

Accelerated failure time model for multivariate two-stage current-status data with parallel and longitudinal correlated random effects

YING-FANG WANG, LIU-CHIH LO AND FUSHING HSIEH*

We develop a parametric accelerated failure time (AFT) model with random effects for analyzing multivariate two-stage current-status survival data. This model structure is motivated by a breeding success study of common ravens in Rostock Germany around the year 1996. Association between land use and the stages of the two breeding events—hatching and fledgling—is of main research interest. Correlation among eggs within the same nest is modeled by a shared bivariate random-effect term, and the correlation of this bivariate random-effect term is designed to account for the dependency between the timing of two breeding events for the same egg. Analytically we construct the likelihood function and derive the maximum likelihood estimate with its asymptotic variance. In regression parameter estimation, the EM algorithm and a Monte Carlo version of the Newton-Raphson maximizer are adapted. A numerical study is also conducted to validate the likelihood based statistical inferences. In the real data analysis, no significant effect for land use was found for either stage. But low nest security in farmland might play some role in the fledgling stage, while food abundance in farmland is typically related positively to the breeding process.

KEYWORDS AND PHRASES: Breeding success, Frailty, Interval censoring, EM algorithm, Metropolis sampler.

1. INTRODUCTION

This methodological development was motivated by an observational study conducted during the common raven's breeding season (from mid-February to the beginning of April in 1996) in a 1570 km² area near Rostock, Mecklenburg-Vorpommern, Germany. This study attempted to gain new information about breeding ecology and habitat selection of common ravens, which were once considerably reduced or extirpated in Central Europe during the early 1900s. Breeding plays a crucial role in the common raven's population dynamics, and a better understanding of the process is very important in the conservational biology of this species. Specifically, such knowledge will allow us to evaluate

the relative importance of difference habitats and management options in the recovery plans. In this study, for each breeding nest, two visits were pre-scheduled to monitor the number of hatched eggs and the number of fledglings that have flown away, respectively, in addition to a prior visit when the clutch size was recorded. One particular interest was to assess habitat suitability with the underlying stages of two serial breeding events—hatching and fledgling. The monitoring time was chosen according to available biological knowledge of this species of raven, and only current-status information at each pre-scheduled visit is obtained. Even though such a sampling schedule is fixed, the duration between the events of interest for each bird can vary to a great extent. A proper survival model is needed to explore the underlying breeding process.

Considering each individual egg to be the analysis unit, there are two correlated current status response variables together with the habitat covariate. Another distinct characteristic of this data is that the clutch size—the number of eggs sharing the same nest—is usually more than 2, and the development of eggs within the same clutch are correlated through several sources. Biologically, breeding is complexly influenced not only by habitat factors, but also by the individual situations of the parental breeding pairs, such as nutrition, health, age, physiological conditions or individual characters, such as aggressive behavior, all of which cannot be readily measured. Therefore, correlation among the eggs within the same nest for reasons that cannot be explained by the habitat factors should also be considered. This dependence carries over from egg to hatching and then to fledgling stage.

Two approaches have typically been discussed in handling the dependency between serial events: marginal model and frailty model. Marginal model focuses on modeling the marginal mean of the cumulative number of events over time, and multiplicative covariate effects are assumed (Ghosh and Lin, 2002 and Chen and Cook, 2004). The dependence among correlated events is implied while the exact dependence structure is left unspecified. Under this semi-parametric setting, the amount of information that needs to be fed into the inference procedure is considerable for the data structure we are targeting. The semi-parametric intensity-based frailty (random effect) models were first in-

*Corresponding author.

troduced by Vaupel et al. (1979), under the Cox Proportional Hazard Model setting. It has been commonly used in explaining the associations among multivariate survival times, whether parallel or longitudinal (Wolfe and Huang, 2004; Zhu et al., 2011). The shared frailty models assume the unobserved factors to be the same within clusters, while correlated random variables governed by a multivariate joint distribution are used to characterize the random effects within each cluster in the correlated frailty model (Yashin and Iachine, 1995; Yashin et al., 1995; Yue and Chan, 1997). But nonparametric estimation of the baseline hazard function is involved in the inference. With observations at only two fixed monitoring times for all subjects, nonparametric estimation will be almost noninformative except for two time points and consequently make the inference procedure very unreliable. The accelerated failure time (AFT) model with additive random effects for multivariate right-censored survival data can be seen in Lambert et al. (2004); Chang (2004); Klein et al. (1999). It has the advantage of being more intuitively and biologically interpretable, and its GLMM-like formulation makes the computation that accommodates the current-status censoring structure feasible. In our analysis, a fully parametric AFT model with additive random effects is developed for serial correlated multivariate current status data. The correlation among eggs within the same nest is explained by a shared random-effects term, and correlated random-effect terms are used to account for the association between the timing of two breeding events for the same egg.

Event counts over specific periods of time provide another way to describe the serial event process when continuous observation of subjects is not practical. He et al. (2008) modeled the multivariate “panel count data” under the assumption of multiplicative covariate effects on the marginal mean functions of the cumulative event counts. In Fiocco, Putter, and Houwelingen (2009), longitudinal counts of recurrent events are modeled by Poisson process with a time-varying gamma frailty term. But in our real data example, not only can we not make continuous observations, but we actually have just one observation for each event process. Trying to model the whole count data process with only one or two observations still seems quite challenging.

Another feature of the breeding process example is the current status data structure, also referred to as “Type-I interval censoring”. It arises when the exact failure time cannot be observed, and each individual is assessed only once at a certain monitoring time. The event of interest is only known to either exceed the monitoring time or not. The monitoring time can generally be random or fixed, as is the case in the other Type I censoring schemes (Klein and Moeschberger, 2003). Under the bivariate copula model setting, Wang and Ding (2000) proposed an inference procedure to estimate the association between two correlated survival events when current status data of a common monitoring time is observed. Sun et al. (2006) developed such

an estimation under the interval censoring scheme and the monitoring times for two events are allowed to be different. But the nonparametric estimation of the survival functions involved in the inference will again make the application almost impossible for our data. Under the Bayesian framework, Komarek and Lesaffre (2006) proposed to jointly model an interval-censored first event time and a doubly interval-censored difference time between first and second events with the AFT model assumption. Association between two subjects of the same cluster is captured by a semi-parametric bivariate error distribution, inspired by the penalized smoothing technique where the shape is partially driven by the data. Komarek and Lesaffre (2008) extended the discussion to a situation when the number of subjects within each cluster can be arbitrarily more than two. A shared random effect term is introduced to the AFT model for the association of the subjects in the same cluster, and a univariate version of the same smoothing technique is used for the random effect and error distribution. Several complications regarding censoring structures are involved in the breeding process example we studied, which signified the uniqueness and necessity of our proposed model. In particular, (1) the two serial events of the same subject do not have a common censoring time; the censoring times are different but in a fixed order (first for hatching and then for fledgling), and (2) the censoring times are not random among subjects; C_{ij1} and C_{ij2} are fixed at 21 days and 42 days for all subjects throughout the whole investigation. The lack of variation in C_{ij1} and C_{ij2} in our real data example makes efficiency extremely important in our modeling attempt, and that’s the main reason why we chose to start with a fully parametric approach over the Bayesian approaches that also requires some modifications in the analysis of breeding process.

The purpose of this paper is to develop a methodology that can be used to analyze multivariate two-stage current-status survival data. It is then to be applied to explore the relationship between habitat selection and breeding dynamics of the common raven. This article is organized as follows: in Section 2, we present the multivariate two-stage AFT model with additive random effects, for current-status data. In Section 3, we discuss the likelihood-based statistical inference. We present the results of a simulation study in Section 4, and real data analysis is reported in Section 5. Some related issues are discussed in Section 6.

2. MULTIVARIATE TWO-STAGE AFT MODEL WITH ADDITIVE RANDOM EFFECTS FOR CURRENT-STATUS DATA

Suppose that in a longitudinal study, the timing of two serial events T_{ij1} and T_{ij2} are of interest, where $i = 1, \dots, n$ is the index of the clusters, and $j = 1, \dots, n_i$ is the index of each individual within cluster i . Without observing the exact timings, instead the status of event 1 and event 2 at two pre-scheduled time points C_{ij1} and $C_{ij1} + C_{ij2}$ are monitored,

respectively. Current-status information, i.e., whether or not an event has yet occurred, is recorded at each time point. Namely, $I_{ij1} = I(T_{ij1} < C_{ij1})$ and $I_{ij2} = I(T_{ij1} + T_{ij2} < C_{ij1} + C_{ij2})$, where $I(\cdot)$ is the indicator function. For both events, we assume an accelerated failure time (AFT) model with an additive random effect term that accounts for the correlation among individuals within the same cluster, aside from what is already explained by the covariates. The AFT models for T_{ij1} and T_{ij2} of each individual are specified as follows:

$$\begin{aligned}\log(T_{ij1}) &= \beta_1 Z_i + \omega_{i1} + \sigma_1 \varepsilon_{ij1} \\ \log(T_{ij2}) &= \beta_2 Z_i + \omega_{i2} + \sigma_2 \varepsilon_{ij2}\end{aligned}$$

where Z_i is a vector consisting of a 1 (to account for the intercept) and the values of the cluster-specific covariates. The distributions of T_{ij1} and T_{ij2} are determined by that of the random error terms, ε_{ij1} and ε_{ij2} . σ_{ij1} and σ_{ij2} are the scale parameters for the error distribution and therefore the variance of ε_{ij1} and ε_{ij2} should be fixed at 1 for identifiability of the parameters. Another way to specify the AFT models is to let σ be absorbed within the distribution of ε . But starting from Section 2.1, our discussion will be focused on the extreme value distribution $f(\varepsilon) = e^\varepsilon e^{-e^\varepsilon}$ that doesn't have a scale parameter specification. We therefore would like to retain σ in the equation here. The pair $(\omega_{i1}, \omega_{i2})$ is the random effect shared by all individuals that belong to cluster i , and is termed *fortitude* for its positive effect on lifetimes (Lambert et al., 2004) in contrast to frailty in the proportional hazards framework. The positive association between the two events for the same individual is accommodated by the correlation between $(\omega_{i1}, \omega_{i2})$. Specifically, we assume that each pair $(\alpha_{i1}, \alpha_{i2}) = (\exp(\omega_{i1}), \exp(\omega_{i2}))$ is a realization from a bivariate positive-valued distribution with marginal mean one and positive correlation. When $(\alpha_{i1}, \alpha_{i2})$ is given, lifetimes T_{ij1} and T_{ij2} are assumed to be independent between each other as well as among different i 's and j 's.

2.1 Likelihood

Throughout we assume censoring to be noninformative about the fortitude. Therefore, independence still holds among the observed data $(C_{ij1}, C_{ij2}, I_{ij1}, I_{ij2})$ for all i and j , given $(\alpha_{i1}, \alpha_{i2})$. The likelihood can be derived as

$$(1) \quad L = \prod_{i=1}^n \int_0^\infty \int_0^\infty \left\{ f_i(C_{ij1}, C_{ij2}, I_{ij1}, I_{ij2} \mid \alpha_{i1}, \alpha_{i2}) \times g(\alpha_{i1}, \alpha_{i2}) \right\} d\alpha_{i1} d\alpha_{i2},$$

where n is the number of clusters, $f_i(C_{ij1}, C_{ij2}, I_{ij1}, I_{ij2} \mid \alpha_{i1}, \alpha_{i2})$ is the joint conditional probability function of the observed data given the fortitude terms for cluster i , and $g(\alpha_{i1}, \alpha_{i2})$ is the probability density function of the fortitude. Four different scenarios can happen

Table 1. Possible scenarios among two serial visits for each individual

A	B	Joint conditional probability representation
Yes	Yes	$P(T_{ij1} < C_{ij1}, T_{ij1} + T_{ij2} < C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})$
Yes	No	$P(T_{ij1} < C_{ij1}, T_{ij1} + T_{ij2} > C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})$
No	Yes	$P(T_{ij1} > C_{ij1}, T_{ij1} + T_{ij2} < C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})$
No	No	$P(T_{ij1} > C_{ij1}, T_{ij1} + T_{ij2} > C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})$

A: Occurrence of Event 1 before the first visit

B: Occurrence of Event 2 before the second visit

among the two serial visits for each individual, as listed in Table 1. Specifically, the joint conditional probability $f_i(C_{ij1}, C_{ij2}, I_{ij1}, I_{ij2} \mid \alpha_{i1}, \alpha_{i2})$ for all individuals in cluster i can be written as

$$(2) \quad \left\{ \prod_{j=1}^{n_i} P(T_{ij1} < C_{ij1}, T_{ij1} + T_{ij2} < C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})^{I_{(I_{ij1}=1, I_{ij2}=1)}} \times P(T_{ij1} < C_{ij1}, T_{ij1} + T_{ij2} > C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})^{I_{(I_{ij1}=1, I_{ij2}=0)}} \times P(T_{ij1} > C_{ij1}, T_{ij1} + T_{ij2} < C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})^{I_{(I_{ij1}=0, I_{ij2}=1)}} \times P(T_{ij1} > C_{ij1}, T_{ij1} + T_{ij2} > C_{ij1} + C_{ij2} \mid \alpha_{i1}, \alpha_{i2})^{I_{(I_{ij1}=0, I_{ij2}=0)}} \right\}$$

where n_i is the number of individuals in each cluster. Conditional on the fortitude, lifetimes follow an accelerated failure time model where the randomness comes from the error terms. While a variety of choices may be considered for the error terms, we adopt the extreme value distribution, which constitutes a Weibull model with random effects. Weibull models have the advantage of satisfying both the accelerated failure time and the proportional hazard assumptions, and it is commonly accepted in modeling survival time. The positive-valued parameters σ_1 and σ_2 are assumed to equal one here simply because the current-status data structure very often does not carry sufficient information to afford complex estimation, as is the case in our breeding success study. Since our main interest is to understand the association between the covariates and the time to two events, measured by β_1 and β_2 , holding other nuisance parameters fixed is a reasonable trade-off between computational simplicity and adequacy. Therefore T_{ij1} and T_{ij2} will reduce to exponential survival times. Explicit expressions of the joint conditional probabilities in Table 1 can be derived as shown in Appendix A, and these probabilities form the basis of our likelihood-based estimation procedure in the next section. A model-checking procedure will be presented in the discussion section to further support the application of our proposed parametric model with this specific setting to the breeding success data.

2.2 Choice for $G(\alpha_1, \alpha_2)$: Bivariate gamma distribution

The Gamma distribution is a positive-valued probability model that is widely applicable in survival analysis, and the two-parameter representation $\text{Gamma}(k, \theta)$ with shape k and scale θ has density

$$g(x; k, \theta) = \frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)}$$

for $x > 0$, and $k, \theta > 0$. This distribution has the property that summation of two independent gamma-distributed random variables with the same scale still follows a gamma distribution, and the scale of a gamma distribution can be adjusted to any value with simple constant multiplication of the original random variable. Therefore, a bivariate gamma distribution whose components are positively correlated can be developed from linear combinations of independent gamma variables (Mathai and Moschopoulos, 1991). In addition, to serve the purpose of random effect in the survival model, the marginal means of both α_1 and α_2 need to be constrained at one. A form of the three-parameter bivariate gamma distribution that can be applied to our proposed model is thus constructed as follows.

Let $V_0 \sim \text{Gamma}(\rho_1, 1)$, $V_1 \sim \text{Gamma}(\rho_2 - \rho_1, 1/\rho_2)$, $V_2 \sim \text{Gamma}(\rho_3 - \rho_1, 1/\rho_3)$ be three mutually independent gamma-distributed random variables with both ρ_2 and ρ_3 greater than ρ_1 , and let

$$A_1 = \frac{1}{\rho_2} V_0 + V_1$$

$$A_2 = \frac{1}{\rho_3} V_0 + V_2$$

Then (A_1, A_2) has a bivariate gamma distribution, whose marginal means both equal one and whose marginal variances are $1/\rho_2$ and $1/\rho_3$, respectively. A_1 and A_2 are positively correlated through a common variable V_0 , and the correlation coefficient equals $0 < \rho_1/\sqrt{\rho_2\rho_3} < 1$. For the same reason as is stated in Section 2.2, throughout the rest of the paper, both ρ_2 and ρ_3 are chosen to be fixed at one to narrow down the number of parameters to be estimated. The joint density of (A_1, A_2) , as derived in Appendix B, involved integration of a polynomial function to an unknown power and does not have an analytical form. Nevertheless, bivariate Gamma random variables defined this way are easy to generate, and this makes our Markov Chain Monte Carlo (MCMC)-based estimation procedure feasible.

3. ESTIMATION PROCEDURE

The likelihood function in (1) cannot be evaluated in closed form because the integration is intractable, and neither can the log-likelihood function. To avoid such computational difficulties, the Expectation-Maximization (EM) algorithm is set up by considering the fortitude terms α_{i1} and

α_{i2} to be the missing data. The log-likelihood based on the complete data $W = (C_1, C_2, I_1, I_2, \alpha_1, \alpha_2)$ is then given by:

$$(3) \quad \log(L_W) = \sum_{i=1}^n \{ \log[f_i(C_{ij1}, C_{ij2}, I_{ij1}, I_{ij2} | \alpha_{i1}, \alpha_{i2})] + \log[g(\alpha_{i1}, \alpha_{i2})] \}$$

The conditional expectation of the complete-data log-likelihood given the observed data $Y = (C_1, C_2, I_1, I_2)$ required in the EM algorithm still cannot be computed directly, because the conditional distribution involves the aforementioned integration. However, Monte Carlo approximation of the conditional expectation can be formed because sampling from the conditional distribution $(\alpha_1, \alpha_2) | Y$ is possible. A particularly neat form of the Metropolis sampler as suggested by McCulloch (1997) is described in Section 3.2.

3.1 Monte Carlo Newton-Raphson procedure

To speed up maximization of the likelihood, the Newton-Raphson procedure is often adapted when the score $U(\beta)$, first partial derivatives of the log-likelihood function, and the observed information $I(\beta)$, negative of the second partial derivatives, can be easily calculated as a function of β . Each $\beta^{(m)}$ is updated by $\beta^{(m+1)} = \beta^{(m)} + [I(\beta^{(m)})]^{-1} U(\beta^{(m)})$ iteratively until convergence. In our EM setting, the primary parameters of interest $\beta = (\beta_1, \beta_2)$ enter only the first term of the complete-data log-likelihood function in (3), and the complete-data score with respect to β can be derived, as shown in Appendix C. The conditional expectation of the complete-data score given the observed data is shown to equal the observed-data score (Gauderman and Navidi, 2001): $U(\beta; Y) = E[U(\beta; W) | Y]$. Therefore, it can be approximated to take place of $U(\beta)$ in the Newton-Raphson procedure, and so can the conditional-expectation expression of the information defined below (Louis, 1982) in place of $I(\beta)$.

$$I(\beta; Y) = E[I(\beta; W) | Y] - \text{Cov}[U(\beta; W) | Y]$$

This is also how the observed information matrix will be extracted in the EM algorithm to obtain the standard error estimates once convergence is achieved. Since both $W_i = (C_{ij1}, C_{ij2}, I_{ij1}, I_{ij2}, \alpha_{i1}, \alpha_{i2})$ and $Y_i = (C_{ij1}, C_{ij2}, I_{ij1}, I_{ij2})$ are independent among $i = 1, \dots, n$, $U(\beta; Y)$ and $I(\beta; Y)$ can be further broken down into the summation of n terms: $U(\beta; Y) = \sum_{i=1}^n E[U_i(\beta; W_i) | Y_i]$ and $I(\beta; Y) = \sum_{i=1}^n E[I_i(\beta; W_i) | Y_i] - \text{Cov}[\sum_{i=1}^n U_i(\beta; W_i) | Y_i]$. Monte Carlo approximations are then constructed in the following

manner,

$$(4) \quad \hat{U}(\beta; Y) = \sum_{i=1}^n \left\{ \frac{1}{K} \sum_{k=1}^K U_i(\beta; W_i^k) \right\}$$

$$(5) \quad \hat{I}(\beta; Y) = \sum_{i=1}^n \left\{ \frac{1}{K} \sum_{k=1}^K I_i(\beta; W_i^k) \right\} \\ - \sum_{i=1}^n \sum_{i'=1}^n \left\{ \frac{1}{K} \sum_{k=1}^K \{U_i(\beta; W_i^k) U_{i'}^T(\beta; W_{i'}^k)\} \right\} \\ + \sum_{i=1}^n \left\{ \frac{1}{K} \sum_{k=1}^K [U_i(\beta; W_i^k)] \right\} \\ \times \sum_{i=1}^n \left\{ \frac{1}{K} \sum_{k=1}^K [U_i(\beta; W_i^k)]^T \right\}$$

where each W_i^k consists of the observed data Y_i and $(\alpha_{i1}^k, \alpha_{i2}^k)$, a random sample drawn from the conditional distribution of $\pi((\alpha_{i1}, \alpha_{i2}) | Y_i)$. An EM-based Monte Carlo Newton-Raphson (MCNR) approach in finding maximum likelihood estimates (MLE) can be carried out using the approximations in (4) and (5), elaborated as the following steps:

Step 1. Select initial values $\beta^{(0)}$ and $\rho^{(0)}$.

Step 2. E step

a) Generate K pairs of random samples, $(\alpha_{i1}^{(1)}, \alpha_{i2}^{(1)})$, $(\alpha_{i1}^{(2)}, \alpha_{i2}^{(2)})$, \dots , $(\alpha_{i1}^{(K)}, \alpha_{i2}^{(K)})$ from the conditional distribution $\pi(\alpha_{i1}, \alpha_{i2} | Y_i; \beta^{(m)}, \rho^{(m)})$ for $i = 1, \dots, n$ using the Metropolis sampler described in Section 3.2.

b) Calculate the Monte Carlo approximations as in (4) and (5).

Step 3. M step

a) Update β value with $\beta^{(m+1)} = \beta^{(m)} + [\hat{I}(\beta^{(m)}; Y)]^{-1} \hat{U}(\beta^{(m)}; Y)$.

b) Choose $\rho^{(m+1)}$ that maximize $1/K \sum_{k=1}^K \log[g_\rho(\alpha_1^k, \alpha_2^k)]$.

Step 4. Repeat Steps 2–4 until convergence and declare $\beta^{(m+1)}$ and $\rho^{(m+1)}$ to be the MLE.

Step 5. Repeat Step 2 one more time to calculate the MC estimates of the observed information matrix. The standard error estimate is then $[\hat{I}(\beta; Y)]^{-1/2}$.

Note: Since our methodology focuses on studying the association between the covariates and the time to two events, the algorithm is mainly developed to estimate β_1 and β_2 . Due to the computational difficulty of the intractable likelihood function of the bivariate gamma distribution, only a point estimate of the nuisance parameter ρ is given, while its precision is unknown.

Table 2. Percentage of each scenario in the simulated data

A	B	Percentage	
		$n = 50$	$n = 100$
Yes	Yes	70%	70%
Yes	No	10%	10.5%
No	Yes	5%	5.5%
No	No	15%	14%

3.2 Metropolis sampler

When generating samples directly from a distribution is not possible, a Markov chain that converges to a stationary distribution of interest can be used instead. Our target distribution here is the bivariate conditional distribution $\pi(\alpha_i | Y_i)$ with α_i denoting $(\alpha_{i1}, \alpha_{i2})$ for notational simplicity. In the $(k+1)^{th}$ step of the Metropolis sampler, a candidate point $\alpha_i^{(k+1)*}$ is sampled from a proposal distribution $q(\alpha_i | \alpha_i^{(k)})$. With probability $A(\alpha_i^{(k)}, \alpha_i^{(k+1)*}) = \min\{1, \frac{\pi(\alpha_i^{(k+1)*} | Y_i) q(\alpha_i^{(k)} | \alpha_i^{(k+1)*})}{\pi(\alpha_i^{(k)} | Y_i) q(\alpha_i^{(k+1)*} | \alpha_i^{(k)})}\}$, the candidate point is accepted to be $\alpha_i^{(k+1)}$; otherwise $\alpha_i^{(k+1)} = \alpha_i^{(k)}$. As suggested by McCulloch (1997), when G the bivariate distribution of the fortitude terms is chosen as the proposal distribution, the acceptance probability involves only the conditional distribution $f_i(Y_i | \alpha_i)$ as derived in (2) and is easy to compute: $A(\alpha_i^{(k)}, \alpha_i^{(k+1)*}) = \min\{1, \frac{f_i(Y_i | \alpha_i^{(k+1)*})}{f_i(Y_i | \alpha_i^{(k)})}\}$. The chain can be repeated for any desirable length. A burn-in period, usually the first 1% ~ 2% of the total iterations, is to be discarded, and the remaining constructs a random sample of $\pi(\alpha_i | Y_i)$.

4. SIMULATION

4.1 Performance of the estimation procedure under correct model assumptions

Simulations were conducted to evaluate the proposed model and the performance of the estimating procedure. We examine samples of two different sizes: $n = 50$ clusters and $n = 100$ clusters, both with cluster size $n_i = 4$. Data are generated based on the model assumption in Section 2, considering one continuous covariate Z following a standard normal distribution, and regression coefficients $\beta_1 = \ln(2)$ for the first stage and $\beta_2 = -\ln(2)$ for the second stage. The fortitude terms (α_1, α_2) are random draws from the bivariate Gamma distribution described in Section 2.2, with marginal means one, marginal variances one, and correlation coefficient 0.5. The error terms, ε_1 and ε_2 , are generated from two independent extreme-value distributions. Fixed censoring times, C_1 and C_2 , are chosen to allow approximately 80% of the data to have event 1 occur before C_1 so that $I_1 = 1$, while 75% of the data have event 2 occur before $C_1 + C_2$ so that $I_2 = 1$. Table 2 shows the approximate percentages

Table 3. Summary of parameter estimation results from 300 simulation replicates

	Proposed model				Modeling separately			
	Bias	T.SE	E.SE	95% CR	Bias	T.SE	E.SE	95% CR
<i>n</i> = 50								
β_1	0.000	0.145 (0.225*)	0.214	82%	-0.142	0.214	0.198	89%
β_2	-0.014	0.179 (0.256*)	0.229	89%	0.688	0.175	0.147	3%
ρ	-0.011		0.054					
<i>n</i> = 100								
β_1	0.016	0.101	0.151	82%	-0.122	0.149	0.128	87%
β_2	0.006	0.122	0.173	82%	0.698	0.121	0.096	0%
ρ	-0.016		0.046					

* numbers in brackets are the Theoretical Standard Error estimated by 30 Bootstrap replicates

of the data that fall into each of the four scenarios given in Table 1.

We generate 300 replicates and fit the proposed model to each. The average bias, average estimated theoretical standard error (T.SE), empirical standard error (E.SE), and coverage probability of the 95% confidence interval are summarized in Table 3. When the sample size is 50, as shown in the top-left corner of the table, the biases of all the parameter estimates are fairly small, concordant with the consistency of the MLEs. The average theoretical standard error estimated by step 5 of the EM-based MCNR algorithm, and the empirical standard error calculated as the sample standard deviation of 300 estimation values, are both reported in order to examine the variation of the β estimates. The theoretical SE is obviously underestimated for several possible reasons, one of which may be ignorance of the variation coming from the estimates for nuisance parameter ρ in step 5 due to computational limitations. Such underestimation of the SEs is also reflected in the 95% coverage rate being lower than the nominal level, at 82% and 89% respectively. Bootstrap can be a potential remedy in such situations that the analytical derivation is intractable. But because our MCNR estimation itself already relies on an iterating procedure, the computational time to combine our MCMC iterations with the bootstrap estimation is not likely to be very practical for most users. For the small sample size setting, we report our simulation summary when the Bootstrap SEs are estimated by 30 resamplings of clusters with replacement (Field and Welsh, 2007). It shows that the bootstrap procedure has the potential to recover much of our underestimation in the theoretical SE (even a slight overestimation is reflected). The estimates will slowly become closer to the sample SE as the number of bootstrap replicates goes up to 100 or even 200, but we stopped at 30 for the purpose of this demonstration.

As shown in the right side of the table, the results are compared to those obtained when modeling the two stages separately using the AFT model with additive random effects by Therneau et al. (2003) in R/SPLUS. Our proposed approach has a dramatic advantage in estimating the regression coefficients at the second stage, demonstrated by a much smaller sample bias and a much higher 95%

coverage rate. Modeling separately suffers from the fact that the starting time of the second stage is unknown and the censoring time of the first stage is naively taken as a substitute, while our proposed model adequately takes into account such a censoring structure. A somewhat smaller bias is also presented by our method in the coefficient estimates of the first stage. From several sets of simulation studies which are not reported here, we believe that the traditional methods were unable to perform satisfactorily even for the first stage mainly because (1) the amount of information carried by our current-status censoring structure with fixed monitoring times is very limited, and (2) the two stages are correlated with each other beyond the sharing of common covariates and modeling jointly has the advantage when such dependency should not be ignored. When we increase the sample size to 100, a similar trend persists, and a higher degree of precision can be reached, as reported in the second part of the table. As a computational remark, we note that the amount of information of ρ becomes small when ρ is away from 0.5 and close to 1 or 0. In those cases the EM-algorithm becomes instable or even not converging.

4.2 Comparison with other existing methods

Our approach is also compared to an existing method proposed by Komarek and Lesaffre (2008). The parametric AFT formulation in their Bayesian approach is very similar to that in our proposed model, and the estimation procedure for both of our methods mainly rely on the MCMC iterations. While they employ the data-driven smoothing techniques to allow for flexible distributional assumptions in the random effect and error terms, their observations were made in periodic visits and the event times were interval censored. We were mainly concerned with the situation when observation is rather difficult or expensive, and only one observation is made for each event which results in the current-status data. With the exact same simulation setting of small sample size ($n = 50$), we applied the Bayesian AFT model in Komarek and Lesaffre (2008) using the R BayesSurv package. Since Current-status data is a special case of interval censoring, the application is straightforward for most cases.

Table 4. Bayesian approach (Komarek and Lesaffre, 2008)

	Bias	SD	MSE
Current-status data with augmented information			
β_1	0.515	0.589	0.612
β_2	0.505	1.019	1.294
Interval censoring with multiple visits			
β_1	-0.023	0.187	0.036
β_2	0.069	0.171	0.034

However, when both events have not happened by the time of each observation ($I_1 = 0$ and $I_2 = 0$), distinction needs to be made whether the first event has already happened or not by the time of the second observation. This is a very natural observation in a dental example, but not necessarily so in other cases when measurement of one event does not come along with measurement of the other event. In the numbers reported here for the purpose of numerical investigation, we have had to add such unobserved information to 15% of the data (the forth group: $I_1 = 0$ and $I_2 = 0$) to make the Bayesian approach applicable. The information being fed into the two methods are not equally fair, and in fact the Bayesian approach has the advantage of using the more detailed data. However, as shown in the upper panel of Table 4, the Bayesian approach cannot give accurate and precise estimates for the regression coefficients. We believe this has more to do with the lack of information due to the current-status censoring structure, rather than the unknown underlying distributions of the true event times. To confirm, we did another set of analysis with the exact same simulated event times, but interval censored by first visit drawn from $N(0.1, 0.02^2)$ and consecutive visits of distance drawn from $N(0.05, 0.005^2)$. The results are shown in the lower panel of Table 4, and the Bayesian approach now performs reasonably well as it was designed to.

4.3 Robustness against random-effect misspecification

While a more relaxed distributional assumption is preferred so that the justification can be less concerned, the cost is very likely to be the information carried by the data and the censoring structure, which is the main pitfall of the type of data we are facing. While the justification of this relatively restrictive assumption is another challenging topic (a crude model checking and diagnostic tool will be briefly discussed in Section 6), below we evaluate through simulations the robustness of our proposed method against misspecification of the random effect distribution in light of its potential to be applied to other scientific settings where repeated measurements are not feasible. The evaluation has mainly focused on the performance of our proposed model when there is a negative dependency between the two event times besides the effect caused by their common covariates of interest. The fortitude terms (α_1, α_2) are random draws from

Table 5. Simulation Comparisons: $n = 50$ clusters with cluster size $n_i = 4$. $Z \sim N(0, 1)$, $\beta_1 = \ln(2)$, $\beta_2 = \ln(2)$, and $(\alpha_1, \alpha_2) \sim$ bivariate lognormal distribution with marginal means one, marginal variances one, and correlation coefficient -0.5 and -0.3 . Both ε_1 and $\varepsilon_2 \sim$ extreme-value distribution. Fixed censoring times, C_1 and C_2 , are chosen to allow approximately 80% of the data to have event 1 occur before C_1 so that $I_1 = 1$, and 75% of the data have event 2 occur before $C_1 + C_2$ so that $I_2 = 1$

	Proposed model			
	Bias	T.SE	E.SE	95% CR
$r = -0.5$				
β_1	0.072	0.148	0.231	79%
β_2	0.226	0.205	0.305	73%
ρ	0.837		0.022	
$r = -0.3$				
β_1	0.072	0.149	0.221	81%
β_2	0.167	0.204	0.307	76%
ρ	0.860		0.029	

the bivariate lognormal distribution with marginal means constrained at one. Marginal variances are chosen to be one as well and the correlation coefficient equal to -0.5 , or -0.3 for the last setting in Table 5. Fixed censoring times, C_1 and C_2 , are chosen to allow for the desired rates of events that have already happened, as specified in more details next to the presentation of the each result. Everything else not mentioned here remains the same as the small sample size ($n = 50$) simulation setting in Table 3.

1. When the common covariates have an effect in an opposite direction:

Our estimation procedure for the regression coefficients (β_1 and β_2), which is the main target of our methodology, is not very sensitive to departure of the positive-correlation assumption in the proposed model when (1) the common covariates have an effect on the event time in an opposite direction that facilitate the negative correlation between the two event times, and (2) the rates of events happened before monitoring times are high. In current-status data, we only observe the status of the event once at the monitoring time. Therefore, the actual event times (T_1, T_2) are unobserved, but only the monitoring time (C_1, C_2) and its status (I_1, I_2) are recorded. High rates of already-happened events refer to relatively late monitoring times during the study. In such cases, many of the unobserved actual event times tend to be much smaller than the monitoring times because the event time (T_1, T_2) grows exponentially under the AFT assumption. We do not however observe exactly how much smaller they are, and that's one of the reason why misspecified random effect distribution can sometimes get away for the purpose of regression estimates. The underlying negative dependency is absorbed by the β_1 and β_2 estimates of opposite direction

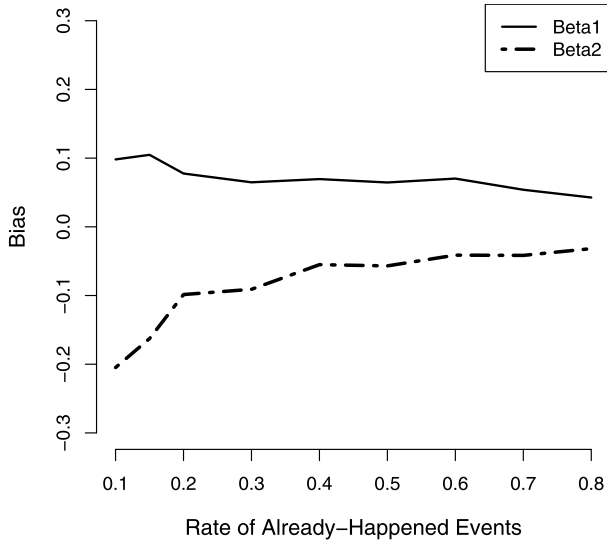


Figure 1. Bias for β_1 and β_2 estimates vs Rate of Already-Happened Events: $n = 50$ clusters with cluster size $n_i = 4$. $Z \sim N(0, 1)$, $\beta_1 = \ln(2)$, $\beta_2 = -\ln(2)$, and $(\alpha_1, \alpha_2) \sim$ bivariate lognormal distribution with marginal means one, marginal variances one, and correlation coefficient -0.5 . Both ε_1 and $\varepsilon_2 \sim$ extreme-value distribution. Fixed censoring times, C_1 and C_2 , are chosen to allow approximately **10% to 80%** of the data to have event 1 occur before C_1 so that $I_1 = 1$, and **10% to 80%** of the data have event 2 occur before $C_1 + C_2$ so that $I_2 = 1$.

and reflected in a reasonable bias away from zero. On the other hand, when rates of already-happened events are low and monitoring times are early, many of which are closer to the actual event times, our model will lose its robustness against negatively-correlated random effects. The bias for β_2 estimates starts to pass beyond 0.1 when the rates of already-happened events are below 30% for both events (see Figure 1).

2. When the common covariates have an effect in the same direction:

Other times, the covariates have an effect on the two events in the same direction while frailty takes effect in the opposite direction, and that’s when our proposed model can become more problematic. β_2 tends to be overestimated, even when the rate of observed events are high (80% for event 1 and 75% for event 2; see the upper of Table 5). When the negative correlation drops to -0.3 , such bias decreases a little bit but is still unignorable (see the lower panel of Table 5).

Several other simulation results also support the robustness of our model when the random effects are generated by a bivariate gamma distribution with various covariance matrix (marginal variances that are not necessarily one), but esti-

Table 6. Frequencies of each scenario in the real data

A	B	Frequency
Yes	Yes	212
Yes	No	12
No	Yes	0
No	No	64

mated by a one-parameter (ρ) bivariate gamma; the tables are omitted here. In summary, this investigation brings to our awareness certain scenarios that our model might lose its robustness against random-effect misspecification, particularly when the bivariate random effect is negatively correlated. The relative direction of the regression coefficients for the common covariates plays a role, as well as the rates of already-happened events. When the common covariates influence the two stages in the same direction and the bivariate random-effect does not, our method might not be able to provide a consistent estimate. But as expected, the bias decreases as the negative correlation increases towards zero.

5. ANALYSIS OF THE BREEDING SUCCESS OF COMMON RAVENS

In the 1996 breeding process study on common ravens in Germany, 65 nests were visited at sparsely-scheduled time points. The average number of eggs in each nest was 4.43, and there were 288 eggs observed in total. One of the visits happened 21 days after the visit when clutch size was counted, followed by another visit which happened 42 days later. Notice that C_{ij1} and C_{ij2} can generally be either random time points or fixed ones, and the example presented here is a special case of current-status data where the monitoring time is fixed. Techniques developed for our proposed model are based on the conditional distribution of the actual event time, given the monitoring time. They can typically be used for random censoring times, but work equally well for fixed ones (Jewell and van der Laan, 2004). At these two visits, current-status information is recorded for each egg regarding hatching and fledgling flown away, respectively. The frequencies of the four scenarios are listed in Table 6. Our analysis focuses on the association between land use of the breeding territory and timing of the two breeding events. The percentage of farmland use in breeding territory is used as the covariate, with a mean of 76.94% and standard deviation of 14.69% among all the observed eggs.

In a previous analysis (2011, unpublished), land use was identified as the single-most important factor in the regression analysis of reproduction, as measured by the number of eggs laid by each female common raven. However, it is not even a significant factor in the logistic regression analysis of breeding success. Breeding is defined as being successful for each nest if at least one fledgling has flown away from the nest in the previous analysis (2011, unpublished), which is

6. DISCUSSION

Table 7. Breeding success data analysis result

	Coefficient Estimates	S.E.	Z	P-value
Proposed Model*				
Time to Egg Hatched				
(Intercept)	2.39	0.58	4.12	0.00
Farmland %	0.72	0.75	0.96	0.34
Time to Fledglings Flown Away				
(Intercept)	3.42	0.69	4.96	0.00
Farmland %	-0.18	0.89	-0.20	0.84
Modeling Separately				
Time to Egg Hatched				
(Intercept)	2.04	0.79	2.57	0.01
Farmland %	0.46	1.02	0.45	0.65
Time to Fledglings Flown Away				
(Intercept)	3.01	0.79	3.83	0.00
Farmland %	0.28	1.02	0.27	0.78

* $\hat{\rho} = 0.6$, but no further inference is made regarding this parameter

related to the second event of each individual egg that we discuss here. Now we have information about the hatching stage, and we want to incorporate the information about both events. When the time to hatching and the time to fledgling were modeled separately using the AFT model with additive random effects, the percentage of farmland use in the breeding territory shows an insignificant positive effect for both events with Z -statistics of 0.45 and 0.27 respectively, as shown in Table 7. In this naive analysis, the time of the first visit was taken as the starting point of the second event. When we consider a more realistic censoring structure and model the breeding process jointly with our proposed approach, the percentage of farmland still does not show a significant effect for the hatching stage, even though the Z -statistic is slightly higher (0.96). From all unpublished previous analyses as well as the current analysis for the hatching stage using our proposed model, the percentage of farmland has always been positively related to reproduction and breeding of common ravens. It is noteworthy here that even though it is also insignificant, a negative coefficient estimate is revealed in the fledgling stage. This data set is relatively small and does not provide sufficient statistical evidence of a negative effect, but such a potential negative effect if valid may bring out an interesting biological implication. For example, a higher percentage of farmland in the breeding territory provides a more abundant food supply, but the raven offspring may therefore be exposed to a higher predation risk. Food abundance primarily contributes to the breeding process through the hatching stage, and nest security seems more important in the fledgling stage. As discussed in the simulation study, a larger sample size might provide better statistical evidence if such a trend indeed exists.

The methodology presented in this paper is developed to handle the multivariate two-stage current-status survival data. Our accelerated failure time (AFT) model with parallel and longitudinal random effects provides a realistic model for the breeding success of common ravens. Both inter-subject correlations at the same stage and intra-subject correlations between stages are modeled through bivariate random effect terms. We develop regression coefficient estimates with their standard errors and the intra-correlation of survival times between the two stages. Our proposed model and its statistical inference have valuable potential in application to other real-world problems in which the timing of two serial events is of interest, and yet only current-status information is observed.

We hereby provide a crude model-checking tool to informally assess the fit of our proposed model to the breeding success data. Cox-Snell residuals (Cox and Snell, 1968; Klein and Moeschberger, 2003) of the exponential model (Weibull with unity scale parameter) defined by $r_{ijk} = \exp(\log(C_{ijk}) - \tilde{\eta}_{ik})$, $k = 1, 2$, where C_{ijk} is the monitoring time of current-status information, and $\tilde{\eta}_{ik} = \hat{\beta}_k Z_i + \tilde{\omega}_{ik}$ is the linear predictor, should approximate to a current-status sample from a standard exponential distribution if the model fits the data (Farrington, 2000). The logarithm of an empirical Bayes estimate ($\tilde{\alpha}_1, \tilde{\alpha}_2$) (the mode of the posterior distribution of (α_1, α_2) given the data) are used to approximate the unobserved fortitude terms involved in the linear predictors (Lambert et al., 2004; Carlin and Louis, 2000). A current-status version of the non-parametric survival function (Turnbull, 1974) is then calculated to compare with the standard exponential distribution, as shown in Figure 2. Because of the lack of information from our data structure, $(\tilde{\omega}_1, \tilde{\omega}_2) = (\log \tilde{\alpha}_1, \log \tilde{\alpha}_2)$ may not have been approximated very accurately. Specifically, when all eggs in one nest are left censored for both events, as is the case for almost half of the observed sample, $(\tilde{\omega}_1, \tilde{\omega}_2)$ tends to be underestimated, and therefore produces large residuals. The dotted/dashed curves seem to correspond well with the solid curve except for the tail of the second stage that might have been caused by the overestimation of some residuals. Given such conditions, no obvious inadequacy is suggested by the figure. With these empirical Bayes estimates ($\tilde{\alpha}_1, \tilde{\alpha}_2$), we can also conduct a crude model checking for the gamma distribution of random effects, using the existing diagnostic tools. Figure 3 presents a Q-Q plot of $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ combined as one sample compared to the marginal gamma distribution of mean one and variance one. The empirical Bayes estimates ($\tilde{\alpha}_1, \tilde{\alpha}_2$) seem well aligned with the theoretical gamma distribution when the value is less than 1.75 which has a cumulative probability of 0.826. There is some discrepancy in the high end of the range, showing that the gamma distribution is more dispersed towards some large values. Again, a possible explanation is the underestimation of the empirical Bayes estimates as discussed above. The sensitivity and rigid use

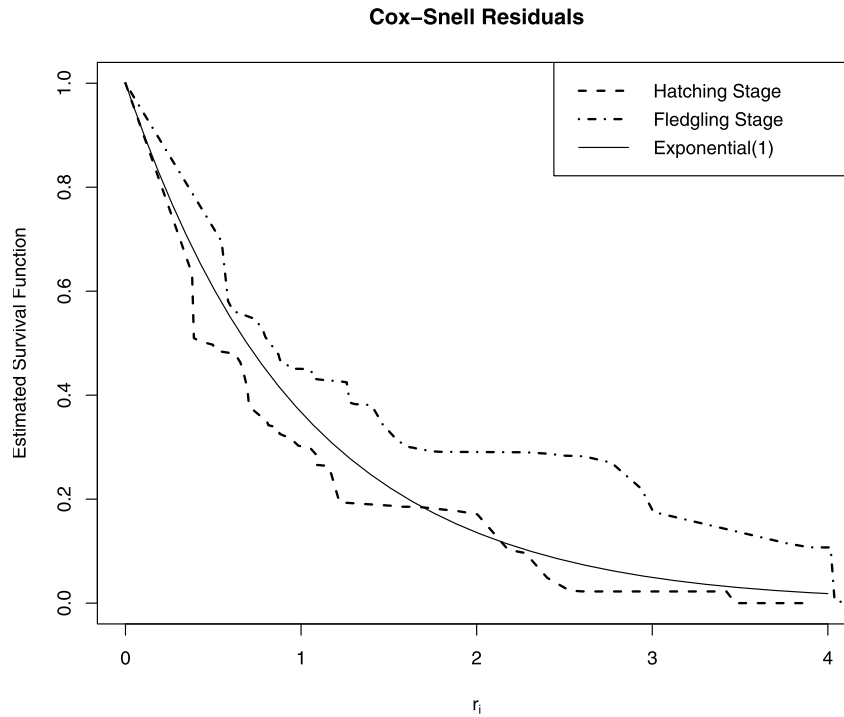


Figure 2. Cox-Snell Residuals for both stages, expected to be close to Exponential(1) if the model fits the data.

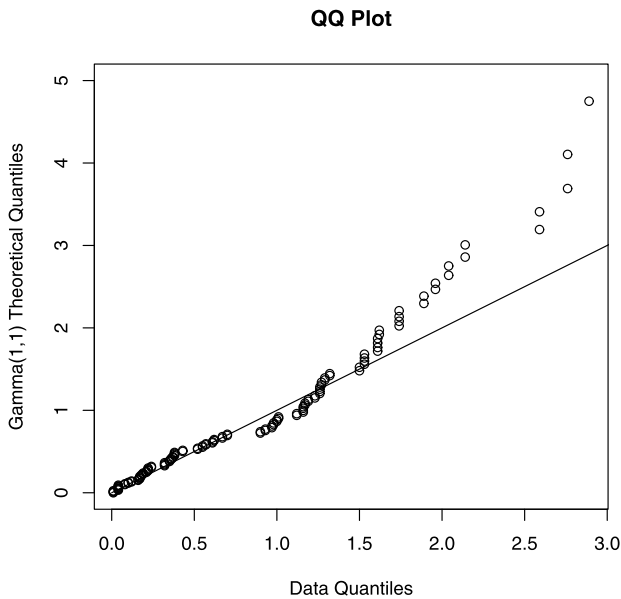


Figure 3. Q-Q plot of the empirical Bayes estimates of fortune terms vs. Gamma distribution, expected to approximately lie on the line $y = x$ (the solid line) if the assumption holds.

of this diagnostic tool needs further confirmation, but rigorous model checking itself is a complicated topic and is beyond the scope of this article.

One might notice that no egg that was still not hatched in the first observation attempt became fledgling in the second observation (frequency being zero in the third row of Table 6). This might have to do with the fact that hatching after a certain point is quite impossible, or just the limitation of our observation method. But in either case, we believe the percentage is rather small or close to zero, and therefore this modeling attempt still provides a reasonable way to study the association between habitat selection and the breeding success. The choice of fitting a frailty model, rather than a logistic mixed-effects model, is cross-validated for a single stage (hatching). The predictive error rates (24.50% for the frailty model and 25.83% for the logistic mixed-effects model, respectively) show slight preference to the frailty model. Furthermore, the frailty model provides additional information about time, and its underlying foundation coincides better with the nature of our biological question. Given the complex data structure and the limited amount of information content, non-significant statistical results are somehow expected even with serious modeling attempts. But the statistical non-significance shall not reduce the biological importance borne by this analysis.

In the breeding process of common raven, we expect the breeding pair who was able to hatch the eggs fast would carry the same ability and tendency to the fledgling stage. Therefore, a positive correlation seems to be a reasonable assumption for the analysis of breeding success data. Although random effects in survival analysis are often modeled by a

Gamma distribution, it would be useful to relax the positive-correlation assumption implied by the bivariate gamma distribution in Section 2.2. Our estimation procedure is directly extendable to any other bivariate positive-value random effect distribution with constraints of unity marginal means, as long as the bivariate random variables can be easily sampled so that the K pairs of (α_1, α_2) from the conditional distribution in Step 2 of the MCNR algorithm can be generated. Bivariate lognormal (Xue and Brookmeyer, 1996) generated by exponential transformation of bivariate normal is an example and it allows for negative correlation. But again, the number of parameters that needs to be estimated has to be somewhat restricted, depending on the information carried by the censoring distribution (for example, random censoring time is usually better than the fixed censoring time in this sense). Our model is directly applicable to a more typical current status data type with random censoring times. It can also be adapted beyond group-specific covariates, such as time-independent covariates of various levels or even hierarchical covariates. The assumption of extreme-value-distributed error terms in this paper implies that the survival time is Weibull-distributed with an additive random effect. This model setup is in fact equivalent to a fully-parametric Cox proportional hazard model with a multiplicative random effect, a so-called frailty model, with a Weibull baseline hazard. When the baseline hazard is arbitrary, Shiboski (1998) provides a semi-parametric method in analyzing current-status data. Apparently the semi-parametric extension to the setting of two-stage current-status survival data with frailty might be worthy of consideration. However the needed computations are expected to be very challenging. The assumption of extreme-value-distributed error terms in this paper implies that the survival time is Weibull-distributed with an additive random effect. This model setup is in fact equivalent to a fully-parametric Cox proportional hazard model with a multiplicative random effect, a so-called frailty model, with a Weibull baseline hazard. When the baseline hazard is arbitrary, Shiboski (1998) provides a semi-parametric method in analyzing current-status data. Apparently the semi-parametric extension to the setting of two-stage current-status survival data with frailty might be worthy of consideration. However the needed computations are expected to be very challenging. Despite several of the restrictions in our model assumptions, we still believe this approach has the potential to findings that cannot be discovered by traditional methods.

APPENDIX A. JOINT CONDITIONAL PROBABILITIES OF $(T_1, T_1 + T_2)$ GIVEN THE FORTITUDES

The error terms with extreme-value distribution have density $f(\varepsilon) = e^\varepsilon e^{-e^\varepsilon}$. Given the fortitudes, lifetime follows

an accelerated failure time model and its conditional density function can be derived as

$$\begin{aligned} f(t_{ijk}|\alpha_{i1}, \alpha_{i2}) &= \alpha_{ik}^{-1} e^{(-\beta_k z_i)} \exp[-\alpha_{ik}^{-1} e^{(-\beta_k z_i)} t_{ijk}], \text{ for } k = 1, 2 \end{aligned}$$

The joint conditional density of $(T_1, T_2|A_1, A_2)$, which are mutually independent, is then

$$\begin{aligned} f_{T_1, T_2|A_1, A_2}(t_1, t_2|\alpha_1, \alpha_2) &= \alpha_1^{-1} \alpha_2^{-1} e^{(-\beta_1 Z - \beta_2 Z)} \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} t_1 \\ &\quad - \alpha_2^{-1} e^{(-\beta_2 Z)} t_2], \text{ for } t_1 > 0, t_2 > 0 \end{aligned}$$

Note that the indices i and j are omitted here for simplicity of notation. An explicit expression of the joint conditional density of $(T_1, T_1 + T_2|A_1, A_2)$ can be obtained by transformation with the Jacobian = 1,

$$\begin{aligned} f_{T_1, T_1+T_2|A_1, A_2}(a, b|\alpha_1, \alpha_2) &= \alpha_1^{-1} \alpha_2^{-1} e^{(-\beta_1 Z - \beta_2 Z)} \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} a \\ &\quad - \alpha_2^{-1} e^{(-\beta_2 Z)} (b - a)], \text{ for } 0 < a < b \end{aligned}$$

The joint conditional probability is then available through integration of the joint conditional density with respect to T_1 and $T_1 + T_2$ within the appropriate range:

(I)

$$\begin{aligned} P(T_1 < C_1, T_1 + T_2 < C_1 + C_2|A_1, A_2) &= \int_0^{C_1} \int_a^{C_1+C_2} f_{T_1, T_1+T_2|A_1, A_2}(a, b|\alpha_1, \alpha_2) db da \\ &= 1 - \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1] \\ &\quad + \frac{\alpha_1^{-1} e^{(-\beta_1 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\ &\quad \times \left\{ \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1 - \alpha_2^{-1} e^{(-\beta_2 Z)} C_2] \right. \\ &\quad \left. - \exp[-\alpha_2^{-1} e^{(-\beta_2 Z)} (C_1 + C_2)] \right\} \end{aligned}$$

(II)

$$\begin{aligned} P(T_1 < C_1, T_1 + T_2 > C_1 + C_2|A_1, A_2) &= \int_0^{C_1} \int_{C_1+C_2}^{\infty} f_{T_1, T_1+T_2|A_1, A_2}(a, b|\alpha_1, \alpha_2) db da \\ &= \frac{\alpha_1^{-1} e^{(-\beta_1 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\ &\quad \times \left\{ \exp[-\alpha_2^{-1} e^{(-\beta_2 Z)} (C_1 + C_2)] \right. \\ &\quad \left. - \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1 - \alpha_2^{-1} e^{(-\beta_2 Z)} C_2] \right\} \end{aligned}$$

(III)

$$\begin{aligned}
& P(T_1 > C_1, T_1 + T_2 < C_1 + C_2 | A_1, A_2) \\
&= \int_{C_1}^{C_1+C_2} \int_{C_1}^b f_{T_1, T_1+T_2 | A_1, A_2}(a, b | \alpha_1, \alpha_2) da db \\
&= \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1] + \frac{\alpha_2^{-1} e^{(-\beta_2 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\
&\quad \times \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} (C_1 + C_2)] \\
&\quad - \frac{\alpha_1^{-1} e^{(-\beta_1 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\
&\quad \times \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1 - \alpha_2^{-1} e^{(-\beta_2 Z)} C_2]
\end{aligned}$$

(IV)

$$\begin{aligned}
& P(T_1 > C_1, T_1 + T_2 > C_1 + C_2 | A_1, A_2) \\
&= \int_{C_1+C_2}^{\infty} \int_{C_1}^b f_{T_1, T_1+T_2 | A_1, A_2}(a, b | \alpha_1, \alpha_2) da db \\
&= \frac{\alpha_1^{-1} e^{(-\beta_1 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\
&\quad \times \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1 - \alpha_2^{-1} e^{(-\beta_2 Z)} C_2] \\
&\quad - \frac{\alpha_2^{-1} e^{(-\beta_2 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\
&\quad \times \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} (C_1 + C_2)]
\end{aligned}$$

APPENDIX B. JOINT DENSITY OF THE BIVARIATE GAMMA DISTRIBUTION

By transformation and integration, we can obtain the joint density of (A_1, A_2) from the joint density of the mutually independent random variables, (V_0, V_1, V_2) . The joint density of (V_0, V_1, V_2) is

$$g_{V_0, V_1, V_2}(v_0, v_1, v_2) = \frac{v_0^{\rho-1} (v_1 v_2)^{-\rho} e^{-(v_0+v_1+v_2)}}{\Gamma(\rho)\Gamma(1-\rho)^2}$$

where $v_0, v_1, v_2 > 0$, and the joint density of (A_0, A_1, A_2) , where $A_0 = V_0$, will then be

$$\begin{aligned}
& g_{A_0, A_1, A_2}(\alpha_0, \alpha_1, \alpha_2) \\
&= g_{V_0, V_1, V_2}(\alpha_0, \alpha_1 - \alpha_0, \alpha_2 - \alpha_0) \\
&= \frac{\alpha_0^{\rho-1} \{(\alpha_1 - \alpha_0)(\alpha_2 - \alpha_0)\}^{-\rho} e^{(\alpha_0 - \alpha_1 - \alpha_2)}}{\Gamma(\rho)\Gamma(1-\rho)^2}
\end{aligned}$$

Notice that both α_1 and α_2 should be greater than α_0 by definition, and therefore α_0 is bounded by an upper limit of the minimum between α_1 and α_2 . The joint density of

(α_1, α_2) is available by integrating out α_0 , i.e.,

$$\begin{aligned}
& g_{A_1, A_2}(\alpha_1, \alpha_2) \\
&= \int_0^{\min(\alpha_1, \alpha_2)} f_{A_0, A_1, A_2}(\alpha_0, \alpha_1, \alpha_2) d\alpha_0 \\
&= \frac{e^{-(\alpha_1 + \alpha_2)}}{\Gamma(\rho)\Gamma(1-\rho)^2} \\
&\quad \times \int_0^{\min(\alpha_1, \alpha_2)} \alpha_0^{\rho-1} \{(\alpha_1 - \alpha_0)(\alpha_2 - \alpha_0)\}^{-\rho} e^{(\alpha_0)} d\alpha_0
\end{aligned}$$

APPENDIX C. SCORE EQUATIONS IN MCNR

The first partial derivatives of the complete-data log-likelihood function with respect to β_1 are computed as

$$\begin{aligned}
\frac{\partial}{\partial \beta_1} \log f &= I_{(\delta_1=1, \delta_2=1)} \left\{ \frac{1}{a-b+1} \right\} \\
&\quad \times \frac{\partial}{\partial \beta_1} (a-b+1) + I_{(\delta_1=1, \delta_2=0)} \left\{ -\frac{1}{a} \right\} \frac{\partial}{\partial \beta_1} (-a) \\
&\quad + I_{(\delta_1=0, \delta_2=1)} \left\{ \frac{1}{b+d} \right\} \frac{\partial}{\partial \beta_1} (b+d) \\
&\quad + I_{(\delta_1=0, \delta_2=0)} \left\{ -\frac{1}{d} \right\} \frac{\partial}{\partial \beta_1} (-d)
\end{aligned}$$

where

$$\begin{aligned}
(6) \quad a &= \frac{\alpha_1^{-1} e^{(-\beta_1 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\
&\quad \times \left\{ \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1 - \alpha_2^{-1} e^{(-\beta_2 Z)} C_2] \right. \\
&\quad \left. - \exp[-\alpha_2^{-1} e^{(-\beta_2 Z)} (C_1 + C_2)] \right\}
\end{aligned}$$

(7)

$$b = \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1]$$

(8)

$$\begin{aligned}
d &= \frac{\alpha_2^{-1} e^{(-\beta_2 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} (C_1 + C_2)] \\
&\quad - \frac{\alpha_1^{-1} e^{(-\beta_1 Z)}}{\alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}} \\
&\quad \times \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1 - \alpha_2^{-1} e^{(-\beta_2 Z)} C_2]
\end{aligned}$$

and $\frac{\partial}{\partial \beta_1} (b) = \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1] \alpha_1^{-1} e^{(-\beta_1 Z)} C_1 Z$. We denote $\frac{\partial}{\partial \beta_1} (a) = \frac{\partial}{\partial \beta_1} \{a1 - a2\}$ and $\frac{\partial}{\partial \beta_1} (d) = \frac{\partial}{\partial \beta_1} \{d1 - a1\}$ where $a1 = \frac{f1}{g}$, $a2 = \frac{f2}{g}$, $d1 = \frac{f3}{g}$, and

$$\begin{aligned}
f1 &= \alpha_1^{-1} e^{(-\beta_1 Z)} \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} C_1 - \alpha_2^{-1} e^{(-\beta_2 Z)} C_2] \\
f2 &= \alpha_1^{-1} e^{(-\beta_1 Z)} \exp[-\alpha_2^{-1} e^{(-\beta_2 Z)} (C_1 + C_2)] \\
f3 &= \alpha_2^{-1} e^{(-\beta_2 Z)} \exp[-\alpha_1^{-1} e^{(-\beta_1 Z)} (C_1 + C_2)] \\
g &= \alpha_1^{-1} e^{(-\beta_1 Z)} - \alpha_2^{-1} e^{(-\beta_2 Z)}
\end{aligned}$$

Then, the first partial derivatives of these sub-functions with respect to β_1 can be further derived as

$$\begin{aligned} \frac{\partial}{\partial \beta_1}(a1) &= \frac{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)} \frac{\partial}{\partial \beta_1}(f1)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ &\quad - \frac{\{\alpha_1^{-1}e^{(-\beta_1 Z)} \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}C_1 - \alpha_2^{-1}e^{(-\beta_2 Z)}C_2]\} \frac{\partial}{\partial \beta_1}(g)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ \frac{\partial}{\partial \beta_1}(a2) &= \frac{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)} \frac{\partial}{\partial \beta_1}(f2)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ &\quad - \frac{\{\alpha_1^{-1}e^{(-\beta_1 Z)} \exp[-\alpha_2^{-1}e^{(-\beta_2 Z)}(C_1 + C_2)]\} \frac{\partial}{\partial \beta_1}(g)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ \frac{\partial}{\partial \beta_1}(d1) &= \frac{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)} \frac{\partial}{\partial \beta_1}(f3)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ &\quad - \frac{\{\alpha_2^{-1}e^{(-\beta_2 Z)} \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}(C_1 + C_2)]\} \frac{\partial}{\partial \beta_1}(g)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \beta_1}(f_1) &= \alpha_1^{-1}e^{(-\beta_1 Z)} \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}C_1 - \alpha_2^{-1}e^{(-\beta_2 Z)}C_2] \\ &\quad \times \alpha_1^{-1}e^{(-\beta_1 Z)}C_1 Z \\ &\quad - \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}C_1 - \alpha_2^{-1}e^{(-\beta_2 Z)}C_2] \\ &\quad \times \alpha_1^{-1}e^{(-\beta_1 Z)} Z \\ \frac{\partial}{\partial \beta_1}(f_2) &= -\exp[-\alpha_2^{-1}e^{(-\beta_2 Z)}(C_1 + C_2)]\alpha_1^{-1}e^{(-\beta_1 Z)} Z \\ \frac{\partial}{\partial \beta_1}(f_3) &= \alpha_2^{-1}e^{(-\beta_2 Z)} \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}(C_1 + C_2)] \\ &\quad \times \alpha_1^{-1}e^{(-\beta_1 Z)}(C_1 + C_2) Z \\ \frac{\partial}{\partial \beta_1}(g) &= -\alpha_1^{-1}e^{(-\beta_1 Z)} Z \end{aligned}$$

The first partial derivatives of the complete-data log-likelihood function with respect to β_2 are

$$\begin{aligned} \frac{\partial}{\partial \beta_2} \log f &= I_{(\delta_1=1, \delta_2=1)} \left\{ \frac{1}{a-b+1} \right\} \frac{\partial}{\partial \beta_2}(a-b+1) \\ &\quad + I_{(\delta_1=1, \delta_2=0)} \left\{ -\frac{1}{a} \right\} \frac{\partial}{\partial \beta_2}(-a) \\ &\quad + I_{(\delta_1=0, \delta_2=1)} \left\{ \frac{1}{b+d} \right\} \frac{\partial}{\partial \beta_2}(b+d) \\ &\quad + I_{(\delta_1=0, \delta_2=0)} \left\{ -\frac{1}{d} \right\} \frac{\partial}{\partial \beta_2}(-d) \end{aligned}$$

where a , b , d are given in (6), (7), (8), and then $\frac{\partial}{\partial \beta_2}(b) = 0$. We again denote $\frac{\partial}{\partial \beta_2}(a) = \frac{\partial}{\partial \beta_2}\{a1 - a2\}$ and $\frac{\partial}{\partial \beta_2}(d) = \frac{\partial}{\partial \beta_2}\{d1 - a1\}$ where $a1$, $a2$, and $d1$ are as previously defined. The first partial derivatives of the sub-functions with respect

to β_2 are thus as follow:

$$\begin{aligned} \frac{\partial}{\partial \beta_2}(a1) &= \frac{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)} \frac{\partial}{\partial \beta_2}(f1)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ &\quad - \frac{\{\alpha_1^{-1}e^{(-\beta_1 Z)} \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}C_1 - \alpha_2^{-1}e^{(-\beta_2 Z)}C_2]\} \frac{\partial}{\partial \beta_2}(g)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ \frac{\partial}{\partial \beta_2}(a2) &= \frac{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)} \frac{\partial}{\partial \beta_2}(f2)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ &\quad - \frac{\{\alpha_1^{-1}e^{(-\beta_1 Z)} \exp[-\alpha_2^{-1}e^{(-\beta_2 Z)}(C_1 + C_2)]\} \frac{\partial}{\partial \beta_2}(g)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ \frac{\partial}{\partial \beta_2}(d1) &= \frac{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)} \frac{\partial}{\partial \beta_2}(f3)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \\ &\quad - \frac{\{\alpha_2^{-1}e^{(-\beta_2 Z)} \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}(C_1 + C_2)]\} \frac{\partial}{\partial \beta_2}(g)}{\alpha_1^{-1}e^{(-\beta_1 Z)} - \alpha_2^{-1}e^{(-\beta_2 Z)^2}} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \beta_2}(f_1) &= \alpha_1^{-1}e^{(-\beta_1 Z)} \exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}C_1 - \alpha_2^{-1}e^{(-\beta_2 Z)}C_2] \\ &\quad \times \alpha_2^{-1}e^{(-\beta_2 Z)}C_2 Z \\ \frac{\partial}{\partial \beta_2}(f_2) &= \alpha_1^{-1}e^{(-\beta_1 Z)} \exp[-\alpha_2^{-1}e^{(-\beta_2 Z)}(C_1 + C_2)] \\ &\quad \times \alpha_2^{-1}e^{(-\beta_2 Z)}(C_1 + C_2) Z \\ \frac{\partial}{\partial \beta_2}(f_3) &= -\exp[-\alpha_1^{-1}e^{(-\beta_1 Z)}(C_1 + C_2)]\alpha_2^{-1}e^{(-\beta_2 Z)} Z \\ \frac{\partial}{\partial \beta_2}(g) &= \alpha_2^{-1}e^{(-\beta_2 Z)} Z \end{aligned}$$

Received 28 September 2011

REFERENCES

- BOOTH, J. G., HOBERT, J. P., and JANK, W. (2001). A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. *Statistical Modeling* **1**, 333–349.
- CARLIN, B. P. and LOUIS, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edn. Chapman and Hall/CRC. [MR1427749](#)
- CHANG, S. H. (2004). Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Analysis* **10**, 175–190. [MR2081720](#)
- CHANG, S. H. and TZENG, S. J. (2006). Nonparametric estimation of sojourn time distributions for truncated serial event data—a weight-adjusted approach. *Lifetime Data Analysis* **12**, 53–67. [MR2255735](#)
- CHEN, B. E. and COOK, R. J. (2004). Tests for multivariate recurrent events in the presence of a terminal event. *Biostatistics* **5**, 129–143.
- COX, D. R. and SNELL, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)* **30**, 248–275. [MR0237052](#)
- FARRINGTON, C. P. (2000). Residuals for proportional hazards models with interval-censored survival data. *Biometrics* **56**, 473–482.

- FIELD, C. A. and WELSH, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 369–390. [MR2323758](#)
- FIOCCO, M., PUTTER, H., and VAN HOUWELINGEN, J. C. (2009). A new serially correlated gamma-frailty process for longitudinal count data. *Biostatistics* **10**, 245–257.
- GAUDERMAN, W. J. and NAVIDI, W. (2001). A Monte Carlo Newton-Raphson procedure for maximizing complex likelihoods on pedigree data. *Computational Statistics and Data Analysis* **35**, 395–415. [MR1819032](#)
- GHOSH, D. and LIN, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12**, 663–688. [MR1929958](#)
- HE, X., TONG, X., SUN, J., and COOK, R. (2008). Regression analysis of multivariate panel count data. *Biostatistics* **9**, 234–248.
- JEWELL, N. P. and VAN DER LAAN, M. J. (2004). Current status data: Review, recent developments and open problems. in: *Advances in Survival Analysis, Handbook of Statistics*, Balakrishnan, N. and Rao, C. R., Eds, 23. Elsevier North-Holland, Chapter 35, pp. 625–643. [MR2065792](#)
- KLEIN, J. P., PELZ, C., and ZHANG, M. J. (1999). Modeling random effects for censored data by a multivariate normal regression model. *Biometrics* **55**, 497–506. [MR1705173](#)
- KLEIN, J. P. and MOESCHBERGER, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd edn. Springer-Verlag, New York.
- KOMAREK, A. and LESAFFRE, E. (2006). Bayesian semi-parametric accelerated failure time model for paired doubly interval-censored data. *Statistical Modelling* **6**, 3–22. [MR2226782](#)
- KOMAREK, A. and LESAFFRE, E. (2008). Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association* **103**, 523–533. [MR2523990](#)
- LAMBERT, P., COLLETT, D., KIMBER, A., and JOHNSON, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine* **23**, 3177–3192.
- LIN, D. Y., SUN, W., and YING, Z. (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* **86**, 59–70. [MR1688071](#)
- LIU, L., WOLFE, R. A., and HUANG, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747–756. [MR2089451](#)
- LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)* **44**, 226–233. [MR0676213](#)
- MATHAI, A. M. and MOSCHOPOULOS, P. G. (1991). On a multivariate Gamma. *Journal of Multivariate Analysis* **39**, 135–153. [MR1128677](#)
- MCCULLOCH, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170. [MR1436105](#)
- SHIBOSKI, S. C. (1998). Generalized additive models for current status data. *Lifetime Data Analysis* **4**, 29–50.
- SUN, L., WANG, L., and SUN, J. (2006). Estimation of the association for bivariate interval-censored failure time data. *Scandinavian Journal of Statistics* **33**, 637–649. [MR2300908](#)
- TANNER, M. A. (1993). *Tools for Statistical Inference: Observed Data and Data Augmentation Methods*, 2nd edn. Springer, Verlag. [MR1096956](#)
- THERNEAU, T. M., GRAMBSCH, P. M., and PANKRATZ, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* **12**, 156–175. [MR1965213](#)
- TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association* **69**, 169–173. [MR0381120](#)
- VAUPEL, J. W., MANTON, K. G., and STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.
- WANG, W. and DING, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika* **87**, 879–893. [MR1813981](#)
- XUE, X. and BROOKMEYER, R. (1996). Bivariate frailty model for the analysis of multivariate survival. *Lifetime Data Analysis* **2**, 277–289.
- YASHIN, A. I. and IACHINE, I. A. (1995). Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. *Genetic Epidemiology* **12**, 529–538.
- 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.
- YUE, H. and CHAN, K. S. (1997). A dynamic frailty model for multivariate survival data. *Biometrics* **53**, 785–793.
- ZHU, L., SUN, J., SRIVASTAVA, D. K., TONG, X., LEISENRING, W., ZHANG, H., and ROBISON, L. L. (2011). Semiparametric transformation models for joint analysis of multivariate recurrent and terminal events. *Statistics in Medicine* **30**, 3010–3023. [MR2851396](#)

Ying-Fang Wang
Department of Statistics
University of California
Davis, CA 95616
USA

Liu-Chih Lo
Department of Geography
National Kaohsiung Normal University
Kaohsiung, 80201
Taiwan

Fushing Hsieh
Department of Statistics
University of California
Davis, CA 95616
USA

E-mail address: fushing@wald.ucdavis.edu