (SE) Penal gamma	Penal lognormal	log-skew-normal
Intercept	49.61(20.23)	49.24(19.72)	49.27(19.75)
Stage	-0.16(0.03)	-0.15(0.03)	-0.16(0.03)
Status at cond. (no CR vs CR)	-0.33(0.08)	-0.31(0.08)	-0.31(0.08)
Serum beta(2)-mic. > 4 vs < 4	-0.28(0.05)	-0.28(0.05)	-0.28(0.05)
Age at transplant (+10 yr)	-0.02(0.03)	-0.02(0.03)	-0.04(0.03)
Year (+10)	-0.23(0.10)	-0.22(0.10)	-0.23(0.10)
Gender (F vs M)	0.13(0.05)	0.12(0.05)	0.13(0.05)
Interval diag.-auto. (+1 yr)	-0.04(0.02)	-0.04(0.02)	-0.04(0.02)
Scale ($\hat{\omega}_e$)	0.62(0.01)	0.63(0.01)	0.64(0.01)
Centre SD ($\hat{\sigma}_b$)	0.27(-)	0.18(-)	0.18(0.07)
Centre skew ($\hat{\gamma}_{1,b}$)			-0.78(0.21)

Acknowledgements

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

Department of Statistical Sciences
University of Padua
Italy