

Modeling Spatial Survival Data Using Semiparametric Frailty Models

Yi Li* and Louise Ryan**

Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute,
Boston, Massachusetts 02115, U.S.A.

* *email*: yili@jimmy.harvard.edu

** *email*: lryan@hsph.harvard.edu

SUMMARY. We propose a new class of semiparametric frailty models for spatially correlated survival data. Specifically, we extend the ordinary frailty models by allowing random effects accommodating spatial correlations to enter into the baseline hazard function multiplicatively. We prove identifiability of the models and give sufficient regularity conditions. We propose drawing inference based on a marginal rank likelihood. No parametric forms of the baseline hazard need to be assumed in this semiparametric approach. Monte Carlo simulations and the Laplace approach are used to tackle the intractable integral in the likelihood function. Different spatial covariance structures are explored in simulations and the proposed methods are applied to the East Boston Asthma Study to detect prognostic factors leading to childhood asthma.

KEY WORDS: Cholesky decomposition; Cox models; Frailty models; Identifiability; Laplace approximation; Monte Carlo simulation; Rank likelihood; Spatial covariance; Survival data.

1. Introduction

Spatially correlated survival data are frequently observed in epidemiological and social behavioral studies. For example, in a study of regional workers' smoking patterns in Massachusetts (Sorensen, Hammond, and Hebert, 1996), interest lies in ascertaining the effects of smoking cessation programs across regions, with the main outcome variable being the time for a smoker to quit smoking; and in a gerontology study conducted by Yale University (Coroni-Huntley, Osterfeld, and Taylor, 1993), the focus is on the relationship between early exposure to violence and the occurrence of disability among the elderly residing in different neighborhoods of New Haven, Connecticut. In these studies, as adjacent neighborhoods usually share similar environmental and social factors, the potential for spatial dependence exists. Hence, correct inference on the association of the main covariates with the event-specific survival times relies on careful consideration of underlying spatial correlations. For other examples of spatial survival problems, see Goldstein (1995).

The project that motivates this article is the East Boston Asthma Study (EBAS), conducted by the Channing Laboratory of Harvard Medical School to understand etiologies of rising prevalence and morbidity of childhood asthma and of the disproportionate burden among urban minority children. Subjects were enrolled at community health clinics in the east Boston area and questionnaire data, documenting asthma status and other environmental factors, were collected during regularly scheduled visits. In addition to basic demographic data, residential addresses were geocoded for each study subject. Geocoding the dataset allowed linkage with various community-level covariates to individuals in the East

Boston data set from U.S. census data at the census block level. Because children residing in nearby census blocks were often exposed to unmeasured similar physical and social environments, the investigator suspected there might exist spatial correlations across different communities. Hence, a major goal of this study was to identify significant risk factors associated with childhood asthma while taking the possible spatial correlations into account.

For clustered survival data commonly observed in biomedical studies (e.g., multicenter clinical trials and familial studies), frailty models (Clayton and Cuzick, 1985) have become increasingly popular. Essentially, these models extend the Cox proportional hazards model (Cox, 1972) by adding random effects into the baseline hazard to model the intracluster correlation. Much work has been done in this area. Nielsen et al. (1992) discussed the use of the EM algorithm for estimation in frailty models, McGilchrist (1993) and Therneau and Grambsch (2000) proposed a penalized partial likelihood approach, and Murphy (1994, 1995) and Parner (1998) studied the theoretical properties of the models; see Oakes and Jeong (1998) for a recent review on frailty models.

There has been, however, virtually no literature dealing with models for spatially correlated survival data. A key assumption on frailties in the clustered survival models is intercluster independence, which may not be adequate enough to model the complicated dependencies in spatial settings. In this article, we propose a new class of spatial survival models by extending the ordinary frailty models to accommodate spatial correlations. To draw inference, we propose a rank likelihood-based inferential procedure that is robust to the misspecification of the baseline hazard.

The rest of this article is structured as follows. We state the model in Section 2 and study the identifiability of the model in Section 3. In Section 4, we derive the marginal rank likelihood on which the inference will be based. In Section 5, we propose using Monte Carlo simulation and the Laplace approximation approach for estimation of unknown parameters. For model assessment and illustration, simulations and an application to the EBAS are performed in Sections 5 and 6. We conclude with a general discussion in Section 7.

2. Spatial Frailty Model

A spatial frailty model is formulated as follows. In each of M geographic regions (e.g., census blocks), a number of subjects, say, $n_i (i = 1, \dots, M)$, are followed until failure or censoring, whichever comes first. For each individual, along with the observed censored time $X_{ij} = \min(T_{ij}, C_{ij})$ and noncensoring indicator $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where T_{ij} and C_{ij} are underlying true survival time and censoring time, respectively, a length- q covariate vector Z_{ij} are also observed. Here $I(\cdot)$ is an indicator function. We assume that the censoring times C_{ij} are independently distributed and are independent of the T_{ij} , given the observed covariates, and that the distributions of C_{ij} do not involve in the parameters of interest. In the following exposition, the covariate Z_{ij} is assumed to be time independent, although, it should be straightforward to extend the results to accommodate time-dependent covariates. In addition, a position marker p_i (e.g., geographic coordinate) is measured for each region. For instance, in the EBAS, $M = 25$, X_{ij} is the observed asthma onset time, Z_{ij} represents the covariates of interest such as maternal asthma status, age, and race, and p_i measures the location of each census block.

Our model specifies that, conditional on the covariates and a region-specific random effect $r(p_i)$, the survival time T_{ij} is independent and has the following intensity function:

$$\lambda_{ij}\{t \mid Z_{ij}, r(p_i)\} = \lambda_0(t)e^{\beta'Z_{ij} + r(p_i)}, \tag{1}$$

where β is the fixed effects vector and $r(\cdot)$ is a mean-zero stationary Gaussian process with some basic properties, such as, for a given region indexed by a marker z , $r(z) \sim N(0, \sigma^2)$ and the covariance between two different regions, indexed by z and w , depends on their geographic distance, i.e., $\text{cov}\{r(w), r(z)\} = \sigma^2\rho(\|w - z\|, \theta)$, where $\|\cdot\|$ is an ordinary Euclidean norm. Here $\rho(\cdot)$ is the correlation function with range $[0, 1]$. We write r_i as an abbreviation for $r(p_i)$ and denote all the variance components by $\Theta = (\sigma^2, \theta)$.

There are many forms for the correlation function $\rho(d, \theta)$ that are often used in spatial settings. For example, the equicorrelated model specifies that

$$\rho(d, \theta) = \begin{cases} 1 & \text{if } d = 0, \\ \theta & \text{if } d > 0, \end{cases} \tag{2}$$

the so-called Gaussian correlation function is

$$\rho(d, \theta) = \exp\{-(d/\theta)^2\}, \tag{3}$$

and the spherical correlation in two dimensions is

$$\rho(d, \theta) = \begin{cases} 1 - \frac{2}{\pi} \left(\frac{d}{\theta} \sqrt{1 - \frac{d^2}{\theta^2}} + \sin^{-1} \frac{d}{\theta} \right) & \text{if } d \leq \theta, \\ 0 & \text{if } d > \theta. \end{cases} \tag{4}$$

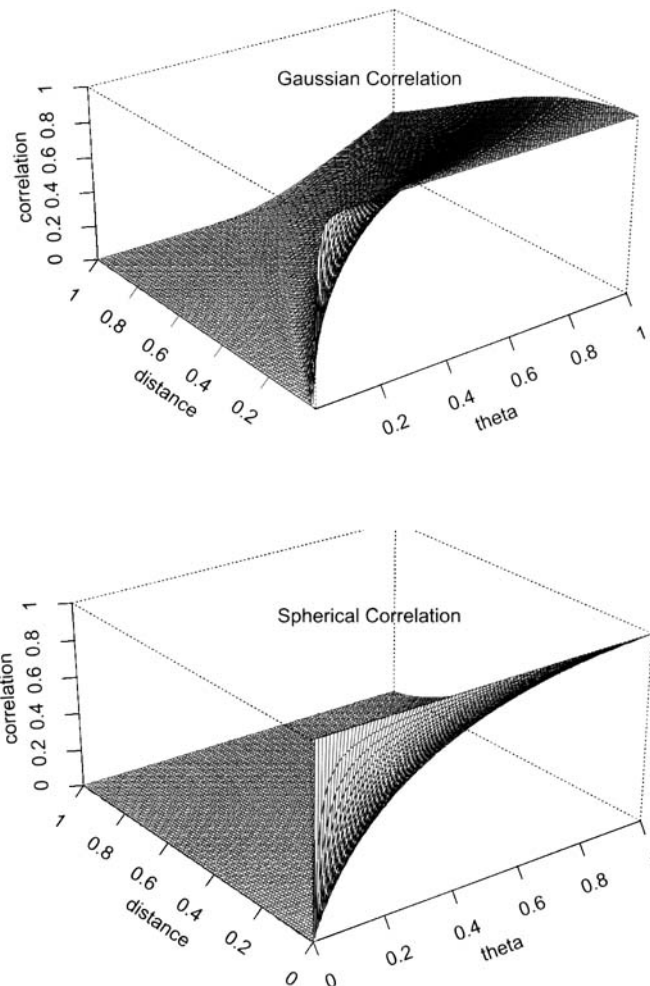


Figure 1. Correlation functions.

Ripley (1981, p. 10) showed that these are valid correlation functions, meaning that the resulting variance-covariance matrices are positive definite in some open parameter sets. One may note that, in models (3) and (4), the interregional correlation decays as the geographic distance increases and that $\theta = 0$ corresponds to interregional independence (see Figure 1). Also note that, in the equicorrelated model, negative correlations among different regions are allowed for negative θ . For a detailed discussion of other correlation functions, see Venables and Ripley (1999, p. 441).

3. Identifiability of the Model

Prior to drawing inference based on model (1), we investigate identifiability of the model, wherein the functional form of the baseline hazard is left unspecified.

The outcomes (X_{ij}, δ_{ij}) can be written in terms of the counting process $N_{ij}(t) = I(X_{ij} \leq t, \delta_{ij} = 1)$ and the at-risk process $Y_{ij}(t) = I(X_{ij} \geq t)$. Introducing the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, letting $\Omega = \{\beta, \Theta, \Lambda_0(t)\}$ be the vector of unknown parameters and Ω_0 be the true parameters, and denoting by pr_Ω the probability measure with respect to Ω , we show that the unknown Ω is

identifiable in the spatial model (1). To proceed, we postulate the following sufficient regularity conditions:

- (a) $\text{pr}_{\Omega_0} \{ \sum_{j=1}^{n_i} Y_{ij}(u) \geq 1 \text{ for all } u \in [0, \infty) \} > 0$;
- (b) $\text{pr}_{\Omega_0} \{ \sum_{j=1}^{n_i} Y_{ij}(0) \geq 2 \} > 0$;
- (c) $\text{pr}_{\Omega_0} \{ \{ \sum_{j=1}^{n_i} Y_{ij}(0) \} \{ \sum_{j=1}^{n_{i'}} Y_{i'j}(0) \} > 0, i \neq i' \} > 0$;
- (d) if $c'Z_{ij} = C$, where C is any constant independent of Z_{ij} , then $c = 0$;
- (e) if $\rho(d, \theta_1) = \rho(d, \theta_2)$ for a $d > 0$, then $\theta_1 = \theta_2$;
- (f) $p_i \neq p_{i'}$ for $i \neq i'$.

Condition (a) is a standard assumption in the conventional proportional hazards model to ensure that we can observe failures in the entire interval $[0, \infty]$ and therefore can estimate $\Lambda_0(t)$ in the entire interval. Condition (b) excludes the case where there is only one subject per region and ensures the identifiability of the variance components σ^2 (Nielsen et al., 1992), while condition (c) excludes that observations are made in only one region to guarantee the identifiability of the measure of the interregion correlation, θ . Condition (d) excludes the trivial situation where the observed covariate Z_{ij} is a constant. Condition (e) ensures the identifiability of the correlation function $\rho(d, \theta)$, and condition (f) indicates that a region is uniquely identified by its position marker. The result is summarized in the following theorem, the proof of which is deferred to the Appendix.

THEOREM 1: *If $\Lambda_0(t)$ is a nonnegative nondecreasing continuous function over $[0, \infty)$ with $\Lambda_0(0) = 0$, then, under the regularity conditions given above, the Kullback–Leibler information is strictly positive for $\Omega \neq \Omega_0$, which implies identifiability.*

4. Marginal Rank Likelihood

Oftentimes, the main objective of an epidemiological study is to investigate significant predictors for a time event, an example being the EBAS study, for which the primary goal was to explore the prognostic factors leading to childhood asthma. In such cases, instead of estimating the baseline hazard simultaneously, we treat it as a nuisance parameter and eliminate it as shown in the ensuing computation. We will report the results of simultaneously estimating the baseline hazard, regression coefficients, and variance components in a subsequent article.

It can be shown that, under model (1), only the relative rankings among all the survival times carry the information about the coefficient β , the random effect $r(p_i)$, and, hence Θ . In fact, for any strictly increasing differentiable transformation of $[0, \infty)$ onto $[0, \infty)$, say, $g(\cdot)$, $T_{ij}^* = g(T_{ij})$, where the distribution of T_{ij} follows model (1), has a conditional intensity

$$\begin{aligned} \lambda_{ij} \{ t^* \mid Z_{ij}, r(p_i) \} &= \lambda_0 \{ g^{-1}(t^*) \} \frac{dg^{-1}(t^*)}{dt^*} e^{\beta' Z_{ij} + r(p_i)} \\ &= \tilde{\lambda}_0(t^*) e^{\beta' Z_{ij} + r(p_i)}. \end{aligned}$$

Hence, (T_{ij}, Z_{ij}, p_i) and $(g(T_{ij}), Z_{ij}, p_i)$ should carry the same amount of information on β and variance component Θ . In other words, the rank statistics of the survival times, combined with the observed covariates and the regional positions, contain all such information. However, due to censoring, one

may not be able to observe the complete ranking among all the exact survival times. We hence work on all the possible rankings consistent with the observed survival times.

Let $T_{(1,0)} < \dots < T_{(L,0)}$ denote L distinct failure times and $T_{(l,k)}$, $k = 1, \dots, c_l$, be the c_l survival times censored immediately preceding $T_{(l,0)}$ but prior to $T_{(l+1,0)}$. It should be noted that, with the assumption of the continuous distribution for the survival time, the probability for ties in the survival times is zero. As indicated by Fleming and Harrington (1991), the probability of all the consistent rankings is equivalent to

$$\begin{aligned} P \{ T_{(1,0)} < \dots < T_{(L,0)}, T_{(l,k)} > T_{(l,0)}, \\ k = 1, \dots, c_l, l = 1, \dots, L \}. \end{aligned} \tag{5}$$

In this formulation, only the relative rankings among the censored survival times are left out.

Before proceeding further, we establish a one-to-one mapping from the rearranged index (l, k) ($k = 0, \dots, c_l, l = 1, \dots, L$) to the original index i, j ($j = 1, \dots, n_i, i = 1, \dots, M$), namely, $(l, k) \leftrightarrow i_{(l,k)}, j_{(l,k)}$. Note the probability that the $T_{(l,k)}$, $k = 1, \dots, c_l$, are longer than $T_{(l,0)}$, conditional on the observed covariates, the unobserved random effects, and that $T_{(l,0)} = t$, is

$$\begin{aligned} h_l(t) &= P \{ T_{(l,k)} > T_{(l,0)}, k = 1, \dots, c_l \mid \\ &\quad T_{(l,0)} = t, Z_{(l,k)}, r_{i_{(l,k)}}, k = 1, \dots, c_l \} \\ &= e^{-\Lambda_0(t)} \sum_{k=1}^{c_l} \exp \{ Z'_{(l,k)} \beta + r_{i_{(l,k)}} \} \end{aligned}$$

and the conditional density for $T_{(l,0)}$ at t is

$$\begin{aligned} f \{ t \mid Z_{(l,0)}, r_{i_{(l,0)}} \} &= \lambda_0(t) \exp \{ \beta' Z_{(l,0)} + r_{i_{(l,0)}} \} \\ &\quad \times e^{-\Lambda_0(t)} \exp \{ \beta' Z_{(l,0)} + r_{i_{(l,0)}} \}. \end{aligned}$$

Hence, some calculation shows that the rank probability (5) can be rewritten as

$$\begin{aligned} &\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_0^{\infty} \int_{t_1}^{\infty} \dots \int_{t_{L-1}}^{\infty} \prod_{l=1}^L f \{ t_l \mid Z_{(l,0)}, r_{i_{(l,0)}} \} \\ &\quad \times h_l(t_l) dt_L \dots dt_1 \\ &\quad \times dF(r_1, \dots, r_M) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{l=1}^L \frac{\exp \{ \beta' Z_{(l,0)} + r_{i_{(l,0)}} \}}{\sum_{(l',k') \in R_l} \exp \{ \beta' Z_{(l',k')} + r_{i_{(l',k')}} \}} \\ &\quad \times dF(r_1, \dots, r_M), \end{aligned}$$

where R_l is the risk set at $T_{(l,0)}$, i.e., $R_l = \{ (j, k) : j = l, \dots, L, k = 0, \dots, c_j \}$, and $F(r_1, \dots, r_M)$ is the joint distribution of the random effects. Interestingly, it turns out that the marginal rank probability is the average partial likelihood over all random effects. Indeed, Fleming and Harrington (1991) have shown that, for untied independent survival data, the partial likelihood can be derived from the rank probability. In a different context, Satten (1996) discussed the use of rank statistics to draw inference for interval-censored survival data.

In terms of risk processes, we can express the likelihood above using the original index as

$$L(\beta, \Theta) = \int \cdots \int \prod_{i=1}^M \prod_{j=1}^{n_i} \left\{ \frac{\exp(\beta' Z_{ij} + r_i)}{\sum_{i',j'} Y_{i',j'}(X_{ij}) \exp(\beta' Z_{i',j'} + r_{i'})} \right\}^{\delta_{ij}} \times dF(r_1, \dots, r_M). \tag{6}$$

If ties of the survival times are observed, we may show that (6) still holds. For example, in a series of survival times, $T_{(1,0)} < \cdots < T_{(i,0)} = T_{(i+1,0)} < \cdots < T_{(L,0)}$, two individuals fail at the same time, $T_{(i,0)} = T_{(i+1,0)}$. We then add a small perturbation $\epsilon > 0$ to $T_{(i+1,0)}$ to form $T_{(1,0)} < \cdots < T_{(i,0)} < T_{(i+1,0)} + \epsilon < \cdots < T_{(L,0)}$ and repeat the previous procedure. In the end, let $\epsilon \rightarrow 0+$, and we may conclude that (6) is valid even in this scenario. Cox (1972) and Fleming and Harrington (1991) relaxed the assumption on the underlying survival distributions to be continuous and drew the same conclusion for independent survival data.

5. Estimation

The estimates of (β, Θ) , i.e., $(\hat{\beta}, \hat{\Theta})$, are obtained by maximizing (6). Explicitly, they solve

$$\frac{\partial}{\partial \beta} \log L(\beta, \Theta) = 0, \quad \frac{\partial}{\partial \Theta} \log L(\beta, \Theta) = 0.$$

The variance estimators for $(\hat{\beta}, \hat{\Theta})$ can be conveniently calculated by inverting the observed information evaluated at $(\hat{\beta}, \hat{\Theta})$.

However, due to the intractability of the integrals, the above equations are difficult to solve using standard numerical methods. We describe below a Monte Carlo simulation method and a Laplace approximation approach to draw inference.

5.1 Monte Carlo Simulation

Our first choice is to use the Monte Carlo method to approximate $L(\beta, \Theta)$. Since r_1, \dots, r_m are not independent, directly applying a Monte Carlo simulation will be difficult. In what follows, we propose a linear transformation of random vectors to bypass this difficulty.

Write $r = (r_1, \dots, r_m)$ and denote by $Q(\Theta) = \text{cov}(r)$, with each entry $Q_{ij}(\Theta) = \text{cov}(r_i, r_j)$. Since Q is nonnegative positive and symmetric, we may perform a Cholesky decomposition of Q , i.e., we find a lower triangular matrix A such that $AA' = Q$ (Rao and Mitra, 1971). Then we write $v = (v_1, \dots, v_M)$ as a linear combination of independent random normal variables, i.e., $v = Au$, where $u = (u_1, \dots, u_M) \sim N(0, I_M)$. Here I_M is an $M \times M$ identity matrix. Since A is lower triangular, immediately $v_i = \sum_{j=1}^i A_{ij} u_j$.

It is obvious that v and r have the same distribution and (6) is equivalent to

$$\int \cdots \int \prod_{i=1}^M \prod_{j=1}^{n_i} C_{ij} \prod_{i=1}^M d\Phi(u_i), \tag{7}$$

where

$$C_{ij} = \left[\exp \left\{ \beta' Z_{ij} + \sum_{k=1}^i A_{ik}(\Theta) u_k \right\} \right.$$

$$\left. \div \sum_{i',j'} Y_{i',j'}(X_{ij}) \exp \left\{ \beta' Z_{i',j'} + \sum_{k=1}^{i'} A_{i',k}(\Theta) u_k \right\} \right]^{\delta_{ij}}$$

and $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable.

This integral still does not have a closed form. However, it can be easily approximated using a Monte Carlo method. Specifically, we independently simulate H length- M standard random normal vectors, say $u^{(b)} = \{u_1^{(b)}, \dots, u_M^{(b)}\}$, $b = 1, \dots, H$, and approximate (6) by

$$\hat{L}(\beta, \Theta) = \frac{1}{H} \sum_{b=1}^H \prod_{i=1}^M \prod_{j=1}^{n_i} D_{ij}, \tag{8}$$

where

$$D_{ij} = \left[\frac{\exp \left\{ \beta' Z_{ij} + \sum_{k=1}^i A_{ik}(\Theta) u_k^{(b)} \right\}}{\sum_{i',j'} Y_{i',j'}(X_{ij}) \exp \left\{ \beta' Z_{i',j'} + \sum_{k=1}^{i'} A_{i',k}(\Theta) u_k^{(b)} \right\}} \right]^{\delta_{ij}}$$

The log likelihood is hence approximated by the log of (8), denoted by $\log \hat{L}(\beta, \Theta)$, with its derivatives analytically available; refer to the Appendix for the formulas of $(\partial/\partial\beta) \times \log \hat{L}(\beta, \Theta)$ and $(\partial/\partial\Theta) \log \hat{L}(\beta, \Theta)$. One may use conventional numerical methods, such as the Newton-Raphson algorithm, to maximize this log likelihood to obtain $\hat{\Omega}$, the estimate of Ω . The variance and covariance matrix of $\hat{\Omega}$ can be estimated by the inverse of the observed information, which is the Hessian matrix evaluated at the final estimates. Notice that the Hessian matrix is a by-product when using Newton's method.

Theoretically, the total sampling variance of the estimator $\hat{\Omega}$ consists of two parts: one due to the use of the Monte Carlo approximation and one due to the data sampling variance (Robert and Casella, 1999). With the number of Monte Carlo simulations, H , sufficiently large, the variation due to the importance sampling can be made arbitrarily small. We will examine this in our simulations.

5.2 Laplace Approximation

As an alternative to Monte Carlo simulations, we also try the analytic likelihood approximation. Given the integral form of (6), a natural choice would be a Laplace approximation.

For notational convenience, we denote by B_i , $i = 1, \dots, M$, a length- M directional vector with the i th coordinate being one. We also introduce

$$S_r^{(l)}(t, \beta, r) = \sum_{i=1}^M \sum_{j=1}^{n_i} B_i^{\otimes l} \exp(Z'_{ij}\beta + B_i'r) Y_{ij}(t)$$

for $l = 0, 1, 2$. Here, for a vector a , $a^{\otimes l} = aa'$ if $l = 2$, $a^{\otimes l} = a$ if $l = 1$, and $a^{\otimes l} = 1$ if $l = 0$. It follows that (6) can be rewritten, up to a multiplicative constant, as

$$L(\beta, \Theta) \sim |Q(\Theta)|^{-1/2} \int e^{f(\beta,r) - \frac{1}{2}r'Q^{-1}(\Theta)r} dr, \tag{9}$$

where

$$f(\beta, r) = \sum_{i=1}^M \sum_{j=1}^{n_i} \delta_{ij} \left[Z'_{ij}\beta + B'_i r - \log \left\{ S_r^{(0)}(X_{ij}, \beta, r) \right\} \right].$$

Now, given β and Θ , we need to find the maximizer, $\hat{r}(\beta, \Theta)$, for the integrand. Explicitly, we may obtain $\hat{r}(\beta, \Theta)$ by solving

$$\frac{\partial}{\partial r} f(r, \beta) - Q^{-1}(\Theta)r = 0, \tag{10}$$

where

$$\frac{\partial}{\partial r} f(r, \beta) = \sum_{i=1}^M \sum_{j=1}^{n_i} \delta_{ij} \left\{ B_i - \frac{S_r^{(1)}(X_{ij}, \beta, r)}{S_r^{(0)}(X_{ij}, \beta, r)} \right\}.$$

Expanding the exponential part of the integrand in (9) around $\hat{r}(\beta, \Theta)$ up to the quadratic term and evaluating a normal density-type integral, we then have the first-order Laplace approximation to the log of integral (6), i.e.,

$$\begin{aligned} \log L(\beta, \Theta) \doteq & \text{const} - \frac{1}{2} \log |I_M - Q(\Theta)D(\hat{r}, \beta)| \\ & + f(\hat{r}, \beta) - \frac{1}{2} \hat{r}' Q^{-1}(\Theta) \hat{r}, \end{aligned} \tag{11}$$

where

$$\begin{aligned} D(r, \beta) &= \frac{\partial^2}{\partial r^2} f(r, \beta) \\ &= \sum_{i=1}^M \sum_{j=1}^{n_i} \delta_{ij} \left[\left\{ \frac{S_r^{(1)}(X_{ij}, \beta, r)}{S_r^{(0)}(X_{ij}, \beta, r)} \right\}^{\otimes 2} - \frac{S_r^{(2)}(X_{ij}, \beta, r)}{S_r^{(0)}(X_{ij}, \beta, r)} \right]. \end{aligned}$$

Hence, maximization of (11) with respect to β and Θ yields the Laplace estimates. In practice, this maximization procedure consists of two iterative steps: (i) starting with some initial values of β and Θ , we solve (10) for an estimate of r , and (ii) we maximize (11) with r being held constant. The iteration continues until convergence.

Following Breslow and Clayton (1993), for step (2), we made further simplifications by ignoring the first term in (11) and chose β to maximize the second because numerically we observed that the absolute value of the first term is negligible compared with that of the second. Hence, the corresponding estimating equations for β and Θ , respectively, have a simpler form,

$$\begin{aligned} \sum_{i=1}^M \sum_{j=1}^{n_i} \delta_{ij} \left\{ Z_{ij} - \frac{S_z^{(1)}(X_{ij}, \beta, \hat{r})}{S_z^{(0)}(X_{ij}, \beta, \hat{r})} \right\} &= 0, \\ -\text{tr} \left[\{I_M - QD(\hat{r}, \beta)\}^{-1} \frac{\partial Q}{\partial \Theta_j} D(\hat{r}, \beta) \right] \\ &+ \hat{r}' Q^{-1} \frac{\partial Q}{\partial \Theta_j} Q^{-1} \hat{r} = 0, \quad j = 1, 2, \end{aligned}$$

where $S_z^{(l)}(t, \beta, r) = \sum_{i=1}^M \sum_{j=1}^{n_i} Z_{ij}^{\otimes l} \exp(Z'_{ij}\beta + B'_i r) Y_{ij}(t)$ for $l = 0, 1$ and $\Theta = (\Theta_1, \Theta_2) = (\sigma^2, \theta)$. Notice that the first equation is indeed the ordinary partial likelihood score equation for the fixed effects with the regional frailties being regarded as a constant offset.

Moreover, the variance estimators for the Laplace estimates

are conveniently obtained by inverting the negative second derivative of (11), though underestimation is expected because we do not take the variability of \hat{r} on β and Θ into account. We will examine the degree of underestimation in later simulations.

An advantage of using the Laplace approximation approach is that we actually obtain the estimates of the random effects as a side product. Given the estimates of the random and fixed effects, we may also estimate, for the purpose of prediction, the cumulative baseline hazard using the Breslow estimator (Breslow, 1974), i.e.,

$$\hat{\Lambda}_0(t) = \sum_i^M \sum_j^{n_i} \int_0^t \frac{dN_{ij}(s)}{S_r^{(0)}(s, \hat{\beta}, \hat{r})},$$

and hence the predicted survivor function for a given region,

$$\hat{S}(t | X_{ij}) = e^{-\exp(X'_{ij}\hat{\beta} + \hat{r}_i)\hat{\Lambda}_0(t)}$$

As correctly pointed out by the referee, these two equalities are essentially the first-order Laplace approximation to the integrals over the distribution of the frailties, conditional on the observed ranks of survival times (Martin, 1993, p. 14). We shall explore making predictions based on model (1) in more detail in a subsequent article.

6. Simulation

Three sets of simulations have been performed to achieve three goals: (I) evaluation of the finite sample performance of the proposed methods with correctly specified frailty covariance structure, (II) study of model robustness with respect to the misspecification of the covariance functions, and (III) choice of reasonable numbers for the Monte Carlo replicates.

I. To evaluate the finite sample performance of the proposed methods, we considered the regions to be 5×5 equally divided square lattices of a unit square $[0, 1]^2$ on a two-dimensional plane. The center of each lattice was used to identify each region. We varied the number of subjects in each region, n_i , uniformly from 2 to 100. Conditional on the regional frailties, the survival times v_{ij} were generated within each region under the hazard $\lambda_{ij}(t) = \lambda_0(t) \exp\{\beta Z_{ij} + r(p_i)\}$ for $j = 1, \dots, n_i$ and $i = 1, \dots, M$, where the Z_{ij} were generated as independent $N(0, 1)$ and the distribution of the regional frailties follows models (2), (3), or (4). Censoring times c_{ij} were generated as independent uniform random variables on $[0, c]$.

We chose true parameter values as follows: the baseline hazard was $\lambda_0(t) = t$; in the frailty covariance function, $\sigma^2 = 0.5, \theta = 0.3$; we took $\beta = 1$; the values of c were chosen to yield four different censoring proportions, 20, 40, 60, and 80%. We generated 500 datasets under each combination of parameter configurations.

For each simulated dataset, we calculated the maximum (rank) likelihood estimates by the Monte Carlo method and the Laplace approach using SAS/IML. When using the Monte Carlo method, we chose the number of replications, H , to be 1000.

The averages of the estimates and the empirical standard errors are displayed in Table 1. It appears that both Monte Carlo simulation and the Laplace methods gave consistent estimates of unknown parameters. We also examined the

Table 1

Comparisons of Monte Carlo estimates and Laplace estimates. The frailty variance function is correctly specified in the calculation. The true regression coefficients are $\beta = 1$, $\sigma^2 = 0.50$, and $\theta = 0.30$. SE_e is the empirical standard error, SE_a is the standard error calculated by inverting the Fisher information.

Censoring (%)		Monte Carlo			Laplace		
		Estimate	SE_e	SE_a	Estimate	SE_e	SE_a
Equicorrelated (Model (1))							
20	β	0.994	0.093	0.091	0.999	0.097	0.087
	σ^2	0.567	0.315	0.305	0.601	0.305	0.238
	θ	0.357	0.307	0.302	0.371	0.317	0.242
40	β	1.002	0.101	0.097	1.006	0.098	0.090
	σ^2	0.537	0.317	0.311	0.557	0.327	0.254
	θ	0.282	0.359	0.351	0.269	0.369	0.269
60	β	1.014	0.103	0.099	1.024	0.108	0.102
	σ^2	0.509	0.429	0.418	0.489	0.419	0.347
	θ	0.367	0.418	0.413	0.350	0.423	0.358
80	β	1.019	0.175	0.164	1.002	0.181	0.167
	σ^2	0.560	0.453	0.437	0.560	0.453	0.377
	θ	0.342	0.433	0.417	0.342	0.423	0.369
Gaussian (Model (2))							
20	β	0.993	0.098	0.092	0.990	0.098	0.091
	σ^2	0.515	0.343	0.327	0.521	0.337	0.246
	θ	0.306	0.317	0.308	0.308	0.324	0.248
40	β	1.016	0.117	0.102	1.012	0.114	0.102
	σ^2	0.462	0.362	0.347	0.473	0.362	0.267
	θ	0.317	0.334	0.320	0.324	0.334	0.259
60	β	1.013	0.159	0.147	1.023	0.156	0.137
	σ^2	0.522	0.382	0.368	0.528	0.392	0.301
	θ	0.354	0.356	0.339	0.361	0.346	0.278
80	β	1.003	0.208	0.190	1.018	0.202	0.189
	σ^2	0.557	0.409	0.389	0.577	0.413	0.324
	θ	0.313	0.422	0.394	0.320	0.402	0.316
Spherical (Model (3))							
20	β	1.007	0.093	0.082	1.017	0.090	0.082
	σ^2	0.514	0.317	0.299	0.532	0.307	0.239
	θ	0.318	0.314	0.298	0.312	0.309	0.248
40	β	1.004	0.102	0.093	1.024	0.101	0.092
	σ^2	0.556	0.352	0.332	0.561	0.357	0.278
	θ	0.305	0.345	0.332	0.292	0.335	0.288
60	β	0.998	0.103	0.099	0.988	0.113	0.099
	σ^2	0.481	0.372	0.363	0.469	0.366	0.301
	θ	0.312	0.355	0.343	0.309	0.365	0.312
80	β	0.998	0.171	0.165	0.982	0.174	0.164
	σ^2	0.476	0.422	0.401	0.462	0.417	0.342
	θ	0.321	0.423	0.411	0.314	0.426	0.367

performance of the standard error estimators calculated by inverting the observed Fisher information. As shown in Table 1, only slight underestimation was observed in the estimated standard errors for the Monte Carlo estimates compared with the empirical standard errors. However, the estimated standard error for the Laplace estimates, especially for the variance components, had somewhat substantial underestimation, which was consistent with Ripatti and Palmgren's (2000) simulation results in the shared and hierarchical frailty models.

II. To assess the robustness of the model with respect to the variance structure, we intentionally misspecified the covariance structure in our calculation. Specifically, with the same parameter configurations as in I, we generated the survival data with the true frailty covariance function following model (4). But in our computation, we specified the covariance function to be of (2). We again calculated the estimates using both the Monte Carlo method and the Laplace approach. The results displayed in Table 2 show that the estimates of the fixed

Table 2

Comparisons of Monte Carlo estimates and Laplace estimates. The frailty variance function is incorrectly specified as model (1) in the calculation, with the true covariance function following model (2). The true regression coefficients are $\beta = 1$, $\sigma^2 = 0.5$, and $\theta = 0.30$. SE_e is the empirical standard error, SE_a is the standard error calculated by inverting the Fisher information.

Censoring (%)		Monte Carlo			Laplace		
		Estimate	SE_e	SE_a	Estimate	SE_e	SE_a
20	β	1.001	0.098	0.083	1.003	0.099	0.083
	σ^2	0.742	0.342	0.302	0.731	0.358	0.278
	θ	0.355	0.257	0.234	0.348	0.252	0.198
40	β	1.001	0.095	0.089	1.004	0.099	0.089
	σ^2	0.757	0.382	0.342	0.752	0.372	0.303
	θ	0.352	0.312	0.288	0.342	0.302	0.226
60	β	1.012	0.104	0.101	1.022	0.106	0.102
	σ^2	0.751	0.389	0.362	0.758	0.392	0.323
	θ	0.367	0.383	0.346	0.374	0.360	0.287
80	β	1.002	0.168	0.157	1.012	0.164	0.156
	σ^2	0.680	0.423	0.388	0.674	0.421	0.378
	θ	0.407	0.411	0.378	0.412	0.421	0.323

effects were quite robust to the misspecification of the covariance structure.

III. For practical purposes, we also tried to give a desirable range for the number of Monte Carlo replicates in the computation. In this set of simulations, the only change in the parameter settings compared with those in I was that we varied H from 30 to 3000. Results were summarized in Table 3. In practice, it appears that any H ranging from 500 to 1000 would be adequate. With such a choice, the inverse of the Fisher information, even without taking the Monte Carlo sampling error into account, gave good approximation to the variances of the estimates.

7. Data Analysis

We applied the proposed methods to analyze the East Boston Asthma Study introduced in Section 1. For our analysis, we focus on the assessment of how the familial history of asthma may have attributed to disparities in disease burden. In particular, the investigator was interested in the relationship between the maternal asthma status (with a variable name *mevast*, coded as 1 = ever had asthma and 0 = never had asthma) and children's asthma status, controlled for children's gender (0 = male and 1 = female) and race (0 = white and 1 = nonwhite).

A total of 753 subjects were enrolled at community health clinics throughout the east Boston area and questionnaire data was collected during regularly scheduled well-baby visits so that the asthma onset time could be identified. During the entire study period, 95 events were observed. Residential addresses were recorded and geocoded for each study subject, which yielded 25 neighborhoods (at census block level).

We fitted a spatial frailty model to the data,

$$\lambda_{ij}(t | b_i) = \lambda_0(t) \exp\{\beta_M \times \text{mevast}_{ij} + \beta_G \times \text{gender}_{ij} + \beta_R \times \text{race}_{ij} + r(p_i)\}, \quad (12)$$

where the subscripts i and j indicate the region (census block) level and the individual level, respectively, and p_i measures

the location of the i th region. For illustration purposes, we varied the covariance function of $r(\cdot)$ and calculated the estimates using both Monte Carlo and Laplace methods under models (1), (2), and (3). In addition, the naive estimates, ignoring the spatial correlations, were also computed using SAS PHREG. The results are displayed in Table 4. It appears that the two proposed methods, the Monte Carlo and the Laplace approaches, yielded similar estimates except that the Laplace approach gave moderately smaller variance estimates. Compared with those in the naive model, the estimates of the regression coefficients in model (1) were magnified with slightly larger standard errors and confirmed the significance of the positive effect of maternal asthma on childhood asthma, i.e., maternal asthma implies higher risk of childhood asthma while adjusting for the effects of gender and race. We also noted that the estimates of the regression coefficients were similar across the three covariance models, which coincides with our simulation results (see II in Section 7). Further research is needed to test for the variance components.

8. Discussion

Both marginal survival models (Wei, Lin, and Weissfeld, 1989) and frailty models are important approaches for multivariate survival data. An advantage of a frailty model over a marginal survival model is that it is often easier to make predictions based on the former model because the correlations or the unobserved latent variables are explicitly modeled through frailties. On the other hand, making predictions or kriging has always been an important component in spatial statistics. For example, in the data example presented, the investigators were also interested in the 1- or 2-year asthma-free rates (survival rates) in a given region or a new region. Hence, in this article, we are motivated to propose a semiparametric frailty model to fit spatially correlated survival data. Specifically, random effects accommodating general correlation structures were allowed to enter into the baseline hazard multiplicatively to account for the potential spatial dependence. We

Table 3

Comparisons of Monte Carlo estimates with different numbers of simulation replicates. The frailty variance function is correctly specified in the calculation. The true regression coefficients are $\beta = 1$, $\sigma^2 = 0.50$, and $\theta = 0.30$. SE_e is the empirical standard error, SE_a is the standard error calculated by inverting the Fisher information.

Censoring (%)		B = 30			B = 1000			B = 3000		
		Estimate	SE _e	SE _a	Estimate	SE _e	SE _a	Estimate	SE _e	SE _a
Equicorrelated (Model (1))										
20	β	1.034	0.123	0.103	0.994	0.093	0.091	0.995	0.092	0.091
	σ^2	0.617	0.435	0.345	0.567	0.315	0.305	0.567	0.315	0.305
	θ	0.407	0.357	0.332	0.357	0.307	0.302	0.353	0.306	0.302
40	β	1.022	0.131	0.112	1.002	0.101	0.097	1.001	0.100	0.097
	σ^2	0.567	0.377	0.331	0.537	0.317	0.311	0.525	0.314	0.310
	θ	0.252	0.405	0.381	0.282	0.359	0.351	0.285	0.357	0.352
60	β	1.041	0.143	0.119	1.014	0.103	0.099	1.012	0.102	0.099
	σ^2	0.549	0.487	0.447	0.509	0.429	0.418	0.506	0.428	0.417
	θ	0.398	0.448	0.423	0.367	0.418	0.413	0.367	0.418	0.413
80	β	1.029	0.189	0.170	1.019	0.175	0.164	1.016	0.174	0.163
	σ^2	0.578	0.495	0.457	0.560	0.453	0.437	0.552	0.451	0.436
	θ	0.367	0.483	0.449	0.342	0.433	0.417	0.337	0.430	0.418
Gaussian (Model (2))										
20	β	1.032	0.123	0.1072	0.993	0.098	0.092	0.995	0.097	0.091
	σ^2	0.555	0.383	0.357	0.515	0.343	0.327	0.514	0.340	0.328
	θ	0.336	0.362	0.338	0.306	0.317	0.308	0.303	0.315	0.307
40	β	1.047	0.137	0.112	1.016	0.117	0.102	1.008	0.115	0.101
	σ^2	0.422	0.412	0.387	0.462	0.362	0.347	0.469	0.360	0.348
	θ	0.336	0.374	0.339	0.317	0.334	0.320	0.317	0.334	0.320
60	β	1.031	0.179	0.157	1.013	0.159	0.147	1.012	0.157	0.148
	σ^2	0.542	0.443	0.408	0.522	0.382	0.368	0.518	0.380	0.369
	θ	0.369	0.396	0.369	0.354	0.356	0.339	0.355	0.353	0.341
80	β	1.043	0.248	0.217	1.003	0.208	0.190	1.002	0.206	0.193
	σ^2	0.577	0.459	0.418	0.557	0.409	0.389	0.546	0.406	0.391
	θ	0.343	0.442	0.407	0.313	0.422	0.394	0.312	0.419	0.397
Spherical (Model (3))										
20	β	1.027	0.113	0.092	1.007	0.093	0.082	1.005	0.092	0.083
	σ^2	0.544	0.357	0.319	0.514	0.317	0.299	0.514	0.317	0.299
	θ	0.338	0.346	0.318	0.318	0.314	0.298	0.314	0.312	0.297
40	β	1.043	0.122	0.099	1.004	0.102	0.093	1.003	0.101	0.094
	σ^2	0.578	0.393	0.362	0.556	0.352	0.332	0.546	0.349	0.330
	θ	0.337	0.385	0.342	0.305	0.345	0.332	0.304	0.342	0.335
60	β	0.978	0.134	0.109	0.998	0.103	0.099	0.997	0.101	0.095
	σ^2	0.451	0.412	0.383	0.481	0.372	0.363	0.486	0.367	0.358
	θ	0.342	0.425	0.364	0.312	0.355	0.343	0.310	0.351	0.346
80	β	1.038	0.189	0.169	0.998	0.171	0.165	0.999	0.170	0.164
	σ^2	0.446	0.452	0.423	0.476	0.422	0.401	0.482	0.419	0.403
	θ	0.351	0.443	0.421	0.321	0.423	0.411	0.317	0.419	0.409

have proven identifiability of the model and have proposed a semiparametric method to draw inference on regression coefficients and variance components. We have derived a marginal rank likelihood based on all the possible rankings of the survival times, which are consistent with what were observed. This may also serve, from the likelihood perspective, as a justification for the use of the penalized partial likelihood approach (McGilchrist, 1993; Therneau and Grambsch, 2000) for the ordinary frailty models.

We have utilized the Monte Carlo simulation method and

the Laplace approach for estimation of the unknown parameters. Our simulation study showed that the Monte Carlo algorithm and Laplace method perform well. In general, the estimated standard errors using the inverse of the observed information agreed well with the empirical standard errors, though underestimation in the variance estimates of the variance components was observed for the Laplace estimates.

The proof of consistency and asymptotic normality of the estimator based on the marginal rank likelihood is an open problem, but our simulation results seem to point to the

Table 4
Analysis results of the Boston Asthma Study data. Estimates were calculated by the naive (ignoring spatial correlations), the Monte Carlo simulation, and the Laplace methods. Numbers inside parentheses are estimated SEs.

Parameter	Naive	Model (1)	Model (2)	Model (3)
		Monte Carlo		
β_M	0.706 (0.283)	0.845 (0.321)	0.837 (0.342)	0.798 (0.352)
β_G	0.173 (0.206)	0.231 (0.256)	0.261 (0.242)	0.257 (0.278)
β_R	-0.135 (0.205)	-0.178 (0.243)	-0.164 (0.226)	0.157 (0.275)
σ^2	—	0.122 (0.243)	0.136 (0.276)	0.178 (0.213)
θ	—	0.321 (0.452)	0.521 (0.342)	0.221 (0.413)
		Laplace		
β_M		0.814 (0.301)	0.817 (0.321)	0.768 (0.331)
β_G		0.211 (0.246)	0.241 (0.221)	0.237 (0.248)
β_R		-0.158 (0.212)	-0.144 (0.216)	0.148 (0.235)
σ^2		0.102 (0.213)	0.113 (0.236)	0.148 (0.183)
θ		0.301 (0.422)	0.481 (0.321)	0.210 (0.391)

asymptotic validity of the proposed method. Another limitation of this study is that the model diagnostics for the proportionality of conditional hazards and the form of frailty covariance functions is not addressed. A further study is thus necessary, though our simulations suggested that the inference about the fixed effects are fairly robust to the misspecification of frailty covariance structures.

Though we only considered scalar regional frailties in the development of our methodology, it would be easy to extend to multidimensional cases. Additionally, with a careful definition of the covariance structure of the random effects, our proposed methods shall find immediate applications in the analysis of hierarchical or multilevel data.

ACKNOWLEDGEMENTS

This work was supported by NIH grants CA57253 (YL) and CA48061 (LR). We would like to thank the editor and an anonymous reviewer for constructive and helpful suggestions that improved the quality of this manuscript. We would also like to thank Rosalind Wright for providing the East Boston Asthma data.

RÉSUMÉ

Nous proposons une nouvelle classe de modèles de fragilité semi-paramétriques pour des données de survie spatiales corrélées. Plus précisément, nous développons une extension des modèles de fragilité habituels en incorporant des effets aléatoires multiplicatifs dans la fonction de risque de base, afin de tenir compte des corrélations spatiales. Nous démontrons que ces modèles sont identifiables et donnons des conditions suffisantes de régularité. Nous proposons de faire de l'inférence à partir d'une vraisemblance marginale de rang. Il n'est pas nécessaire de supposer une forme paramétrique pour le risque de base dans cette approche semi-paramétrique. Des simulations de Monte-Carlo et la technique de Laplace sont utilisées pour résoudre l'intégrale insoluble dans la fonction de vraisemblance. Différentes structures de corrélations spatiales sont examinées par simulations. Ces méthodes sont appliquées à l'étude sur l'asthme de Boston Est pour identifier les facteurs pronostiques de la survenue d'asthme dans l'enfance.

REFERENCES

- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.
- Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A* **148**, 82–117.
- Coroni-Huntley, J., Ostfeld, A. M., and Taylor, J. O. (1993). Established populations for epidemiologic studies of the elderly: Study design and methodology. *Aging (Milano)* **5**, 27–37.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- Goldstein, H. (1995). *Multilevel Statistical Models*. New York: Edward Arnold.
- Martin, A. T. (1993). *Tools for Statistical Inference*. New York: Springer-Verlag.
- McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics* **49**, 221–225.
- Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Annals of Statistics* **22**, 712–731.
- Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Annals of Statistics* **23**, 182–198.
- Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25–43.

Oakes, D. and Jeong, J. (1998). Frailty models and rank tests. *Lifetime Data Analysis* **4**, 209–228.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Annals of Statistics* **26**, 183–214.

Rao, C. R. and Mitra, S. (1971). *Generalized Inverse of Matrices and Its Applications*. New York: Wiley.

Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalty partial likelihood. *Biometrics* **56**, 1016–1022.

Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.

Satten, G. A. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika* **83**, 355–370.

Sorensen, G., Hammond, S. K., and Hebert, J. (1996). Double jeopardy: Workplace hazards and behavioral risks for crafts persons and laborers. *American Journal of Health Promotion* **10**, 355–363.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Venables, W. N. and Ripley, B. D. (1999). *Modern Applied Statistics with S-plus*. New York: Springer.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.

Received June 2001. Revised December 2001.
Accepted December 2001.

APPENDIX

Proof of Identifiability

Suppose that the true parameters are $\Omega_0 = \{\beta_0, \sigma_0^2, \theta_0, \Lambda_0^0(t)\}$. Let pr_Ω denote the probability measure with respect to Ω . It is well known that the Kullback–Leibler information is nonnegative (Bickel et al., 1993) and is equal to zero only when $\text{pr}_\Omega = \text{pr}_{\Omega_0}$.

Let $N_i(t) = \{N_{i1}(t), \dots, N_{in_i}(t)\}'$ and $Y_i(t) = \{Y_{i1}(t), \dots, Y_{in_i}(t)\}'$. Denote $(N_1, \dots, N_M)'$ by N and define Y similarly. The identifiability can thus be proven by showing that the joint distribution of (N, Y) is uniquely determined by the parameter Ω_0 , i.e., we need to show that, if the joint distributions of (N, Y) are equal under two sets of parameters, Ω_0 and $\Omega_1 = \{\beta_1, \sigma_1^2, \theta_1, \Lambda_0^1(t)\}$, then $\Omega_0 = \Omega_1$.

For an $i \in \{1, 2, \dots, M\}$, we consider the marginal intensities $N_{ij}(\cdot)$ with respect to the filtration $\sigma\{N_{ij}(s), Y_{ij}(s) \mid 0 \leq s \leq u\}$ before u reaches a failure, i.e., $u \leq \inf\{s \mid N_{ij}(s) > 0\}$. Under the two sets of parameters Ω_0 and Ω_1 , we have that $\lambda(u \mid Z_{ij}; \Omega_0) = \lambda(u \mid Z_{ij}; \Omega_1)$, i.e.,

$$E_{\Omega_0} \left\{ \lambda_0^0(u) e^{\beta_0' Z_{ij} + r_i} \mid T_{ij} \geq u, Z_{ij} \right\} = E_{\Omega_1} \left\{ \lambda_0^1(u) e^{\beta_1' Z_{ij} + r_i} \mid T_{ij} \geq u, Z_{ij} \right\}, \tag{13}$$

where T_{ij} is the true survival time for the j th observation.

Under regularity condition (b), we can consider the joint counting process for two individuals within the same region. The joint intensity of $N_{i1}(\cdot)$ and $N_{i2}(\cdot)$ with respect to the filtration $\sigma\{N_{ij}(s_j), Y_{ij}(s_j), Z_{ij}, j = 1, 2 \mid 0 \leq s_j \leq u_j\}$ for u_1 and u_2 such that $\max(u_1, u_2) \leq \inf\{s \mid N_1(s) + N_2(s) > 0\}$ is defined as

$$\lim_{\Delta u_1 \rightarrow 0, \Delta u_2 \rightarrow 0} \frac{E\{N_{i1}(u_1 + \Delta u_1)N_{i2}(u_2 + \Delta u_2) \mid \sigma\{N_{ij}(s_j), Y_{ij}(s_j), Z_{ij}, 0 \leq s_j \leq u_j, j = 1, 2\}\}}{\Delta u_1 \Delta u_2}$$

Then, under Ω_0 and Ω_1 , we have that

$$E_{\Omega_0} \left\{ \lambda_0^0(u_1)\lambda_0^0(u_2) e^{\beta_0'(Z_{i1}+Z_{i2})+2r_i} \mid T_{ij} \geq u_j, Z_j, j = 1, 2 \right\} = E_{\Omega_1} \left\{ \lambda_0^1(u_1)\lambda_0^1(u_2) e^{\beta_1'(Z_{i1}+Z_{i2})+2r_i} \mid T_{ij} \geq u_j, Z_{ij}, j = 1, 2 \right\}. \tag{14}$$

Besides, under regularity condition (c), we can consider the joint counting process for two individuals in two different regions, say i_1 and i_2 , where $i_1 \neq i_2$. The joint intensity of $N_{i_1,1}(\cdot)$ and $N_{i_2,1}(\cdot)$ with respect to the filtration $\sigma\{N_{i_1,1}(s_1), Y_{i_1,1}(s_1), N_{i_2,1}(s_2), Y_{i_2,1}(s_2), \mid 0 \leq s_j \leq u_j, j = 1, 2\}$ for u_1 and u_2 such that $\max(u_1, u_2) \leq \inf\{s \mid N_{i_1,1}(s) + N_{i_2,1}(s) > 0\}$ is defined as

$$\lim_{\Delta u_1 \rightarrow 0, \Delta u_2 \rightarrow 0} \frac{E\{N_{i_1,1}(u_1 + \Delta u_1)N_{i_2,1}(u_2 + \Delta u_2) \mid \sigma\{N_{i_k,1}(s_k), Y_{i_k,1}(s_k), Z_{i_k,1}, 0 \leq s_k \leq u_k, k = 1, 2\}\}}{\Delta u_1 \Delta u_2}$$

Then, under Ω_0 and Ω_1 , we have that

$$\begin{aligned} E_{\Omega_0} \left\{ \lambda_0^0(u_1)\lambda_0^0(u_2) e^{\beta_0'(Z_{i_1,1}+Z_{i_2,1})+r_{i_1}+r_{i_2}} \mid T_{i_k,1} \geq u_k, Z_{i_k,1}, k = 1, 2 \right\} \\ = E_{\Omega_1} \left\{ \lambda_0^1(u_1)\lambda_0^1(u_2) e^{\beta_1'(Z_{i_1,1}+Z_{i_2,1})+r_{i_1}+r_{i_2}} \mid T_{i_k,1} \geq u_k, Z_{i_k,1}, k = 1, 2 \right\}. \end{aligned} \tag{15}$$

Let $u, u_1, u_2 \rightarrow 0$ in (13), (14), and (15). We have that

$$\lambda_0^0(0)e^{\beta_0' Z_{ij}} E_{\Omega_0} \left(e^b \right) = \lambda_0^1(0)e^{\beta_1' Z_{ij}} E_{\Omega_1} \left(e^{r_i} \right), \tag{16}$$

$$\{\lambda_0^0(0)\}^2 e^{\beta'_0(Z_{i1}+Z_{i2})} E_{\Omega_0}(e^{2r_i}) = \lambda_0^1(0)\}^2 e^{\beta'_1(Z_{i1}+Z_{i2})} E_{\Omega_1}(e^{2r_i}), \tag{17}$$

$$\{\lambda_0^0(0)\}^2 e^{\beta'_0(Z_{i1,1}+Z_{i2,1})} E_{\Omega_0}(e^{r_{i1}+r_{i2}}) = \{\lambda_0^1(0)\}^2 e^{\beta'_1(Z_{i1,1}+Z_{i2,1})} E_{\Omega_1}(e^{r_{i1}+r_{i2}}). \tag{18}$$

Comparing the coefficients of Z_{ij} in (16), we have that $\beta_0 = \beta_1$. In addition, we have that, for $k = 0, 1$,

$$E_{\Omega_k}(e^{r_i}) = e^{\sigma_k^2/2}, \quad E_{\Omega_k}(e^{2r_i}) = e^{2\sigma_k^2}, \quad E_{\Omega_k}(e^{r_{i1}+r_{i2}}) = e^{\sigma_k^2\{1+\rho(\|p_{i1}-p_{i2}\|, \theta_k)\}}.$$

If we substitute these two identities into (16), (17), and (18) and use conditions (e) and (f), simple calculations give that $\sigma_0^2 = \sigma_1^2$ and $\theta_1 = \theta_2$.

We now have finished proving the identifiability of part of finite dimensional parameters in the unknown parameter vector Ω . We next show that $\Lambda_0^0(\cdot) = \Lambda_0^1(\cdot)$. Consider the survival function under Ω_k ($k = 0, 1$),

$$S(t | Z_{ij}; \Omega_k) = E_{\Omega_0} \left\{ e^{-\Lambda_0^k(t) \exp(\beta'_k Z_{ij} + r_i)} | Z_{ij} \right\} = LP \left\{ \Lambda_0^k(t) \right\},$$

where $LP(\cdot)$ is the Laplace transformation of the random variable $\exp(\beta'_k Z_{ij} + r_i)$ conditional on Z_{ij} . Then, by the invertibility of the Laplace transformation, $\Lambda_0^0(t) = \Lambda_0^1(t)$. This completes the proof of identifiability.

Derivative Formulas

Write $\hat{L}(\beta, \Theta) = \frac{1}{H} \sum_{b=1}^H \prod_{i=1}^m \prod_{j=1}^{n_i} L_{ij}^{(b)}$, where

$$L_{ij}^{(b)} = \left[\frac{\exp \left\{ \beta' Z_{ij} + \sum_{k=1}^i A_{ik}(\Theta) u_k^{(b)} \right\}}{\sum_{i',j'} Y_{i'j'}(X_{ij}) \exp \left\{ \beta' Z_{i'j'} + \sum_{k=1}^{i'} A_{i'k}(\Theta) u_k^{(b)} \right\}} \right]^{\delta_{ij}}$$

Some calculation gives

$$\frac{\partial}{\partial \beta} \log L_{ij}^{(b)} = \delta_{ij} \left[Z_{ij} - \frac{\sum_{i',j'} Y_{i'j'}(X_{ij}) Z_{i'j'} \exp \left\{ \beta' Z_{i'j'} + \sum_{k=1}^{i'} A_{i'k}(\Theta) u_k^{(b)} \right\}}{\sum_{i',j'} Y_{i'j'}(X_{ij}) \exp \left\{ \beta' Z_{i'j'} + \sum_{k=1}^{i'} A_{i'k}(\Theta) u_k^{(b)} \right\}} \right]$$

and

$$\frac{\partial}{\partial \Theta} \log L_{ij}^{(b)} = \delta_{ij} \left[\sum_{k=1}^i \frac{\partial}{\partial \Theta} A_{ik}(\Theta) u_k^{(b)} - \frac{\sum_{i',j'} Y_{i'j'}(X_{ij}) \left(\sum_{k=1}^{i'} \frac{\partial}{\partial \Theta} A_{i'k}(\Theta) u_k^{(b)} \right) \exp \left\{ \beta' Z_{i'j'} + \sum_{k=1}^{i'} A_{i'k}(\Theta) u_k^{(b)} \right\}}{\sum_{i',j'} Y_{i'j'}(X_{ij}) \exp \left\{ \beta' Z_{i'j'} + \sum_{k=1}^{i'} A_{i'k}(\Theta) u_k^{(b)} \right\}} \right]$$

Let $L^{(b)} = \prod_{i=1}^m \prod_{j=1}^{n_i} L_{ij}^{(b)}$. It then follows that

$$\frac{\partial}{\partial \beta} \log \hat{L}(\beta, \Theta) = \frac{1}{H \cdot \hat{L}(\beta, \Theta)} \sum_{b=1}^H L^{(b)} \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{\partial}{\partial \beta} \log L_{ij}^{(b)}$$

and

$$\frac{\partial}{\partial \Theta} \log \hat{L}(\beta, \Theta) = \frac{1}{H \cdot \hat{L}(\beta, \Theta)} \sum_{b=1}^H L^{(b)} \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{\partial}{\partial \Theta} \log L_{ij}^{(b)}.$$