

Semiparametric Normal Transformation Models for Spatially Correlated Survival Data

Yi Li and Xihong Lin

There is an emerging interest in modeling spatially correlated survival data in biomedical and epidemiologic studies. In this article we propose a new class of semiparametric normal transformation models for right-censored spatially correlated survival data. This class of models assumes that survival outcomes marginally follow a Cox proportional hazard model with unspecified baseline hazard, and their joint distribution is obtained by transforming survival outcomes to normal random variables, whose joint distribution is assumed to be multivariate normal with a spatial correlation structure. A key feature of the class of semiparametric normal transformation models is that it provides a rich class of spatial survival models where regression coefficients have population average interpretation and the spatial dependence of survival times is conveniently modeled using the transformed variables by flexible normal random fields. We study the relationship of the spatial correlation structure of the transformed normal variables and the dependence measures of the original survival times. Direct nonparametric maximum likelihood estimation in such models is practically prohibited due to the high-dimensional intractable integration of the likelihood function and the infinite-dimensional nuisance baseline hazard parameter. We hence develop a class of spatial semiparametric estimating equations, which conveniently estimate the population-level regression coefficients and the dependence parameters simultaneously. We study the asymptotic properties of the proposed estimators and show that they are consistent and asymptotically normal. The proposed method is illustrated with an analysis of data from the East Boston Asthma Study, and its performance is evaluated using simulations.

KEY WORDS: Asymptotic normality; Consistency; Cox model; Cross ratio; Dependence measure; Likelihood; Population-average interpretation; Semiparametric estimating equation; Spatial dependence; Spatial survival data.

1. INTRODUCTION

Biomedical and epidemiological studies have spawned an increasing interest in and practical need for developing statistical methods for modeling time-to-event data that are subject to spatial dependence. Our motivating example, the East Boston Asthma Study (EBAS) conducted by the Channing Laboratory of Harvard Medical School, aimed at understanding the etiology of the rising prevalence and morbidity of childhood asthma and the disproportionate burden among urban minority children. Subjects were enrolled at community health clinics in the east Boston area, and questionnaire data, documenting age at onset of childhood asthma and other environmental factors, were collected during regularly scheduled visits. Apart from the basic demographic data, residential addresses were geocoded for each study subject, so that the latitudes and longitudes were available. Residents of east Boston are mainly relatively low-income working families. Children residing in this area have similar social economical backgrounds and are often exposed to similar physical and social environments. These environmental factors are important triggers of asthma but are often difficult to measure in practice. Age at onset of asthma of the children in this study was hence likely to be subject to spatial correlation. The statistical challenge is to identify significant risk factors associated with age at onset of childhood asthma while taking the possible spatial correlation into account.

Prevailing modeling techniques, such as marginal models (see, e.g., Wei, Lin, and Weissfeld 1989; Prentice and Cai 1992) and frailty models (see, e.g., Murphy 1995; Parner 1998; Oakes 1989), have been successfully developed for handling clustered survival data, where individuals are grouped into independent clusters. In a marginal survival model, survival outcomes

are assumed to marginally follow a Cox proportional hazard model, whereas the within-cluster correlation is regarded as a nuisance parameter. In contrast, a frailty model directly models the within-cluster correlation using random effects or frailties, and regression coefficients typically do not have a population-average interpretation (Kalbfleisch and Prentice 2002, p. 306). However, there has been virtually no literature on modeling spatially correlated survival data, where both population-level regression coefficients and spatial dependence parameters are of interest.

Over the past two decades, spatial statistical methods have been well established for normally distributed data (Cressie 1993; Haining, Griffith, and Bennett 1989) and discrete data (Journel 1983; Cressie 1993; Carlin and Louis 1996; Diggle, Tawn, and Moyeed 1998). Statistical models for such uncensored data are often fully parameterized, and inference procedures are based on maximum likelihood (Clayton and Kaldor 1987; Cressie 1993), penalized maximum likelihood (Breslow and Clayton 1993), and Markov chain Monte Carlo (Besag, York, and Mollie 1991; Waller, Carlin, Xia, and Gelfand 1997).

Little work has been done for modeling survival data that are subject to spatial correlation, however. We are interested in developing a semiparametric likelihood model for spatially correlated survival outcomes, where observations marginally follow the Cox proportional hazard model and regression coefficients have a population-level interpretation and their joint distribution can be specified using a likelihood function that allows for flexible spatial correlation structures. But it is not straightforward to extend the existing models used for clustered survival data to spatial survival data with these features. Specifically, for clustered survival data, a semiparametric model that allows regression coefficients to have a population-level interpretation can be specified using a copula model (Oakes 1989) or a frailty model with a positive-stable frailty distribution (Hougaard 1986). Such models allow for only a simple constant correlation structure and are difficult to extend to allow

Yi Li is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115 (E-mail: yili@jimmy.harvard.edu). Xihong Lin is Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: mlin@hsph.harvard.edu). This work was supported in part by National Cancer Institute grants R01 CA95747 (to Li) and R01 CA76404 (to Lin). The authors thank the editor, an associate editor, and reviewer for their insightful comments. We also thank Dr. Rosalind Wright for providing the East Boston Asthma dataset.

for a flexible spatial correlation. For example, it is very difficult to specify a multivariate positive-stable frailty distribution in frailty models. Hence one needs to seek an alternative route to specify a semiparametric likelihood model that allows for regression coefficients to have a marginal interpretation and to allow for a flexible spatial correlation structure. From the Bayesian perspective of conditional modeling, Banerjee and Carlin (2003) and Banerjee, Carlin, and Gelfand (2004, chap. 9) considered hierarchical frailty spatial survival models. But in their models regression coefficients do not have a population-level interpretation.

In contrast to the existing methodology, in this article we develop a semiparametric normal transformation model for spatial survival data, where observations marginally follow a Cox proportional hazard model and their joint distribution is specified by transforming observations into normally distributed variables and assuming a multivariate normal distribution for the resulting transformed variables. A key feature of this model is that it provides a rich class of models for which regression coefficients have a population-level interpretation and the spatial dependence of survival times is conveniently modeled using flexible normal random fields. We investigate the relationship of the spatial correlation of the transformed normal variables and the dependence measures of the original survival times. As in the conventional Cox model, the baseline hazard function is left unspecified and is regarded as nuisance in semiparametric normal transformation models. In view of the high-dimensional integration of the likelihood function and the infinite-dimensional baseline hazard, we develop an estimation procedure for regression coefficients and spatial dependence parameters using unbiased spatial semiparametric estimating equations, in a similar spirit to the composite likelihood approach in parametric settings (Lindsay 1988; Heagerty and Lele 1998). Recently, Parner (2001) applied the composite likelihood approach to clustered survival data under a fully parameterized survival model.

The rest of the article is structured as follows. In Section 2 we introduce a semiparametric normal transformation model for spatially correlated survival data. In Section 3 we study the dependence measures of survival times under this model. In Section 4 we develop spatial semiparametric estimating equations for regression coefficients and spatial correlation parameters, and also study the asymptotic properties for the resulting estimators. In Section 5 we evaluate through simulations the finite-sample performance of the proposed method. We apply the proposed method to the analysis of data from the East Boston Asthma Study in Section 6, followed by a discussion in Section 7.

2. THE SEMIPARAMETRIC NORMAL TRANSFORMATION SPATIAL SURVIVAL MODEL

2.1 The Model

Consider in a spatial region of interest a total of m subjects who are followed up to failure or being censored, whichever comes first. For individual i ($i = 1, \dots, m$), we observe a $r \times 1$ vector of covariates \mathbf{Z}_i , an observed event time $X_i = \min(T_i, C_i)$, and a noncensoring indicator $\delta_i = I(T_i \leq C_i)$, where T_i and C_i are underlying true survival time and censoring

time and $I(\cdot)$ is an indicator function. We assume independent censoring; that is, the censoring times C_i are independent of the survival times T_i , given the observed covariates, and the distributions of C_i do not involve parameters of the true survival time model. We also assume that the maximum follow-up time is $\tau > 0$. The covariates \mathbf{Z}_i are assumed to be a predictable time-dependent process. Each individual's geographic location \mathbf{a}_i (e.g., latitude and longitude) is also documented.

Our model specifies that the survival time, T_i , marginally follows the Cox proportional hazard model,

$$\lambda\{t|\mathbf{Z}_i(\cdot)\} = \lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{Z}_i(t)}, \quad (1)$$

where $\boldsymbol{\beta}$ is a regression coefficient vector and $\lambda_0(t)$ is an unspecified baseline hazard function. The marginal model refers to the assumption that the hazard function (1) is, with respect to each individual's own filtration, $\mathcal{F}_{i,t} = \sigma\{I(X_i \leq s, \delta_i = 1), I(X_i > s), \mathbf{Z}_i(s), 0 \leq s \leq t\}$, the sigma field generated by the survival and covariate information up to time t . The regression coefficients $\boldsymbol{\beta}$ hence have a population-level interpretation.

We are interested in specifying a spatial joint likelihood model for T_1, \dots, T_m that allows T_i to marginally follow the Cox model (1) and allows for a flexible spatial correlation structure among the T_i 's. Denote by $\Lambda_i(t) = \int_0^t \lambda_i(s|\mathbf{Z}_i) ds$ the cumulative hazard and $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$ the cumulative baseline hazard. Then $\Lambda_i(T_i)$ marginally follows a unit exponential distribution, and its probit-type transformation,

$$T_i^* = \Phi^{-1}\{1 - e^{-\Lambda_i(T_i)}\}, \quad (2)$$

follows the standard normal distribution marginally, where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution. We can then conveniently impose a spatial structure on the underlying random fields of $\mathbf{T}^* = \{T_i^*, i = 1, \dots, m\}$ within the traditional Gaussian geostatistical framework. Hence such a normal transformation of the cumulative hazard provides a general framework to construct a flexible joint likelihood model for spatial survival data by preserving the Cox proportional hazards model for each individual marginally. This also provides a convenient way to generate spatially correlated survival data whose marginal distributions follow the Cox model.

Specifically, we assume \mathbf{T}^* to be a Gaussian random field, a special case of the Gibbs field (Winkler 1995), such that \mathbf{T}^* follows a joint multivariate normal distribution as

$$\mathbf{T}^* = \{T_i^*, i = 1, \dots, m\} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Gamma}), \quad (3)$$

where $\boldsymbol{\Gamma}$ is a positive definite matrix with diagonal elements being 1. Denote by θ_{ij} the (i, j) th element of $\boldsymbol{\Gamma}$. We assume that the correlation θ_{ij} between a pair of normalized survival times, say T_i^* and T_j^* , depends on their geographic locations \mathbf{a}_i and \mathbf{a}_j , that is,

$$\text{corr}(T_i^*, T_j^*) = \theta_{ij} = \theta_{ij}(\mathbf{a}_i, \mathbf{a}_j), \quad (4)$$

for $i \neq j$ ($i, j = 1, \dots, m$), where $\theta_{ij} \in (-1, 1)$. Generally a parametric model is assumed for θ_{ij} , which depends on a parameter vector $\boldsymbol{\alpha}$ as $\theta_{ij}(\boldsymbol{\alpha})$. We discuss common choices of models for $\theta_{ij}(\boldsymbol{\alpha})$ in Section 2.2.

Under noninformative censoring, the likelihood function for the unknown parameters $\{\Lambda_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\alpha}\}$, based on the observed data $(X_i, \delta_i, \mathbf{Z}_i), i = 1, \dots, m$, is

$$(-1)^{\delta_1 + \dots + \delta_m} \times \frac{\partial^{\delta_1 + \dots + \delta_m}}{\partial t_1^{\delta_1} \dots \partial t_m^{\delta_m}} \int_{\Phi^{-1}\{1 - e^{-\Lambda_m(t_m)}\}}^{\infty} \dots \int_{\Phi^{-1}\{1 - e^{-\Lambda_1(t_1)}\}}^{\infty} \psi(x_1, \dots, x_m; \boldsymbol{\Gamma}) dx_1 \dots dx_m, \tag{5}$$

evaluated at (X_1, \dots, X_m) , where $\psi(x_1, \dots, x_m; \boldsymbol{\Gamma})$ is the density function for an m -dimensional normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{\Gamma}$. A direct application of maximal likelihood estimation procedure is very difficult, if not infeasible, because of the high dimensionality of the intractable integral involved in the likelihood function and the infinite dimensionality of the nuisance baseline hazard $\Lambda_0(\cdot)$. As an alternative, we explore a spatial semiparametric estimating equation approach to draw inference in Section 4.

2.2 Specifications of the Spatial Correlation of the Transformed Times \mathbf{T}^*

Because the transformed times \mathbf{T}^* are normally distributed, a rich class of models can be used to model the spatial dependence by specifying a parametric model for θ_{ij} . For instance, we may parameterize $\theta_{ij}(\boldsymbol{\alpha}) = \rho(d_{ij}, \boldsymbol{\alpha})$, an isotropic correlation function that decays as the Euclidean distance d_{ij} between two individuals increases. A widely adopted choice for the correlation function is the Matérn function

$$\rho(d, \boldsymbol{\alpha}) = \frac{\alpha_1}{2^{\alpha_3 - 1} \Gamma(\alpha_3)} (2\alpha_2 \sqrt{\alpha_3} d)^{\alpha_3} K_{\alpha_3}(2\alpha_2 \sqrt{\alpha_3} d), \tag{6}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, α_1 is a scale parameter that corresponds to the ‘‘partial sill’’ as described by Waller and Gotway (2004, p. 279), α_2 measures the correlation decay with the distance, α_3 is a smoothness parameter, $\Gamma(\cdot)$ is the conventional gamma function, and $K_{\alpha_3}(\cdot)$ is the modified Bessel function of the second kind of order α_3 (see, e.g., Abramowitz and Stegun 1965). This spatial correlation model is rather general, with special cases including the exponential function $\rho(d, \boldsymbol{\alpha}) = \alpha_1 \exp(-d\alpha_2)$ when the smoothness parameter $\alpha_3 = .5$ and the ‘‘Gaussian’’ correlation function $\rho(d, \boldsymbol{\alpha}) = \alpha_1 \exp\{-d^2\alpha_2^2\}$ when $\alpha_3 \rightarrow \infty$ (see, e.g., Waller and Gotway 2004, p. 279). In all of these formulations, we require $0 \leq \alpha_1 \leq 1$ and $\alpha_2, \alpha_3 \geq 0$. Note that such spatial dependence models distinguish local and global spatial effects, where α_1 measures local correlation [i.e., $\alpha_1 = \lim_{d \rightarrow 0^+} \rho(d, \boldsymbol{\alpha})$], whereas α_2 controls the spatial decay over the distance. The smoothness parameter α_3 characterizes the behavior of the correlation function near the origin, but its estimation is difficult, because it requires dense space data and may even run into identifiability problems. Stein (1999) has argued that data cannot distinguish between $\alpha_3 = 2$ and $\alpha_3 > 2$. Hence we follow the strategy adopted by common spatial software (e.g., geoR) by fixing α_3 to estimate the other parameters and performing a sensitivity analysis by varying α_3 in data analysis.

3. DEPENDENCE MEASURES OF THE ORIGINAL SURVIVAL TIMES \mathbf{T}

The correlation coefficient θ_{ij} conveniently specifies the spatial correlation of the normally transformed survival times T_i^* and T_j^* through the conventional spatial correlation structure. It is of substantial interest to understand how such a correlation of the transformed times T_i^* and T_j^* implies for the dependence structure of the original survival times T_i and T_j , that is, how the dependence between the original survival times T_i and T_j depends on θ_{ij} . Two types of bivariate dependence are commonly used to describe multivariate survival times: local dependence and global dependence (Hougaard 2000). In this section we investigate these dependence measures under the semiparametric transformation model.

3.1 The Local Time Dependence Measure: The Cross-Ratio Function

Let T_1 and T_2 be arbitrary bivariate survival times. A common local dependence measure of T_1 and T_2 is the *cross-ratio*, defined as follows (Kalbfleisch and Prentice 2002):

$$c_{12}(t_1, t_2) = \frac{\lambda_1(t_1|T_2 = t_2)}{\lambda_1(t_1|T_2 \geq t_2)} = \frac{\lambda_2(t_2|T_1 = t_1)}{\lambda_2(t_2|T_1 \geq t_1)},$$

where $\lambda(\cdot|\cdot)$ denotes the conditional hazard function for a pair of survival times [e.g., (T_1, T_2)]; more specifically,

$$\lambda_1(t_1|t_2) = \lim_{dt \downarrow 0} (dt)^{-1} P(t_1 < T_1 \leq t_1 + dt | T_1 > t_1, T_2 = t_2).$$

The cross-ratio $c_{12}(t_1, t_2)$ measures the dependence of T_1 and T_2 at the time point (t_1, t_2) . If $c_{12}(t_1, t_2) = 1$, then T_1 and T_2 are independent at (t_1, t_2) . If $c_{12}(t_1, t_2) > 1$, then T_1 and T_2 are positively correlated at (t_1, t_2) , and vice versa. If $c_{12}(t_1, t_2)$ is a constant, then (T_1, T_2) follows the Clayton model (Clayton 1978).

Under the general spatial model (4) for the transformed survival times T_i^* , we are interested in investigating how the cross-ratio of any arbitrary survival time pairs T_i and T_j depends on their marginal survival functions and the spatial correlation θ_{ij} of the transformed survival times T_i^* and T_j^* . Specifically, under (4), one can easily calculate the joint tail probability function for the normally transformed survival time pair (T_i^*, T_j^*) as

$$\begin{aligned} \Psi(z_1, z_2; \theta_{ij}) &= P(T_i^* > z_1, T_j^* > z_2; \theta_{ij}) \\ &= \int_{z_1}^{\infty} \int_{z_2}^{\infty} \Phi_2\{dx_1, dx_2; \theta_{ij}\}, \end{aligned}$$

where $\Phi_2(\cdot, \cdot; \sigma)$ is the cdf for a bivariate normal vector with mean $(0, 0)$ and covariance matrix $\begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}$. It follows that the bivariate survival function for the original survival time pair (T_i, T_j) is

$$\begin{aligned} S_{ij}(t_1, t_2; \theta_{ij}) &= P(T_i > t_1, T_j > t_2; \theta_{ij}) \\ &= \Psi[\Phi^{-1}\{F_i(t_1)\}, \Phi^{-1}\{F_j(t_2)\}; \theta_{ij}], \tag{7} \end{aligned}$$

where $F_i(\cdot), F_j(\cdot)$ are the marginal cdf’s of T_i and T_j .

Equation (7) shows that the joint bivariate survival function is a functional of two marginal distributions. It follows that model (7) belongs to the common copula family (Hougaard 1986). In particular, when $\theta_{ij} = 0$, (7) becomes $\{1 - F_i(t_1)\}\{1 - F_j(t_2)\}$, corresponding to the independent case. One can easily

show that the bivariate survival function (7) approaches the upper Fréchet bound $\min\{1 - F_i(t_1), 1 - F_j(t_2)\}$ as $\theta_{ij} \rightarrow 1^-$, the independent case when $\theta_{ij} \rightarrow 0$, and the lower Fréchet bound $\max\{1 - F_i(t_1) - F_j(t_2), 0\}$ as $\theta_{ij} \rightarrow -1^+$.

Using the Cholesky decomposition and variable transformation, we can rewrite the two-dimensional integral in (7) as

$$S_{ij}(t_1, t_2; \theta_{ij}) = 1 - F_i(t_1) - \int_{\Phi^{-1}\{F_i(t_1)\}}^{\infty} \Phi \left\{ \frac{\Phi^{-1}\{F_j(t_2)\} - \theta_{ij}y}{(1 - \theta_{ij}^2)^{1/2}} \right\} d\Phi(y).$$

Some calculations show that the cross-ratio function is given by the survival functions

$$c_{ij}(t_1, t_2) = \frac{\lambda_i(t_1|T_j = t_2)}{\lambda_i(t_1|T_j \geq t_2)} = \frac{\frac{\partial^2}{\partial t_1 \partial t_2} S_{ij}(t_1, t_2; \theta_{ij}) \times S_{ij}(t_1, t_2; \theta_{ij})}{\frac{\partial}{\partial t_1} S_{ij}(t_1, t_2; \theta_{ij}) \times \frac{\partial}{\partial t_2} S_{ij}(t_1, t_2; \theta_{ij})},$$

where standard calculus gives

$$\frac{\partial}{\partial t_1} S_{ij}(t_1, t_2; \theta_{ij}) = -F_i^{(1)}(t_1) \left[1 - \Phi \left\{ \frac{\Phi^{-1}\{F_j(t_2)\} - \theta_{ij}\Phi^{-1}\{F_i(t_1)\}}{(1 - \theta_{ij}^2)^{-1/2}} \right\} \right],$$

$$\frac{\partial}{\partial t_2} S_{ij}(t_1, t_2; \theta_{ij}) = -F_j^{(1)}(t_2) \left[1 - \Phi \left\{ \frac{\Phi^{-1}\{F_i(t_1)\} - \theta_{ij}\Phi^{-1}\{F_j(t_2)\}}{(1 - \theta_{ij}^2)^{-1/2}} \right\} \right],$$

and

$$\frac{\partial^2}{\partial t_1 \partial t_2} S_{ij}(t_1, t_2; \theta_{ij}) = \frac{F_i^{(1)}(t_1)F_j^{(1)}(t_2)}{(1 - \theta_{ij}^2)^{1/2}\phi\{\Phi^{-1}\{F_j(t_2)\}\}} \times \phi \left[\frac{\Phi^{-1}\{F_j(t_2)\} - \theta_{ij}\Phi^{-1}\{F_i(t_1)\}}{(1 - \theta_{ij}^2)^{1/2}} \right].$$

Here $\phi(\cdot)$ is the density function of a standard normal random variable, and for an arbitrary function $H(\cdot)$, $H^{(1)}(\cdot)$ denotes the first derivative. These results show that the cross-ratio is fully determined by the marginal survival functions and θ_{ij} , the correlation of the corresponding normally transformed variables T_i^* and T_j^* .

To numerically illustrate the functional dependence of the cross-ratio $c_{ij}(t_1, t_2)$ on the spatial correlation coefficient of the transformed survival times θ_{ij} , Figure 1 shows the cross-ratio curve as a function of θ_{ij} when the marginal survival functions are assumed to be exponential 1. One can see that the cross-ratio $c_{ij}(t_1, t_2)$ is a nonlinear monotone increasing function of θ_{ij} . As $\theta_{ij} \rightarrow 0$, $c_{ij}(t_1, t_2) \rightarrow 1$, indicating independence of T_i and T_j .

3.2 The Global Time-Dependence Measures

An alternative measure of the dependence of an arbitrary pair of the original bivariate survival time is based on global measures, which measure the overall dependence of a pair of individuals over the entire lifespan by integrating over time. Kendall's τ and Spearman's ρ are the commonly used global

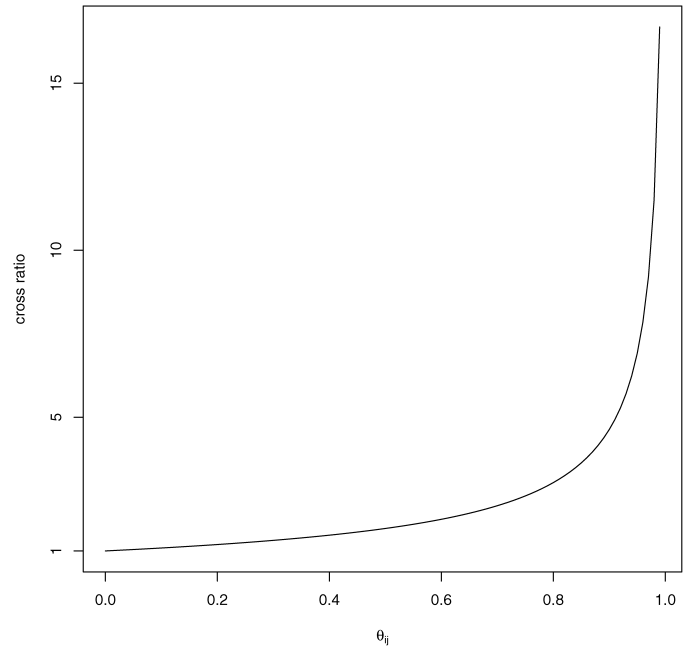


Figure 1. The Cross-Ratio at $(t_1, t_2) = (.5, .5)$ as a Function of the Correlation Coefficient of the Transformed Survival Times θ_{ij} . The individual survival times marginally follow a unit exponential distribution.

dependence measures. Both are based on concordance and discordance, and hence do not depend on the parametric forms of baseline hazard functions. They lie in $[-1, 1]$, where the value 1 corresponds to perfect concordance and the value -1 corresponds to complete discordance. Hence they are parallel to the classical correlation coefficient. However, as a global dependence measure, they are not informative about how the correlation varies with time.

Consider a copula function $C(u_1, u_2)$ such that $P(T_1 > t_1, T_2 > t_2) = C\{F_1(t_1), F_2(t_2)\}$, for a pair of nonnegative random variables T_1 and T_2 , where $F_i(\cdot)$ is the marginal cdf of T_i ($i = 1, 2$). Kendall's τ and Spearman's ρ are defined as (Kalbfleisch and Prentice 2002)

$$\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2)C(du_1, du_2) - 1$$

and

$$\rho = 12 \int_0^1 \int_0^1 C(u_1, u_2) du_1 du_2 - 3.$$

As shown in Section 2.2, the bivariate survival function of T_i and T_j under the semiparametric normal transformation model belongs to the copula family. Hence we can easily use (7) to calculate the relationships between the Kendall's τ and Spearman's ρ of the original survival times T_i and T_j and the spatial correlation θ_{ij} of the transformed time T_i^* and T_j^* as

$$\tau(\theta_{ij}) = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(z_1, z_2; \theta_{ij})\Phi_2(dz_1, dz_2; \theta_{ij}) - 1$$

and

$$\rho(\theta_{ij}) = 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(z_1, z_2; \theta_{ij})\Phi(dz_1)\Phi(dz_2) - 3,$$

where $\Psi(\cdot)$ and $\Phi_2(\cdot)$ are defined in (7). Hence Kendall's τ and Spearman's ρ are uniquely determined by the marginal survival

functions of T_i and T_j and the spatial correlation coefficient θ_{ij} of the transformed times T_i^* and T_j^* . Although the expressions of $\tau(\theta_{ij})$ and $\rho(\theta_{ij})$ do not have closed forms, both can be easily evaluated numerically. Note that both $\tau(\theta_{ij})$ and $\rho(\theta_{ij})$ approach 0 as $\theta_{ij} \rightarrow 0$, approach 1 as θ_{ij} increases to ∞ , and approach -1 as θ_{ij} decreases to $-\infty$.

4. THE SEMIPARAMETRIC ESTIMATION PROCEDURE

The likelihood function in (5) involves a high-dimensional integration, and the dimension of the required integration is the same as the sample size. In view of the numerical difficulties of directly maximizing the likelihood function, we consider spatial semiparametric estimating equations, constructed using the first two moments of individual survival times and all pairs of survival times, to estimate the regression coefficients β and the spatial correlation parameters α in $\theta_{ij}(\alpha)$.

4.1 The Martingale Covariance Rate Function

We first derive the martingale covariance rate function under the semiparametric normal transformation model (2)–(3). We define the counting process $N_i(t) = I(X_i \leq t, \delta_i = 1)$ and the at-risk process $Y_i(t) = I(X_i \geq t)$. We define a martingale, which is adapted to the filtration $\mathcal{F}_{i,t} = \sigma(N_i(s), Y_i(s), \mathbf{Z}_i(s), 0 \leq s < t)$, as

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\beta' \mathbf{Z}_i(s)} d\Lambda_0(s).$$

To relate the correlation parameters α to the counting processes, we need to consider the joint counting process of two individuals. We define the conditional martingale covariance rate function for the joint counting process of two individuals, a multidimensional generalization of the conditional hazard function, as (Prentice and Cai 1992)

$$A_{i,j}(dt_1, dt_2) = E\{M_i(dt_1)M_j(dt_2) | T_i > t_1, T_j > t_2\}.$$

Then we have

$$E\left\{M_i(t_1)M_j(t_2) - \int_0^{t_1} \int_0^{t_2} Y_i(s_1)Y_j(s_2)A_{i,j}(ds_1, ds_2)\right\} = 0.$$

We denote by $\tilde{S}_{ij}(v_1, v_2)$ the joint survival function of $\Lambda_i(T_i)$ and $\Lambda_j(T_j)$, the exponential transformations of the original survival times. Then

$$\begin{aligned} \tilde{S}_{ij}(v_1, v_2; \theta_{ij}) &= P\{\Lambda_i(T_i) > v_1, \Lambda_j(T_j) > v_2; \theta_{ij}\} \\ &= S_{ij}\{\Lambda_i^{-1}(v_1), \Lambda_j^{-1}(v_2); \theta_{ij}\}, \end{aligned} \tag{8}$$

where $S_{ij}(\cdot)$ is defined in (7). Following Prentice and Cai (1992), we can show that the covariance rate can be written as

$$A_{i,j}(dt_1, dt_2; \theta_{ij}) = A_0\{\Lambda_i(t_1), \Lambda_j(t_2); \theta_{ij}\} \Lambda_i(dt_1) \Lambda_j(dt_2),$$

where

$$\begin{aligned} A_0(v_1, v_2; \theta) &= \left\{ \frac{\partial^2}{\partial v_1 \partial v_2} \tilde{S}_{ij}(v_1, v_2; \theta) + \tilde{S}_{ij}(v_1, v_2; \theta) \right. \\ &\quad \left. + \frac{\partial}{\partial v_1} \tilde{S}_{ij}(v_1, v_2; \theta) + \frac{\partial}{\partial v_2} \tilde{S}_{ij}(v_1, v_2; \theta) \right\} \\ &\quad / \tilde{S}_{ij}(v_1, v_2; \theta). \end{aligned}$$

As a special case, $A_0(v_1, v_2; \theta = 0) \equiv 0$. A first-order approximation to $A_0(v_1, v_2; \theta)$ when θ is near 0 is given in the Appendix. It is also shown in the Appendix that as $\theta \rightarrow 0^+$, $A_0(v_1, v_2; \theta)$ converges to 0 uniformly at the same rate as that when (v_1, v_2) lies in a compact set.

4.2 The Semiparametric Estimating Equations

We simultaneously estimate the regression coefficients β (an $r \times 1$ vector) and the correlation parameters α (a $q \times 1$ vector) by considering the first two moments of the martingale vector (M_1, \dots, M_m) . In particular, for a predetermined constant $\tau > 0$ such that it is within the support of the observed failure time, that is, $P(\tau < C_i \wedge T_i) > 0$ (in practice, τ is usually the study duration), we consider the following unbiased estimating functions for $\Theta = \{\beta, \alpha\}$ for an arbitrary pair of two individuals, indexed by u and v :

- If $u = v$, then

$$\mathbf{U}_{u,u}(\Theta) = \begin{bmatrix} \int_0^\tau \mathbf{Z}_u(s) W_{(u,u)}(s) dM_u(s) \\ \mathbf{v}_{uu}\{M_u^2(\tau) - \int_0^\tau Y_u(s) d\Lambda_u(s)\} \end{bmatrix},$$

where $W_{(u,u)}(s)$ (a scalar) and \mathbf{v}_{uu} (a length- q vector) are nonrandom weights.

- If $u \neq v$, then

$$\mathbf{U}_{u,v}(\Theta) = \begin{bmatrix} \int_0^\tau \mathbf{Z}_{u,v}(s) \mathbf{W}_{(u,v)}(s) d\mathbf{M}_{u,v}(s) \\ \mathbf{v}_{uv}\{M_u(\tau)M_v(\tau) - A_{uv}\} \end{bmatrix},$$

where $\mathbf{Z}_{u,v}(s) = \{\mathbf{Z}_u(s), \mathbf{Z}_v(s)\}$, $d\mathbf{M}_{u,v}(s) = \{dM_u(s), dM_v(s)\}'$, $\mathbf{W}_{(u,v)}(s) = \{w_{ij}^{(u,v)}\}_{2 \times 2}$, and \mathbf{v}_{uv} (a length- q vector) are nonrandom weights, and

$$\begin{aligned} A_{uv} &= \int_0^\tau \int_0^\tau Y_u(s)Y_v(t) \\ &\quad \times A_0\{\Lambda_u(s), \Lambda_v(t); \theta_{uv}\} d\Lambda_u(s) d\Lambda_v(t) \\ &= \int_0^{\Lambda_u(X_u \wedge \tau)} \int_0^{\Lambda_v(X_v \wedge \tau)} A_0\{t_1, t_2; \theta_{uv}\} dt_1 dt_2. \end{aligned}$$

We show in the Appendix that A_{uv} , the covariance of martingales, decays to 0 at the same rate as the spatial correlation parameter θ_{uv} . In the Appendix we also provide a first-order approximation to A_{uv} when θ_{uv} is small.

It can be easily shown that $\mathbf{U}_{u,v}$ is an unbiased estimating function, because $E\{\mathbf{U}_{u,v}(\Theta_0)\} = 0$, where the expectation is taken under the true $\Theta_0 = (\beta_0, \alpha_0)$ and the true cumulative hazard function $\Lambda_0(\cdot)$. Note that the first component of $\mathbf{U}_{u,v}$, which is the estimating equation for β , is unbiased even when the spatial correlation structure is misspecified. Hence the regression coefficient estimator $\hat{\beta}$ is robust to misspecification of the spatial correlation structure.

Nonetheless, this estimator is not immediately computable, because $\Lambda_0(t)$ in the estimating equations is unknown. A natural alternative is to substitute it with the Breslow estimator,

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^m dN_i(s)}{\sum_{i=1}^m Y_i(s) e^{\beta' \mathbf{Z}_i(s)}}.$$

As a result, the parameters of interest $\Theta = (\beta, \alpha)$ are estimated by solving the following estimating equations, which are constructed by weightedly pooling individual martingale residuals and weightedly pooling all pairs of martingale residuals:

$$\mathbf{G}_m = m^{-1} \sum_{u \geq v} \hat{\mathbf{U}}_{u,v}(\Theta) = \mathbf{0}. \tag{9}$$

Note that $\hat{\mathbf{U}}(\cdot)$ is used to reflect the fact that $\Lambda_0(t)$ is estimated by $\hat{\Lambda}_0(t)$.

Using the matrix notation, we can conveniently express (9) as

$$m^{-1} \left[\int_0^\tau \mathbf{Z}(s) \mathbf{W} d\hat{\mathbf{M}}(s) \right. \\ \left. \hat{\mathbf{M}}'(\tau) \mathbf{V}_1 \hat{\mathbf{M}}(\tau) - \text{tr}(\mathbf{V}_j \hat{\mathbf{A}}) \right] = \mathbf{0}, \tag{10}$$

where $j = 1, \dots, q$, \mathbf{W} and \mathbf{V}_j are weight matrices, $\hat{\mathbf{M}} = (\hat{M}_1, \dots, \hat{M}_n)'$, $\mathbf{Z}(s) = \{\mathbf{Z}_1(s), \dots, \mathbf{Z}_n(s)\}'$, $\hat{\mathbf{A}}$ is an $n \times n$ matrix whose uv th ($u \neq v$) entry is \hat{A}_{uv} obtained from A_{uv} with $\Lambda_0(t)$ replaced by $\hat{\Lambda}_0(t)$, and $\hat{A}_{uu} = \int_0^\tau Y_u(s) d\hat{\Lambda}_u(s)$.

The weight matrices \mathbf{W} and $\mathbf{V}_1, \dots, \mathbf{V}_q$ are introduced to improve efficiency and convergence of the estimator of β and α . In particular, to specify \mathbf{W} , following Cai and Prentice (1997) in clustered survival data, we can specify \mathbf{W} as $(\mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2})^{-1}$, the inverse of the correlation matrix of the martingale vector $\mathbf{M}(\tau)$, where $\mathbf{D} = \text{diag}(A_{11}, \dots, A_{mm})$. In the absence of spatial dependence, \mathbf{W} is an identity matrix, and hence the first set of equations of (10) is reduced to the ordinary partial likelihood score equation for regression coefficients β . To specify \mathbf{V}_j ($j = 1, \dots, q$), we could assume that $\mathbf{V}_j = \mathbf{A}^{-1} (\partial \mathbf{A} / \partial \alpha_j) \mathbf{A}^{-1}$. Under this specification, the second set of estimating equations in (10) resembles the score equations of the variance components α if the “response” $\hat{\mathbf{M}}$ follows a multivariate normal distribution $N(\mathbf{0}, \mathbf{A})$ (Cressie 1993, p. 483).

For numerical considerations, a modification of the spatial estimating equation (10) is created by adding a penalty term,

$$\mathbf{G}_m^*(\Theta) = \mathbf{G}_m(\Theta) - \frac{1}{m} \mathbf{\Omega} \Theta,$$

where $\mathbf{\Omega}$ is a positive definite matrix, acting like a penalty term. This penalized version of the spatial estimating equation (10) can be motivated from the perspective of ridge regression or, from Bayesian perspectives, by putting a Gaussian prior $N(\mathbf{0}, \mathbf{\Omega}^{-1})$ on Θ , and results in stabilized variance component estimates of α , for example, for moderate sample sizes, and is likely to force the resulting estimates to lie in the interior of the parameter space (Heagerty and Lele 1998). Therefore, in our simulations, especially when the sample size is not large, we consider using a small penalty, $\mathbf{\Omega} = \omega \mathbf{I}$, where $0 < \omega < 1$, to ensure numerical stability. Note as the sample size m goes to ∞ , we have $\frac{1}{m} \mathbf{\Omega} \Theta \rightarrow 0$. Therefore, $\mathbf{G}_m(\Theta)$ and $\mathbf{G}_m^*(\Theta)$ are asymptotically equivalent, and the large-sample results of the original and penalized estimating equations are equivalent.

4.3 Asymptotic Properties and Variance Estimation

In this section we study the asymptotic properties of the estimators proposed in Section 4.2 and propose a finite-sample covariance estimate. Under the regularity conditions listed in the Appendix, the estimators obtained by solving $\mathbf{G}_m(\Theta) = \mathbf{0}$

exist and are consistent for the true values of $\Theta_0 = (\beta_0, \alpha_0)$, and $n^{1/2}\{\hat{\Theta} - \Theta_0\}$ is asymptotic normal with mean 0 and a covariance matrix that can be easily estimated using a sandwich estimator. The results are stated formally in Proposition 1.

Proposition 1. Assume that the true Θ_0 is an interior point of an compact set, say, $\mathcal{B} \times \mathcal{A} \in \mathbb{R}^{r+q}$, where r is the dimension of β and q is the dimension of α . Under the regularity conditions 1–5 in the Appendix, when m is sufficiently large, the estimating equation $\mathbf{G}_m(\Theta) = \mathbf{0}$ has a unique solution in a neighborhood of Θ_0 with probability tending to 1, and the resulting estimator $\hat{\Theta}$ is consistent for Θ_0 . Furthermore, $\sqrt{m}\{\Sigma^{(2)}\}^{-1/2} \Sigma\{(\hat{\beta}, \hat{\alpha})' - (\beta_0, \alpha_0)'\} \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$, where \mathbf{I} is an identity matrix whose dimension is equal to that of Θ_0 , and

$$\Sigma = \frac{1}{m} \sum_{u \geq v} E \left\{ \frac{\partial}{\partial \Theta} \mathbf{U}_{u,v}(\Theta) \right\}$$

and

$$\Sigma^{(2)} = \frac{1}{m^2} \sum_{u_1 \geq v_1} \sum_{u_2 \geq v_2} E \{ \mathbf{U}_{u_1, v_1}(\Theta_0) \mathbf{U}_{u_2, v_2}(\Theta_0) \}.$$

It follows that the covariance of $\hat{\Theta}$ can be estimated in finite samples by

$$\mathbf{I}_m^{-1} = \hat{\Sigma}^{-1} \hat{\Sigma}^{(2)} \{ \hat{\Sigma}^{-1} \}' \tag{11}$$

where $\hat{\Sigma}$ and $\hat{\Sigma}^{(2)}$ are estimated by replacing $\mathbf{U}_{uv}(\cdot)$ by $\hat{\mathbf{U}}_{uv}(\cdot)$ and evaluated at $\hat{\Theta}_0$.

Although each $E\{\hat{\mathbf{U}}_{u_1, v_1}(\Theta_0) \hat{\mathbf{U}}_{u_2, v_2}(\Theta_0)\}$ could be evaluated numerically, the total number of these calculations would be prohibitive, especially when the sample size m was large. To numerically approximate $\hat{\Sigma}^{(2)}$, we explore the resampling techniques of Carlstein (1986) and Sherman (1996). Specifically, under the assumption that, asymptotically,

$$m \times E\{\mathbf{G}_m \mathbf{G}_m'\} \rightarrow \Sigma_\infty,$$

we can estimate Σ_∞ by averaging K randomly chosen subsets of size m_j ($j = 1, \dots, K$) from the m subjects as

$$\hat{\Sigma}_\infty = K^{-1} \sum_{j=1}^K m_j \{ \hat{\mathbf{G}}_{m_j} \hat{\mathbf{G}}_{m_j}' \},$$

where $\hat{\mathbf{G}}_{m_j}$ is obtained by substituting Θ by $\hat{\Theta}$ in \mathbf{G}_{m_j} . The m_j is often chosen to be proportional to m so as to capture the spatial covariance structure. In our later simulations we chose m_j to be roughly 1/5 of the total population. Given the estimates $\hat{\Sigma}_\infty$ and $\hat{\Sigma}$, the covariance of $\hat{\Theta}$ can be estimated by $\hat{\Sigma}^{-1} [1/m \times \hat{\Sigma}_\infty] (\hat{\Sigma}^{-1})'$. For the covariance estimate of the penalized estimator obtained by solving $\mathbf{G}_m^*(\Theta) = \mathbf{0}$, $\hat{\Sigma}$ is replaced by $\hat{\Sigma} - \frac{1}{m} \mathbf{\Omega}$. A similar procedure was adopted by Heagerty and Lele (1998) for analyzing spatial binary data.

5. SIMULATION STUDY

We performed a simulation study to evaluate the finite-sample performance of the proposed methods. The locations of subjects were sampled uniformly over region $[0, m]^2$, where m is the number of subjects. The survival times T were generated marginally under the hazard model

$$\lambda(t) = \exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3\}$$

and models (2) and (3), where Z_1 and Z_2 were generated independently from the uniform distribution over $[-2, 2]$ and Z_3 was generated as a binary variable taking 0 or 1 with equal probability. The spatial dependence between two arbitrary individuals, i and j , was specified by the Matèrn function (6), where $d_{ij} = \|\mathbf{a}_i - \mathbf{a}_j\|$, $\mathbf{a}_i = (x_i, y_i)$ are the two-dimensional coordinates for subject i and $\|\cdot\|$ is the Euclidean distance. In particular, we first generated the T_{ij}^* using the multivariate normal model (3) under the Matèrn covariance matrix, then transformed the T_{ij}^* back to the original survival time scale to obtain T_{ij} using (2) and the foregoing marginal Cox model.

We set the true values $\beta_1 = 1$, $\beta_2 = .5$, $\beta_3 = .5$, $\alpha_1 = .5$, and $\alpha_2 = 2.5$. We varied α_3 in (6) to be .5 and 1. We generated censoring times c_{ij} as independent uniform random variables on $[0, 1]$ and $[0, 2]$, resulting in 70% and 50% censoring. For each set of parameters, we considered the number of subjects (m) to

be 100 and 200. We also considered $m = 400$ with $\alpha_3 = .5$ and 70% censoring. In our calculations we set the penalty parameter as $\omega = .1$. As indicated in the previous section, this penalty term was introduced to increase numerical stability by forcing the estimate to be in the interior of the parameter space.

A total of 500 simulated datasets were generated for each configuration, and the averages of the point estimates and their standard errors (SEs) were calculated, along with the coverage rates of the corresponding 95% confidence intervals. The results, summarized in Table 1, show that our estimator performed well in finite samples. The finite-sample biases of the regression coefficient estimates β were negligible, and the SE estimates agreed well with their empirical counterparts, although the coverage rates were a little below the nominal level. For the spatial correlation parameters, the performance of the estimator of α_1 was very good and similar to that of β . The

Table 1. Simulation Results Based on 500 Runs

Sample size	α_3	Censoring	Parameter	Estimate	SE_e	SE_a	Coverage probability
100	.5	70%	β_1	.9909	.2491	.2456	91.5%
			β_2	.5068	.2138	.1933	99.0%
			β_3	.5044	.2149	.1916	92.6%
			α_1	.4789	.1827	.1915	89.2%
			α_2	2.0555	.9275	.7994	73.0%
	.5	50%	β_1	.9920	.1971	.2033	92.9%
			β_2	.5134	.1731	.1628	92.5%
			β_3	.4831	.1702	.1548	90.4%
			α_1	.4656	.1520	.1533	90.8%
			α_2	2.1292	.9958	.8916	79.0%
	1	70%	β_1	.9836	.2511	.2467	90.3%
			β_2	.5112	.2127	.1897	89.1%
			β_3	.5113	.2066	.1936	91.5%
			α_1	.4767	.1814	.1935	90.6%
			α_2	2.3043	.9685	.7941	71.9%
	1	50%	β_1	1.007	.2052	.2114	91.2%
			β_2	.5139	.1845	.1659	90.6%
			β_3	.4986	.1699	.1566	91.2%
			α_1	.4796	.1539	.1507	88.0%
			α_2	2.358	1.0239	.8176	74.3%
200	.5	70%	β_1	.9869	.1556	.1702	94.7%
			β_2	.4940	.1341	.1312	92.8%
			β_3	.4882	.1421	.1323	92.8%
			α_1	.4902	.1400	.1397	92.2%
			α_2	2.3575	1.0620	.9496	80.4%
	.5	50%	β_1	.9819	.1300	.1441	95.1%
			β_2	.4951	.1133	.1100	93.8%
			β_3	.4792	.1223	.1100	90.2%
			α_1	.4990	.1011	.1094	92.2%
			α_2	2.4218	.9966	.8051	82.4%
	1	70%	β_1	.9993	.1740	.1688	91.9%
			β_2	.4909	.1387	.1245	92.3%
			β_3	.4960	.1303	.1296	92.5%
			α_1	.5118	.1395	.1380	90.0%
			α_2	2.6356	1.1288	.9811	80.1%
	1	50%	β_1	.9838	.1437	.1395	92.1%
			β_2	.4830	.1204	.1066	90.2%
			β_3	.4803	.1174	.1076	90.2%
			α_1	.5136	.1113	.1042	92.7%
			α_2	2.5033	1.0106	.7608	84.1%
400	.5	70%	β_1	.9616	.1055	.1285	94.0%
			β_2	.4960	.1040	.1032	96.0%
			β_3	.5083	.0970	.0990	94.0%
			α_1	.5077	.1093	.0963	94.2%
			α_2	2.4439	1.2866	1.0205	90.4%

NOTE: Estimates were calculated using the spatial semiparametric estimating equation method assuming the Matèrn correlation structure with 70% and 50% censoring proportions. The true parameters are $\beta_1 = 1$, $\beta_2 = \beta_3 = .5$, $\alpha_1 = .5$, and $\alpha_2 = 2.5$. Both the empirical (SE_e) and estimated (SE_a) SEs are reported, along with the 95% coverage probabilities.

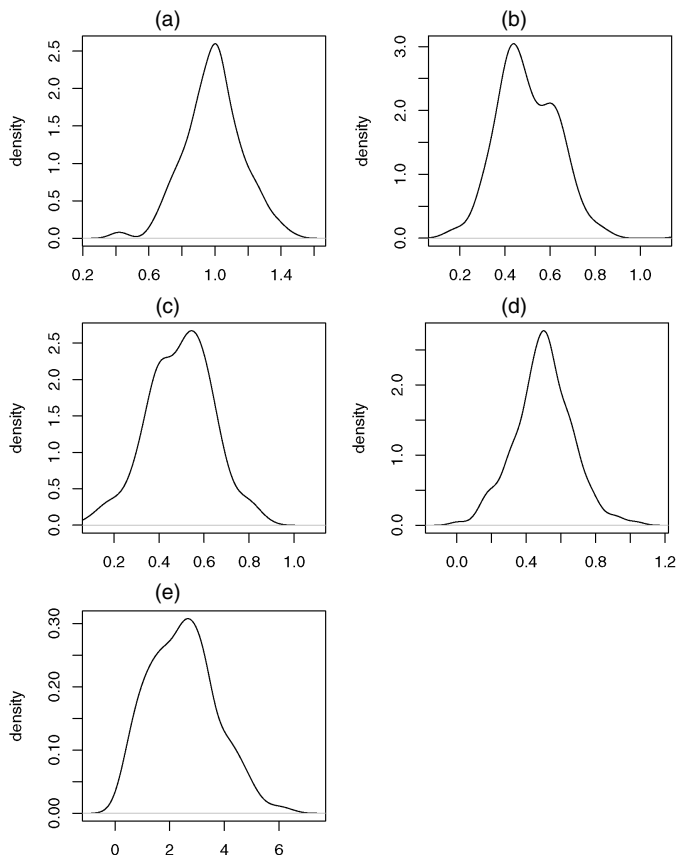


Figure 2. Empirical Density Plots of the Model Parameter Estimates From the Simulation Study With $m = 200$, $\alpha_3 = .5$, and Censoring Proportion = 50%. (a) β_1 ; (b) β_2 ; (c) β_3 ; (d) α_1 ; (e) α_2 .

estimate of α_2 had slightly more bias, and its estimated SE underestimated its true SE, resulting in a worse coverage probability. This indicates α_2 is more difficult to estimate for small samples. As the sample size increased, the biases decreased, and all of the estimates quickly approached the true values, the estimated and empirical SEs became very close, and the coverage rates became closer to the nominal level. Figure 2 depicts the estimated density plots of the parameter estimates when $m = 200$, $\alpha_3 = .5$, and the censoring proportion = 70%. They indicated that the estimates were approximately normally distributed in finite samples. These empirical results support our asymptotic findings.

To assess the robustness of the model with respect to the parameterization of the spatial dependence, we conducted an additional simulation study by intentionally misspecifying the correlation model (4) in our calculations. Specifically, using the same parameter configurations as before, with $m = 100$ and censoring proportion = 70%, we generated the survival data with the spatial dependence specified by the “spherical” correlation,

$$\rho(d) = .5 \left(1 - \frac{3d}{4} + \frac{d^3}{8} \right) I(d \leq 2),$$

but assumed the Matérn correlation (6) in our estimation. Although the estimates of the spatial dependence parameters were biased due to the misspecification of the spatial correlation structure, the estimates of the regression coefficients were still

close to the true values. The averages of the point estimates were .9950, .5232, and .5028, which were close to the true values. These results support our theoretical findings.

6. ANALYSIS OF THE EAST BOSTON ASTHMA DATA

We applied the proposed method to analyze the East Boston Asthma data introduced in Section 1. For our analysis, we focused on assessing how the familial history of asthma may have attributed to disparity in disease burden. In particular, the investigator was interested in the relationship between the low respiratory index (LRI) in the first year of life, ranging from 0 to 16, with high values indicating worse respiratory functioning, and the age at onset of childhood asthma, controlling for maternal asthma status (MEVAST), which was coded as 1 = ever had asthma and 0 = never had asthma, and log-transformed maternal cotinine levels (LOGMCOT). Such an investigation would help the investigator better understand the natural history of asthma and its associated risk factors and thus develop future intervention programs.

Subjects were enrolled at community health clinics throughout the east Boston area, and questionnaire data were collected during regularly scheduled well-baby visits, so that the age at onset of asthma could be identified. Residential addresses were recorded and geocoded. The geographic distance was calculated in kilometers. A total of 606 subjects with complete information on latitude and longitude were included in the analysis, with 74 events observed at the end of the study. The median follow-up was 5 years. East Boston is a residential area of relatively low-income working families. Participants in this study were largely white and Hispanic children ranging in age from infancy to age 6 years. Asthma is a disease strongly affected by environmental triggers. Because the children had similar backgrounds and living environments and were exposed to similar unmeasured physical and social environments, their age at onset of asthma was likely to be subject to spatial correlation.

We considered the spatial semiparametric normal transformation model and assumed that the age at onset of asthma marginally followed the Cox model,

$$\lambda(t) = \lambda_0(t) \exp\{\beta_L \times \text{LRI} + \beta_M \times \text{MEVAST} + \beta_C \times \text{LOGMCOT}\}. \quad (12)$$

We assumed the Matérn model (6) for the spatial dependence. We estimated the regression coefficients and the correlation parameters using the spatial semiparametric estimating equation approach proposed in Section 4.2, and calculated the associated SE estimates (11). For checking the robustness of the method, we also varied the smoothness parameter α_3 in (6) to be .5, 1, and 1.5.

Because the East Boston Asthma Study was conducted in a fixed region, to examine the performance of the variance estimator in (11), which was developed under the increasing-domain asymptotic, we also calculated the variance using a “delete-a-block” jackknife method (see, e.g., Kott 1998). Specifically, we divided the samples into B nonoverlapping blocks based on their geographic proximity and then formed B jackknife replicates, where each replicate was formed by deleting one of the blocks from the entire sample. For each replicate, we computed the estimates based on the semiparametric

estimating equations developed in Section 4.2 and obtained the jackknife variance as

$$\text{var}_{\text{jackknife}} = \frac{B-1}{B} \sum_{j=1}^B (\hat{\Theta}_j - \hat{\Theta})(\hat{\Theta}_j - \hat{\Theta})', \quad (13)$$

where $\hat{\Theta}_j$ is the estimate produced from the jackknife replicate with the j th “group” deleted and $\hat{\Theta}$ is the estimate based on the entire population. We chose $B = 40$, which appeared large enough to render a reasonably good measure of variability. This jackknife scheme, in a spirit similar to that of a subsampling scheme proposed by Carlstein (1986, 1988), treated each block as approximately independent and seemed plausible for this dataset, especially in the presence of weak spatial dependence. Loh and Stein (2004) called this scheme the *splitting method* and found that it worked even better than more-complicated block-bootstrapping methods (e.g., Kunsch 1998; Liu and Singh 1992; Politis and Romano 1992; Bullmann and Kunsch 1995). Other advanced resampling schemes for spatial data are also available, including a double-subsampling method (Lahiri, Kaiser, Cressie, and Hsu 1999; Zhu and Morgan 2004) and linear estimating equation jackknifing (Lele 1991), but these are subject to a much greater computational burden compared with the simple that jackknife scheme we used.

The results are presented in Table 2, with the large-sample SEs (SE_a) computed using the method described in Section 4.3 and the jackknife SEs (SE_j) computed using (13). The estimates of the regression coefficients and their SEs were almost constant with various choices of the smoothness parameter α_3 and indicated that the regression coefficient estimates were not sensitive to the choice of α_3 in this dataset. The SEs obtained from the large-sample approximation and the jackknife method were reasonably similar. Low respiratory index was highly associated with the age at onset of asthma, for example, $\hat{\beta}_L = .3121$ ($SE_a = .0440$, $SE_j = .0357$) when $\alpha_3 = .5$, $\hat{\beta}_L = .3118$ ($SE_a = .0430$, $SE_j = .0369$) when $\alpha_3 = 1.0$, and $\hat{\beta}_L = .3124$ ($SE_a = .0432$, $SE_j = .0349$) when $\alpha_3 = 1.5$, indicating that a child with a poor respiratory function was more likely to develop asthma, after controlling for maternal asthma, and maternal cotinine levels, and accounting for the spatial variation. No significant association was found between age at onset of asthma and maternal asthma and cotinine levels. The estimates of the spatial dependence parameters, α_1 and α_2 , varied slightly with the choice of α_3 . The scale parameter α_1 corresponds to the partial sill (Waller and Gotway 2004, p. 279) and measures the correlation between subjects in close geographic proximity. Our analysis showed that such a correlation was small. The parameter α_2

measures global spatial decay of dependence with the spatial distance (measured in kilometers). For example, when $\alpha_3 = .5$ (i.e., under the exponential model), $\alpha_2 = 2.2977$ means that the correlation decays by $1 - \exp(-2.2977 \times 1) \doteq 90\%$ for every 1-km increase in distance. As pointed out by a reviewer, the value of α_2 should be interpreted with caution, because its interpretation depends on the unit of distance.

7. DISCUSSION

We have proposed a semiparametric normal transformation model for spatial survival data. Although statistical methods for clustered survival data and noncensored spatial data have been well developed, little work has been done on modeling censored spatial survival data. However, direct extensions of models for clustered survival data to censored spatial survival data are difficult to use in constructing a semiparametric likelihood to allow each survival outcome to marginally follow the Cox proportional hazard model. An attractive feature of our semiparametric normal transformation models is that they provide a general semiparametric likelihood framework within which to generate censored spatial survival data with a flexible spatial correlation structure and individual observations marginally following the Cox proportional hazard model. Hence such models provide an elegant connection between classical spatial models for normal continuous spatial outcomes and the traditional Cox model for censored survival data, and allow the regression coefficients to have marginal interpretations. To our knowledge, this article is the first attempt to develop such semiparametric marginal models for spatial survival data.

In view of the intractable high-dimensional integration required by maximum likelihood estimation and the presence of the infinite-dimensional nuisance baseline hazard parameter in the likelihood function, we develop a class of spatial semiparametric estimating equations using individual and pairwise survival times. The proposed method is computationally easy and is shown to yield consistent and asymptotically normal estimators and to yield a regression coefficient estimator that is robust to misspecification of the correlation structure. Our simulation study shows that the proposed method performs well in finite samples.

The estimating equation for the spatial correlation parameter α mimics the normal-likelihood score equation for martingale residuals. It would be of interest to develop quasi-likelihood-type estimating equations to improve the efficiency of the estimator of α as it characterizes the underlying spatial dependence, which is sometimes of practical interest. Such a quasi-likelihood-type estimating equation for α , however,

Table 2. Results of Analysis of the East Boston Asthma Study Under the Normal Transformation Model Assuming the Matérn Correlation and the Marginal Cox Model

Parameters	$\alpha_3 = .5$			$\alpha_3 = 1$			$\alpha_3 = 1.5$		
	Estimate	SE_a	SE_j	Estimate	SE_a	SE_j	Estimate	SE_a	SE_j
β_L	.3121	.0440	.0357	.3118	.0430	.0369	.3124	.0432	.0349
β_M	.2662	.3314	.3222	.2644	.3289	.3309	.2676	.3283	.3340
β_C	.0294	.1394	.1235	.02521	.1270	.1063	.0277	.1288	.1083
α_1	1.68E-3	9.8E-3	.0127	.74E-3	5.0E-3	7.1E-3	.72E-3	5.5E-3	4.8E-3
α_2	2.2977	4.974	3.708	2.1917	4.7945	4.1988	1.8886	6.5005	5.01617

NOTE: Estimates were calculated by the spatial semiparametric estimating equation method, the large-sample SEs (SE_a) were computed using the method described in Section 4.3, and the jackknife SEs (SE_j) were computed using the formulation (13) in Section 6.

would involve third- and fourth-order moments of the martingale residuals $M_{uv}(s)$, the computation of which can be difficult. It would be of future research interest to investigate the efficiency loss of the proposed estimator of α relative to such a more-complicated quasi-likelihood estimator.

Although rather computationally demanding, it might be feasible to develop a full nonparametric maximum likelihood estimator (MLE) of the regression coefficient estimator β and the spatial correlation parameter α based on the semiparametric normal transformation likelihood (5), with the baseline hazard estimated nonparametrically by a step function with jumps at distinct failure times. For example, an EM-type analysis under (5) might be possible by viewing the censoring-prone survival times as missing values. It would be of future research interest to study the theoretical properties of such nonparametric MLEs and compare the efficiency and robustness of the spatial semiparametric estimating equation-based estimators in this article with the nonparametric MLEs. It is likely that the nonparametric MLEs of the regression coefficients are sensitive to the misspecification of the spatial correlation structure, whereas the spatial semiparametric estimating equation-based estimators are robust to such misspecifications. But if the semiparametric normal transformation model is a true model, then the spatial semiparametric estimating equation-based estimators may be less efficient than the nonparametric MLEs. More research is needed.

In this article we have focused on normal transformation models assuming a marginal Cox proportional hazard model in view of the popularity of the Cox model in health sciences research and the attractive interpretation of regression coefficients. We may extend the normal transformation model to the accelerated failure time models that specify

$$\log T_i = -\beta'Z_i + \epsilon_i, \quad i = 1, \dots, m,$$

where ϵ_i follows an unspecified distribution. This model is equal to, marginally, $T_i \sim S_0(t \exp(\beta'Z_i))$, where $S_0(t)$ is an unspecified survival function. Then we define the normal transformation as $T_i^* = \Phi^{-1}\{1 - S_0(t \exp(\beta'Z_i))\}$. Hence T_i^* follows the standard normal distribution marginally. We can then conveniently impose a spatial structure on the underlying random fields of $\mathbf{T}^* = \{T_i^*, i = 1, \dots, m\}$ within the traditional Gaussian geostatistical framework as described in Section 2. However, further research is needed for drawing inference based on this new class of models, because the proposed martingale-based estimating equations in Section 4.2 are not directly available to fit this model, especially in the presence of unknown baseline survival function $S_0(\cdot)$. A rank-based procedure along the line of that of Jin, Lin, and Wei (2003) may need to be adopted. We will pursue this idea in future work.

APPENDIX: TECHNICAL DETAILS

A.1 A First-Order Expansion of the Martingale Covariance Rate Function

Following Moran (1983) and Kotz, Balakrishnan, and Johnson (2000, eq. 45.89), some algebra shows that when θ is sufficiently small, we can approximate the following bivariate tail probability:

$$\Psi(z_1, z_2; \theta) = \int_{z_1}^{\infty} \int_{z_2}^{\infty} \Phi_2(dx_1, dx_2; \theta)$$

by

$$\Psi(z_1, z_2; \theta) = \{1 - \Phi(z_1)\}\{1 - \Phi(z_2)\} + \theta\phi(z_1)\phi(z_2) + o(\theta),$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and density function for a standard normal distribution and $o(\theta)$ holds uniformly with respect to (z_1, z_2) in any bounded set. Then

$$\frac{\partial}{\partial z_1} \Psi(z_1, z_2; \theta) = -[\phi(z_1)\{1 - \Phi(z_2)\} + \theta z_1 \phi(z_1)\phi(z_2)] + o(\theta),$$

$$\frac{\partial}{\partial z_2} \Psi(z_1, z_2; \theta) = -[\phi(z_2)\{1 - \Phi(z_1)\} + \theta z_2 \phi(z_1)\phi(z_2)] + o(\theta),$$

and

$$\frac{\partial^2}{\partial z_1 \partial z_2} \Psi(z_1, z_2; \theta) = \phi(z_1)\phi(z_2) + \theta z_1 z_2 \phi(z_1)\phi(z_2) + o(\theta). \quad (A.1)$$

Using a copula representation and a first-order Taylor expansion, Sungur (1990) also derived (A.1) for approximating the standard bivariate normal density function.

Hence, from (7) and (8), we can approximate $\tilde{S}_{ij}(t_1, t_2; \theta)$, the joint survival function of the exponential transformations of the original survival times, by

$$\begin{aligned} \tilde{S}_{ij}(t_1, t_2; \theta) &= \Psi\{\Phi^{-1}(1 - e^{-t_1}), \Phi^{-1}(1 - e^{-t_2}); \theta\} \\ &= e^{-(t_1+t_2)} + \theta\phi(x_1)\phi(x_2) + o(\theta), \end{aligned}$$

where $x_k = \Phi^{-1}(1 - e^{-t_k})$ ($k = 1, 2$), $o(\theta)$ holds uniformly for $(t_1, t_2) \in [\epsilon_1, M_1] \times [\epsilon_2, M_2]$ for any $0 < \epsilon_k < M_k < \infty$, $k = 1, 2$. Then, by the chain rule,

$$\begin{aligned} \frac{\partial}{\partial t_1} \tilde{S}_{ij}(t_1, t_2; \theta) &= \frac{\partial}{\partial x_1} \Psi(x_1, x_2; \theta) \frac{dx_1}{dt_1} \\ &= -e^{-(t_1+t_2)} - \theta x_1 e^{-t_1} \phi(x_2) + o(\theta), \end{aligned}$$

$$\frac{\partial}{\partial t_2} \tilde{S}_{ij}(t_1, t_2; \theta) = -e^{-(t_1+t_2)} - \theta x_2 e^{-t_2} \phi(x_1) + o(\theta),$$

and

$$\frac{\partial^2}{\partial t_1 \partial t_2} \tilde{S}_{ij}(t_1, t_2; \theta) = e^{-(t_1+t_2)} + \theta x_1 x_2 + o(\theta).$$

Then it follows that the martingale covariance function is

$$A_0(t_1, t_2; \theta) = \theta\{e^{t_1}\phi(x_1) - x_1\}\{e^{t_2}\phi(x_2) - x_2\} + o(\theta).$$

Again, $o(\theta)$ holds uniformly for $(t_1, t_2) \in [\epsilon_1, M_1] \times [\epsilon_2, M_2]$.

Hence, for any $\tau_1 < M_1 < \infty$ and $\tau_2 < M_2 < \infty$, writing the double integral $\int_0^{\tau_1} \int_0^{\tau_2}$ as

$$\int_0^{\tau_1} \int_0^{\tau_2} = \int_{\epsilon_1}^{\tau_1} \int_{\epsilon_2}^{\tau_2} + \int_{\epsilon_1}^{\tau_1} \int_0^{\epsilon_2} + \int_0^{\epsilon_1} \int_{\epsilon_2}^{\tau_2} + \int_0^{\epsilon_1} \int_0^{\epsilon_2},$$

we have

$$\begin{aligned} &\int_0^{\tau_1} \int_0^{\tau_2} \theta^{-1} |A_0(t_1, t_2; \theta) \\ &\quad - \theta(e^{t_1}\phi(x_1) - x_1)(e^{t_2}\phi(x_2) - x_2)| dt_1 dt_2 \\ &\leq \int_{\epsilon_1}^{\tau_1} \int_{\epsilon_2}^{\tau_2} \theta^{-1} |A_0(t_1, t_2; \theta) \\ &\quad - \theta(e^{t_1}\phi(x_1) - x_1)(e^{t_2}\phi(x_2) - x_2)| dt_1 dt_2 \\ &\quad + \int_{\epsilon_1}^{\tau_1} \int_0^{\epsilon_2} \theta^{-1} |A_0(t_1, t_2; \theta)| dt_1 dt_2 \\ &\quad + \int_0^{\epsilon_1} \int_{\epsilon_2}^{\tau_2} |e^{t_1}\phi(x_1) - x_1| dt_1 \int_0^{\epsilon_2} |e^{t_2}\phi(x_2) - x_2| dt_2 \\ &\quad + \int_0^{\epsilon_1} \int_0^{\epsilon_2} \theta^{-1} |A_0(t_1, t_2; \theta)| dt_1 dt_2 \end{aligned}$$

$$\begin{aligned}
& + \int_0^{\epsilon_1} |e^{t_1} \phi(x_1) - x_1| dt_1 \int_0^{\epsilon_2} |e^{t_2} \phi(x_2) - x_2| dt_2 \\
& + \int_0^{\epsilon_1} \int_{\epsilon_2}^{\tau_2} \theta^{-1} |A_0(t_1, t_2; \theta)| dt_1 dt_2 \\
& + \int_0^{\epsilon_1} |e^{t_1} \phi(x_1) - x_1| dt_1 \int_{\epsilon_2}^{\tau_2} |e^{t_2} \phi(x_2) - x_2| dt_2.
\end{aligned}$$

It can be shown that $A_0(t_1, t_2; \theta)$ is integrable over any finite rectangle $[0, \tau_1] \times [0, \tau_2]$ and that $e^{t_k} \phi(x_k) - x_k$ ($k = 1, 2$) are integrable over any finite interval $[0, \tau_k]$. Therefore, using $(\epsilon - \delta)$ -type arguments, we can show that all of the foregoing components converge to 0 as $\theta \rightarrow 0$. Hence

$$\begin{aligned}
& \int_0^{\tau_1} \int_0^{\tau_2} A_0(t_1, t_2; \theta) dt_1 dt_2 \\
& = \theta \int_0^{\tau_1} \{e^{t_1} \phi(x_1) - x_1\} dt_1 \int_0^{\tau_2} \{e^{t_2} \phi(x_2) - x_2\} dt_2 + o(\theta).
\end{aligned}$$

Furthermore, integration by parts yields, for $k = 1, 2$,

$$\begin{aligned}
& \int_0^{\tau_k} \{e^{t_k} \phi(x_k) - x_k\} dt_k \\
& = \tau_k \Phi^{-1}(1 - e^{-\tau_k}) - e^{\tau_k} \phi\{\Phi^{-1}(1 - e^{-\tau_k})\} \\
& \quad + \int_{-\infty}^{\Phi^{-1}(1 - e^{-\tau_k})} \left[\log\{1 - \Phi(x)\} - \frac{x\phi(x)}{1 - \Phi(x)} \right] dx.
\end{aligned}$$

Hence when the spatial dependence is weak, we are able to approximate the two-dimensional integral of the martingale covariance by the product of two univariate integrals, which greatly facilitates computation. This result also indicates that the covariance between two martingales decays to 0 at the same rate as the spatial correlation parameter θ , warranting the large-sample theory.

A.2 Regularity Conditions

For the asymptotic properties of the estimator, we assume that the spatial domain is increasing regularly in the sense of Guyon (1995). That is, we consider increasing-domain asymptotics, wherein the domain $D_m \subset \mathcal{R}^2$ is a sequence of increasing domains over which the data are collected. Let $|D_m|$ be the associated cardinalities and assume that there exists an $a > 0$ and m_n , a strictly increasing sequence of integers such that

$$\sum_{n \geq 1} n^a |D_{m_n}|^{-1} < \infty$$

and

$$\sum_{n \geq 1} \left(\frac{|D_{m_{n+1}}|/|D_{m_n}|}{|D_{m_n}|} \right)^2 < \infty.$$

Another commonly used asymptotic framework in spatial statistics is in-fill asymptotics, which has been found to be most useful when considering the asymptotics of kriging. Because we are mainly concerned with the asymptotic behavior of the estimates of the population-level regression parameters as well as correlation parameters, we have adopted the increasing-domain asymptotics in the following derivations. In practice, increasing-domain asymptotics are appropriate when the spatial domain of interest is extendable, and new observations are added beyond existing ones, generating an expanding surface.

Next, we state the other sufficient regularity conditions, which warrant the large-sample theory on a random field:

1. Stability. Denote by $s^{(k)}(\boldsymbol{\beta}, t) = E\{Y_j(t) \mathbf{Z}_j^{\otimes k}(t) e^{\boldsymbol{\beta}' \mathbf{Z}_j(t)}\}$ for $k = 0, 1, 2$. Assume that these functions exist and are bounded in

$\mathcal{B} \times [0, \tau]$. In particular, $s^{(0)}(\boldsymbol{\beta}, t)$ is bounded away from 0. Moreover,

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} |s^{(k)}(\boldsymbol{\beta}, t) - s^{(k)}(\boldsymbol{\beta}, t)| \xrightarrow{P} 0$$

for $k = 0, 1, \dots, 3$.

Assume that all of the covariates \mathbf{Z}_i are uniformly bounded and that the weight functions, $W_{(i,j)}(t)$, are chosen such that there exist (bounded) functions $sw_{ij}^{(k)}(\boldsymbol{\beta}, t)$, $i, j = 1, 2$, which satisfy, for $k = 0, 1$,

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} \left\| m^{-1} \sum_{i \geq j} \mathbf{Z}_i(t) w_{11}^{(i,j)}(t) Y_i(t) e^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \otimes \mathbf{Z}_i^k(t) - sw_{11}^{(k)}(\boldsymbol{\beta}, t) \right\| \xrightarrow{P} 0,$$

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} \left\| m^{-1} \sum_{i \geq j} \mathbf{Z}_i(t) w_{12}^{(i,j)}(t) Y_j(t) e^{\boldsymbol{\beta}' \mathbf{Z}_j(t)} \otimes \mathbf{Z}_j^k(t) - sw_{12}^{(k)}(\boldsymbol{\beta}, t) \right\| \xrightarrow{P} 0,$$

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} \left\| m^{-1} \sum_{i \geq j} \mathbf{Z}_j(t) w_{21}^{(i,j)}(t) Y_i(t) e^{\boldsymbol{\beta}' \mathbf{Z}_i(t)} \otimes \mathbf{Z}_i^k(t) - sw_{21}^{(k)}(\boldsymbol{\beta}, t) \right\| \xrightarrow{P} 0,$$

and

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} \left\| m^{-1} \sum_{i \geq j} \mathbf{Z}_j(t) w_{22}^{(i,j)}(t) Y_j(t) e^{\boldsymbol{\beta}' \mathbf{Z}_j(t)} \otimes \mathbf{Z}_j^k(t) - sw_{22}^{(k)}(\boldsymbol{\beta}, t) \right\| \xrightarrow{P} 0.$$

Here, for two column vectors, say, \mathbf{a} and \mathbf{b} , $\mathbf{a} \otimes \mathbf{b} = \mathbf{ab}'$.

2. Boundedness. Assume that the covariate processes $\mathbf{Z}_i(\cdot)$ and weights $w_{ij}^{(uv)}(\cdot)$, \mathbf{v}_{ij} are uniformly bounded. Also assume that $\Lambda_0(\tau) < \infty$ for $\tau < \infty$.
3. Differentiability. Assume that the covariance function A_0 is at least twice differentiable.
4. Positive definiteness of the information. Assume that matrix $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{ij})_{2 \times 2}$, has positive eigenvalues, where

$$\begin{aligned}
\boldsymbol{\Sigma}_{11} & = \sum_{p=1}^2 \sum_{q=1}^2 \int_0^\tau \left\{ sw_{pq}^{(1)}(\boldsymbol{\beta}_0, t) \right. \\
& \quad \left. - sw_{pq}^{(0)}(\boldsymbol{\beta}_0, t) \otimes \frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right\} d\Lambda_0(t),
\end{aligned}$$

$$\boldsymbol{\Sigma}_{12} = 0,$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{21} & = \int_0^\tau \lim_{m \rightarrow \infty} m^{-1} \\
& \quad \times \sum_{i \geq j} \mathbf{v}_{ij} \otimes \left[\left\{ \mathbf{Z}_i(t) - \frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right\} \right. \\
& \quad \times Y_i(t) e^{\boldsymbol{\beta}'_0 \mathbf{Z}_i(t)} \{M_j(X_j \wedge \tau) + A_{ij}^{(100)}\} \\
& \quad \left. + \int_0^\tau \left\{ \mathbf{Z}_j(t) - \frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right\} \right. \\
& \quad \left. \times Y_j(t) e^{\boldsymbol{\beta}'_0 \mathbf{Z}_j(t)} \{M_i(X_i \wedge \tau) + A_{ij}^{(010)}\} \right] d\Lambda_0(t).
\end{aligned}$$

and

$$\Sigma_{22} = \lim_{m \rightarrow \infty} m^{-1} \sum_{i>j} \mathbf{v}_{ij} \otimes \mathbf{A}_{ij}^{(001)},$$

for $i = j$;

$$A_{ii}^{(100)} = A_{ii}^{(010)} = 0,$$

for $i \neq j$; and

$$A_{ij}^{(100)} = \int_0^{\Lambda_i(\tau \wedge X_i)} \int_0^{\Lambda_j(\tau \wedge X_j)} A_0^{(100)}\{s_1, s_2; \theta_{ij}(\boldsymbol{\alpha}_0)\} ds_1 ds_2,$$

$$A_{ij}^{(010)} = \int_0^{\Lambda_i(\tau \wedge X_i)} \int_0^{\Lambda_j(\tau \wedge X_j)} A_0^{(010)}\{s_1, s_2; \theta_{ij}(\boldsymbol{\alpha}_0)\} ds_1 ds_2,$$

and

$$\begin{aligned} \mathbf{A}_{ij}^{(001)} &= \int_0^{\Lambda_i(\tau \wedge X_i)} \int_0^{\Lambda_j(\tau \wedge X_j)} A_0^{(010)}\{s_1, s_2; \theta_{ij}(\boldsymbol{\alpha}_0)\} \\ &\quad \times \frac{\partial}{\partial \boldsymbol{\alpha}} \theta_{ij}(\boldsymbol{\alpha}_0) ds_1 ds_2. \end{aligned}$$

Here $\Lambda_i(t) = \int_0^t e^{\boldsymbol{\beta}'_0 \mathbf{Z}_i(s)} d\Lambda_0(s)$, $A_0^{(100)}(u, v; \theta) = \frac{\partial}{\partial u} A_0(u, v; \theta)$, $A_0^{(010)}(u, v; \theta) = \frac{\partial}{\partial v} A_0(u, v; \theta)$, and $A_0^{(001)}(u, v; \theta) = \frac{\partial}{\partial \theta} A_0(u, v; \theta)$. All of the foregoing limits are the probabilistic limits (provided existence) when $m \rightarrow \infty$.

5. At $\boldsymbol{\Theta}_0$, $\sup_{i,j} E(\hat{\mathbf{U}}_{i,j} \hat{\mathbf{U}}'_{i,j}) < \infty$ and $\boldsymbol{\Sigma}^{(2)} = mE(\mathbf{G}_m \mathbf{G}'_m)$ is bounded below by a positive definite matrix and $\sup_m \boldsymbol{\Sigma}^{(2)} < \infty$.

A.3 Sketchy Proof of Proposition 1

In this section we provide a sketchy proof of Proposition 1. A detailed proof is given in a technical report that can be obtained from the authors. We first apply the inverse function theorem (see, e.g., Foutz 1977) to prove consistency. Specifically, we need to check the three sufficient conditions based on a straightforward extension of the work of Foutz (1977): (1) asymptotic unbiasedness of the estimating equation, that is, $\mathbf{G}_m(\boldsymbol{\Theta}_0) \xrightarrow{P} \mathbf{0}$; (2) existence, continuity, and uniform convergence of the partial derivatives of the estimating equations in a neighborhood of the true parameters, that is, $(\partial/\partial \boldsymbol{\Theta})\mathbf{G}_m(\boldsymbol{\Theta})$ converges uniformly in a neighborhood of $\boldsymbol{\Theta}_0$; and (3) the negative definiteness of the partial derivatives of the estimating equations at the true values, that is, $(\partial/\partial \boldsymbol{\Theta})\mathbf{G}_m(\boldsymbol{\Theta}_0)$ converges in probability to a matrix with strictly negative eigenvalues.

Rewrite $\mathbf{G}_m(\boldsymbol{\Theta}) = \{\hat{\mathbf{U}}_1^{(1)} + \hat{\mathbf{U}}_2^{(1)} + \hat{\mathbf{U}}_3^{(1)} + \hat{\mathbf{U}}_4^{(1)}, \hat{\mathbf{U}}^{(2)}\}'$, where

$$\hat{\mathbf{U}}_1^{(1)} = m^{-1} \sum_{u \geq v} \int_0^\tau \mathbf{Z}_u(t) w_{11}^{(uv)}(t) d\hat{M}_u(t),$$

$$\hat{\mathbf{U}}_2^{(1)} = m^{-1} \sum_{u \geq v} \int_0^\tau \mathbf{Z}_u(t) w_{12}^{(uv)}(t) d\hat{M}_v(t),$$

$$\hat{\mathbf{U}}_3^{(1)} = m^{-1} \sum_{u \geq v} \int_0^\tau \mathbf{Z}_v(t) w_{21}^{(uv)}(t) d\hat{M}_u(t),$$

$$\hat{\mathbf{U}}_4^{(1)} = m^{-1} \sum_{u \geq v} \int_0^\tau \mathbf{Z}_v(t) w_{22}^{(uv)}(t) d\hat{M}_v(t),$$

and

$$\hat{\mathbf{U}}^{(2)} = m^{-1} \sum_{u \geq v} \mathbf{v}_{uv} \{\hat{M}_u(\tau) \hat{M}'_v(\tau) - \hat{A}_{uv}\},$$

where \hat{A}_{uv} is as defined in (10).

We next show that $\mathbf{G}_m(\boldsymbol{\Theta}_0) \rightarrow \mathbf{0}$ in probability. Consider $\hat{\mathbf{U}}_1^{(1)}(\boldsymbol{\Theta}_0)$, with $\Lambda_0(t)$ substituted by its Breslow estimator,

$$\begin{aligned} \hat{\mathbf{U}}_1^{(1)} &= m^{-1} \sum_{u \geq v} \int_0^\tau \mathbf{Z}_u(t) w_{11}^{(uv)}(t) dM_u(t) \\ &\quad - m^{-1} \sum_{u \geq v} \int \mathbf{Z}_u(t) Y_u(t) e^{\boldsymbol{\beta}' \mathbf{Z}_u(t)} w_{11}^{(uv)}(t) \\ &\quad \times \frac{\sum_{i=1}^m dM_i(t)}{\sum_{i=1}^m Y_i(t) e^{\boldsymbol{\beta}'_0 \mathbf{Z}_i(t)}} \end{aligned} \tag{A.2}$$

$$= \frac{1}{m} \sum_{i=1}^m \int_0^\tau \left\{ \mathbf{Z}_i(t) \left\{ \sum_{j \leq i} w_{11}^{(ij)}(t) \right\} - \frac{sw_{11}^{(0)}(t)}{s^{(0)}(t)} \right\} dM_i(t) \tag{A.3}$$

$$\begin{aligned} &- \frac{1}{m} \sum_{i=1}^m \int_0^\tau \frac{1}{s^{(0)}(t)} \left\{ m^{-1} \sum_{u \geq v} \mathbf{Z}_u(t) Y_u(t) e^{\boldsymbol{\beta}'_0 \mathbf{Z}_u(t)} w_{11}^{(uv)}(t) \right. \\ &\quad \left. - sw_{11}^{(0)}(\boldsymbol{\beta}_0, t) \right\} dM_i \end{aligned} \tag{A.4}$$

$$\begin{aligned} &- \frac{1}{m} \sum_{i=1}^m \int sw_{11}^{(0)}(t) \left[\left(m^{-1} \sum_{i=1}^m Y_u(t) e^{\boldsymbol{\beta}'_0 \mathbf{Z}_u(t)} \right)^{-1} \right. \\ &\quad \left. - \{s^{(0)}(t)\}^{-1} \right] dM_i(t) \end{aligned} \tag{A.5}$$

$$\begin{aligned} &- \frac{1}{m} \sum_{i=1}^m \int_0^\tau \left\{ m^{-1} \sum_{u \geq v} \mathbf{Z}_u(t) Y_u(t) e^{\boldsymbol{\beta}'_0 \mathbf{Z}_u(t)} w_{11}^{(uv)}(t) \right. \\ &\quad \left. - sw_{11}^{(0)}(t) \right\} \end{aligned} \tag{A.6}$$

$$\begin{aligned} &\times \left[\left(m^{-1} \sum_{i=1}^m Y_u(t) e^{\boldsymbol{\beta}'_0 \mathbf{Z}_u(t)} \right)^{-1} - \{s^{(0)}(t)\}^{-1} \right] dM_i(t), \end{aligned} \tag{A.7}$$

where $M_i(u) = N_i(u) - \int_0^u Y_i(u) \exp(\boldsymbol{\beta}'_0 \mathbf{Z}_i) d\Lambda_0(t)$, a martingale with respect to the filtration generated by each individual's own survival status and covariate processes. It can be shown that (A.2)–(A.5) all converge to 0 in probability. Similarly, it can be shown that $\hat{\mathbf{U}}_2^{(1)}, \hat{\mathbf{U}}_3^{(1)}$, and $\hat{\mathbf{U}}_4^{(1)}$ all converge to 0 in probability. It can also be shown that $\hat{\mathbf{U}}^{(2)}$ is asymptotically equivalent to

$$\begin{aligned} \mathbf{U}^{(2)} &= \frac{1}{m} \sum_{u \geq v} \mathbf{v}_{uv} \left\{ M_u(\tau) M'_v(\tau) \right. \\ &\quad \left. - \int_0^\tau \int_0^\tau Y_u(s) Y_v(t) \right. \\ &\quad \left. \times A_0\{\Lambda_u(s), \Lambda_v(t)\} d\Lambda_u(s) d\Lambda_v(t) \right\}, \end{aligned}$$

which also converges to 0 in probability under the mixing condition using the Chebyshev inequality. Therefore, we conclude that $\mathbf{G}_m(\boldsymbol{\Theta}_0)$ converges to 0 in probability.

For any fixed m , the continuity of $\partial \mathbf{G}_m(\boldsymbol{\Theta})/\partial \boldsymbol{\Theta}$ in $\boldsymbol{\Theta}$ follows from the smoothness assumption of the covariance rate function $A_0(\cdot)$. We then consider the large-sample behavior for $\partial \mathbf{G}_m(\boldsymbol{\Theta})/\partial \boldsymbol{\Theta}$ in a small neighborhood of $\boldsymbol{\Theta}_0$. For example, we can show that $\partial \hat{\mathbf{U}}_1^{(1)}/\partial \boldsymbol{\beta}$ converges uniformly in a neighborhood of $\boldsymbol{\Theta}_0$ and, at $\boldsymbol{\beta}_0$,

$$\frac{\partial \hat{\mathbf{U}}_1^{(1)}}{\partial \boldsymbol{\beta}} \rightarrow \int_0^\tau \left\{ sw_{11}^{(2)}(\boldsymbol{\beta}_0, t) - sw_{11}^{(0)}(\boldsymbol{\beta}_0, t) \otimes \frac{s^{(1)}(\boldsymbol{\beta}_0, t)}{s^{(0)}(\boldsymbol{\beta}_0, t)} \right\} d\Lambda_0(t)$$

in probability. The same arguments also apply to $\partial\hat{U}_2^{(1)}/\partial\beta$, $\partial\hat{U}_3^{(1)}/\partial\beta$, and $\partial\hat{U}_4^{(1)}/\partial\beta$, which converge uniformly in a neighborhood of Θ_0 . Hence, in particular, at Θ_0 , the (1, 1)th block of $\partial\mathbf{G}_m(\Theta)/\partial\Theta$ converges to $-\Sigma_{11}$. Similarly, we can show that other blocks of $\partial\mathbf{G}_m(\Theta)/\partial\Theta$ converge uniformly at Θ_0 to $-\Sigma$, which has negative eigenvalues by condition 4. Thus it follows from the inverse function theorem (Foutz 1977) that when n is sufficiently large, in a neighborhood of Θ_0 there exists a unique sequence of $\hat{\Theta} = (\hat{\beta}, \hat{\alpha})'$ such that $\mathbf{G}_m(\hat{\Theta}) = \mathbf{0}$ with probability going to 1 and $\hat{\Theta} \xrightarrow{P} \Theta_0 = (\beta'_0, \alpha'_0)'$.

We now consider the asymptotic normality of $\hat{\Theta}$. A Taylor expansion of $\mathbf{G}_m(\hat{\Theta})$ at the true value Θ_0 gives

$$-\sqrt{m}\left\{\frac{\partial}{\partial\Theta}\mathbf{G}_m(\Theta^*)\right\}(\hat{\Theta} - \Theta_0) = \sqrt{m}\mathbf{G}_m(\Theta_0),$$

where Θ^* is on the segment between Θ_0 and $\hat{\Theta}$. With condition 5 and the assumed spatial dependence, a central limit theorem (Guyon 1995, chap. 3) applies to the sequence of $\mathbf{G}_m(\Theta_0)$ such that

$$\sqrt{m}\{\Sigma^{(2)}\}^{-1/2}\mathbf{G}_m(\Theta_0) \rightarrow N(\mathbf{0}, \mathbf{I})$$

in distribution. Note that $\partial\mathbf{G}_m(\Theta^*)/\partial\Theta$ converges to $-\Sigma$ in probability. Application of the Slutsky theorem gives

$$\sqrt{m}\{\Sigma^{(2)}\}^{-1/2}\Sigma(\hat{\Theta} - \Theta_0) \rightarrow N(\mathbf{0}, \mathbf{I})$$

in distribution.

[Received August 2004. Revised September 2005.]

REFERENCES

- Abramowitz, M., and Stegun, I. A. (eds.) (1965), *Handbook of Mathematical Functions*, New York: Dover.
- Banerjee, S., and Carlin, B. P. (2003), "Semiparametric Spatio-Temporal Frailty Modeling," *Environmetrics*, 14, 523–535.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC Press.
- Besag, J., York, J., and Mollie, A. (1991), "Bayesian Image Restoration, With Two Applications in Spatial Statistics," *Annals of the Institute of Statistical Mathematics*, 43, 1–20.
- Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.
- Bulmann, P., and Kunsch, H. (1995), "The Blockwise Bootstrap for General Parameters of a Stationary Time Series," *Scandinavian Journal of Statistics*, 22, 35–54.
- Cai, J., and Prentice, R. L. (1997), "Regression Estimation Using Multivariate Failure Time Data and a Common Baseline Hazard Function Model," *Lifetime Data Analysis*, 3, 197–213.
- Carlin, B. P., and Louis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall.
- Carlstein, E. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic From a Stationary Sequence," *The Annals of Statistics*, 14, 1171–1179.
- (1988), "Law of Large Numbers for the Subseries Values of a Statistic From a Stationary Sequence," *Statistics*, 19, 295–299.
- Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.
- Clayton, D., and Kaldor, J. (1987), "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping," *Biometrics*, 43, 671–681.
- Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998) "Model-Based Geostatistics," *Journal of the Royal Statistical Society, Ser. C*, 47, 299–326.
- Foutz, R. V. (1977), "On the Unique Consistent Solution to the Likelihood Equations," *Journal of the American Statistical Association*, 72, 147–148.
- Guyon, X. (1995), *Random Fields on a Network: Modeling, Statistics, and Applications*, New York: Springer-Verlag.
- Haining, R., Griffith, D., and Bennett, R. (1989), "Maximum Likelihood Estimation With Missing Spatial Data and With an Application to Remotely Sensed Data," *Communications in Statistics: Theory and Methods*, 18, 1875–1894.
- Heagerty, P. J., and Lele, S. R. (1998), "A Composite Likelihood Approach to Binary Spatial Data," *Journal of the American Statistical Association*, 93, 1099–1111.
- Hougaard, P. (1986), "A Class of Multivariate Failure Time Distributions," *Biometrika*, 73, 671–678.
- (2000), *Analysis of Multivariate Survival Data*, New York: Springer-Verlag.
- Jin, Z., Lin, D., and Wei, L. J. (2003), "Rank-Based Inference for the Accelerated Failure Time Model," *Biometrika*, 90, 341–353.
- Journel, A. G. (1983), "Geostatistics," *Encyclopedia of Statistical Sciences*, 3, 424–431.
- Kalbfleisch, J., and Prentice, R. (2002), *The Statistical Analysis of Failure Time Data*, Hoboken, NJ: Wiley.
- Kott, P. S. (1998), "Using the Delete-a-Group Jackknife Variance Estimator in Practice," in *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 763–768.
- Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000), *Continuous Multivariate Distributions, Vol. 1: Models and Applications*, New York: Wiley.
- Kunsch, H. R. (1998), "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics*, 17, 1217–1241.
- Lahiri, S. N., Kaiser, M. S., Cressie, N., and Hsu, N. (1999), "Prediction of Spatial Cumulative Distribution Functions Using Subsampling," *Journal of the American Statistical Association*, 94, 86–97.
- Lele, S. (1991), "Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes," *Journal of the Royal Statistical Society, Ser. B*, 53, 253–267.
- Lindsay, B. G. (1988), "Composite Likelihood Methods," in *Statistical Inference From Stochastic Processes*, ed. N. U. Prabhu, Providence, RI: American Mathematical Society, pp. 221–239.
- Liu, R., and Singh, K. (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of Bootstrap*, eds. R. LePage and L. Billard, New York: Wiley, pp. 225–248.
- Loh, J. M., and Stein, M. L. (2004), "Bootstrapping a Spatial Point Process," *Statistica Sinica*, 14, 69–101.
- Moran, P. A. (1983), "A New Expansion for the Multivariate Normal Distribution," *The Australian Journal of Statistics*, 25, 339–344.
- Murphy, S. A. (1995), "Asymptotic Theory for the Frailty Model," *The Annals of Statistics*, 23, 182–198.
- Oakes, D. (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association*, 84, 487–493.
- Parner, E. (1998), "Asymptotic Theory for the Correlated Gamma-Frailty Model," *The Annals of Statistics*, 26, 183–214.
- (2001), "A Composite Likelihood Approach to Multivariate Survival Data," *Scandinavian Journal of Statistics*, 28, 295–302.
- Politis, D. N., and Romano, J. P. (1992), "A General Resampling Scheme for Triangular Arrays of α -Mixing Random Variables With Application to the Problem of Spectral Density Estimation," *The Annals of Statistics*, 20, 1985–2007.
- Prentice, R. L., and Cai, J. (1992), "Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data," *Biometrika*, 79, 495–512.
- Sherman, M. (1996), "Variance Estimation for Statistics Computed From Spatial Lattice Data," *Journal of the Royal Statistical Society, Ser. B*, 58, 509–523.
- Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory of Kriging*, New York: Springer-Verlag.
- Sungur, E. A. (1990), "Dependence Information in Parameterized Copulas," *Communications in Statistics, Part B—Simulations and Computation*, 19, 1339–1360.
- Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997), "Hierarchical Spatio-Temporal Mapping of Disease Rates," *Journal of the American Statistical Association*, 92, 607–617.
- Waller, L., and Gotway, C. (2004), *Applied Spatial Statistics for Public Health Data*, New York: Wiley.
- 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.
- Winkler, G. (1995), *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: A Mathematical Introduction*, New York: Springer-Verlag.
- Zhu, J., and Morgan, G. D. (2004), "Comparison of Spatial Variables Over Subregions Using a Block Bootstrap," *Journal of Agricultural, Biological, and Environmental Statistics*, 9, 91–104.