Semiparametric Normal Transformation Models for Spatially Correlated Survival Data

Yi Li and Xihong Lin *

Abstract

There is an emerging interest in modeling spatially correlated survival data in biomedical and epidemiological studies. In this paper, 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 models; Cross ratio; Dependence measures; Likelihood; Population-average interpretation; Semiparametric estimating equations; Spatial survival data; Spatial dependence.

*Yi Li is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115, and Xihong Lin is Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115. This work was partially supported by U. S. National Cancer Institute grants R01 CA95747 (Li) and R01 CA76404 (Lin). We thank the editor, an associate editor and reviewer for their insightful comments.

1 Introduction

Biomedical and epidemiological studies have spawned an increasing interest and practical need in 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 etiology of 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 ages 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. Ages at onset of asthma of the children in this study were 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 while 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 (p.306, 2002)). There has been, however, 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, et al., 1989) and discrete data (Journel, 1983; Cressie, 1993; Carlin and Louis, 1996; Diggle et al., 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, Mollie, 1991; Waller et al., 1997).

Little work however has been done for modeling survival data that are subject to spatial correlation. 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. It is however 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 only allow for a simple constant correlation structure, and are difficult to be extended to allow 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 (Ch 9, 2004) considered hierarchical frailty spatial survival models. But, regression coefficients in their models do not have a population-level interpretation.

In contrast to the existing methodology, we develop in this article 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 where 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. We develop in Section 4 spatial semiparametric estimating equations for regression coefficients and spatial correlation parameters, and study the asymptotic properties for the resulting estimators. In Section 5 we evaluate via 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 discussions 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 ($i = 1, \cdots, m$), we observe a $r \times 1$ vector of covariates $\mathbf{Z}_i$, and an observed event time $X_i = min(T_i, C_i)$ and a non-censoring indicator $\delta_i = I(T_i \leq C_i)$, where T_i and C_i are underlying true survival time and censoring time respectively, and $I(\cdot)$ is an indicator function. We assume independent censoring, i.e., 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 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

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

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, $\mathcal{F}_{i,t} = \sigma\{I(X_i \le s, \delta = 1), I(X_i \ge s), \mathbf{Z}_i(s), 0 \le s \le t\}$, 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 $T_1, \cdots, T_m$ 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

$$T_i^* = \Phi^{-1}\left\{1 - e^{-\Lambda_i(T_i)}\right\} \tag{2}$$

follows the standard normal distribution marginally, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We can then conveniently impose a spatial structure on the underlying random fields of $\mathbf{T}^* = \{T_i^*, i = 1, \ldots, m\}$ 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

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

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 $\mathbf{a}_j$, i.e.

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

for $i \neq j$ $(i, j = 1, \ldots, m)$, 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 non-informative censoring, the likelihood function for the unknown parameters $\{\Lambda_0(\cdot), \boldsymbol{\beta}, \boldsymbol{\alpha}\}$, based on the observed data $(X_i, \delta_i, \mathbf{Z}_i), i = 1, \ldots, m$, is

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

evaluated at $(X_1, \ldots, X_m)$, where $\psi(x_1, \ldots, x_m; \boldsymbol{\Gamma})$ is the density function for an m-dimensional normal distribution with mean 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 will explore a spatial semiparametric estimating equation approach to draw inference in Section 4.

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

Since 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 which decays as the Euclidean distance d_{ij} between two individuals increases. A widely adopted choice for the correlation function is the Matèrn function

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

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, α_1 is a scale parameter and corresponds to the 'partial sill' as described in Waller and Gotway (2004, p.279), α_2 measures the correlation decay with the distance and α_3 is a smoothness parameter, $\Gamma(\cdot)$ is the conventional Gamma function, $K_{\alpha_3}(\cdot)$ 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, special cases including the exponential function $\rho(d, \boldsymbol{\alpha}) = \alpha_1 \exp(-d\alpha_2)$ when the smoothness parameter $\alpha_3 = 0.5$, and the "Gaussian" correlation function $\rho(d, \boldsymbol{\alpha}) = \alpha_1 \exp\{-d^2\alpha_2^2\}$ when $\alpha_3 \to \infty$ (see, e.g., Waller and Gotway, 2004, p. 279). In all these formulations, we require $0 \le \alpha_1 \le 1$ and $\alpha_2, \alpha_3 \ge 0$. Note that such spatial dependence models distinguish local and global spatial effects, where α_1 measures local correlation (i.e. $\alpha_1 = \lim_{d \to 0+} \rho(d, \boldsymbol{\alpha})$), while 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 as it requires dense space data and may even run into identifiability problems.

Stein (1999) has argued that data can not 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 T

The correlation coefficient θ_{ij} conveniently specifies the spatial correlation of the normally transformed survival times T_i^* and T_j^* via 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, i.e., 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). We investigate in this section 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 $c_{12}(t_1, t_2) = 1$, T_1 and T_2 are independent at (t_1, t_2). If $c_{12}(t_1, t_2) > 1$, T_1 and T_2 are positively correlated at (t_1, t_2), and vise versa. If $c_{12}(t_1, t_2)$ is a constant, (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

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

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

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

where $F_i(\cdot), F_j(\cdot)$ are the marginal CDFs of T_i and T_j respectively.

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 Frechet bound $\min\{1 - F_i(t_1), 1 - F_j(t_2)\}$ as $\theta_{ij} \to 1^-$, the independent case when $\theta_{ij} \to 0$, and the lower Frechet bound $\max\{1 - F_i(t_1) - F_j(t_2), 0\}$ as $\theta_{ij} \to -1^+$.

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

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

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 one. One can see that the cross ratio $c_{ij}(t_1, t_2)$ is a nonlinear monotone increasing function of θ_{ij}. As $\theta_{ij} \to 0$, $c_{ij}(t_1, t_2) \to 1$, 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 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. They hence are parallel to the classical correlation coefficient. However, as a global dependence measure, they are not informative about how the correlation varies with times.

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

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

respectively.

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. We hence can easily use equation (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 $\Psi(\cdot), \Phi_2(\cdot)$ are defined in (7). Hence Kendall's τ and Spearman's ρ are uniquely determined by the marginal survival 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 to 0 as $\theta_{ij} \to 0$, approach to 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 $\boldsymbol{\beta}$ and the spatial correlation parameters $\boldsymbol{\alpha}$ in $\theta_{ij}(\boldsymbol{\alpha})$.

4.1 The Martingale Covariance Rate Function

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

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

To relate the correlation parameters $\boldsymbol{\alpha}$ to the counting processes, one needs to consider the joint counting process of two individuals. Define the conditional martingale covariance rate function for the joint counting process of two individuals, a multi-dimensional 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

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

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

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

where $S_{ij}(\cdot)$ is defined in (7). Following Prentice and Cai (1992), one 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

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

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 $\theta \to 0+$, $A_0(v_1, v_2; \theta)$ 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 $\boldsymbol{\beta}$ (an $r \times 1$ vector) and the correlation parameters $\boldsymbol{\alpha}$ (a $q \times 1$ vector) by considering the first two moments of the martingale vector $(M_1, \ldots, M_m)$. In particular, for a pre-determined constant $\tau > 0$ such that it is within the support of the observed failure time, i.e $P(\tau < C_i \wedge T_i) > 0$ (in practice τ is usually the study duration), we consider the following unbiased estimating functions for $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ for an arbitrary pair of two individuals, indexed by u and v:

- if $u = v$,

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

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

- if $u \neq v$,

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

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

$$\begin{aligned} A_{uv} &= \int_0^\tau \int_0^\tau Y_u(s)Y_v(t)A_0\{\Lambda_u(s), \Lambda_v(t); \theta_{uv}\}d\Lambda_u(s)d\Lambda_v(t) \\ &= \int_0^{\Lambda_u(X_u\wedge\tau)} \int_0^{\Lambda_v(X_v\wedge\tau)} A_0\{t_1, t_2; \theta_{uv}\}dt_1dt_2. \end{aligned}$$

  We show in the Appendix that A_{uv}, the covariance of martingales, decay to 0 at the same rate as the spatial correlation parameter θ_{uv}. We also provide in the Appendix 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, since $E\{\mathbf{U}_{u,v}(\mathbf{\Theta}_0)\} = 0$, where the expectation is taken under the true $\mathbf{\Theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)$ 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 $\boldsymbol{\beta}$, is unbiased even when the spatial correlation structure is misspecified. Hence the regression coefficient estimator $\widehat{\boldsymbol{\beta}}$ is robust to misspecification of the spatial correlation structure.

It is however not immediately computable as $\Lambda_0(t)$ in the estimating equations is unknown. A natural alternative is to substitute it with the Breslow estimator

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

As a result, the parameters of interest $\mathbf{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ 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 respectively

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

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

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

$$m^{-1} \begin{bmatrix} \int_0^\tau \mathbf{Z}(s)\mathbf{W}d\hat{\mathbf{M}}(s) \\ \hat{\mathbf{M}}'(\tau)\mathbf{V}_1\hat{\mathbf{M}}(\tau) - tr(\mathbf{V}_j\hat{\mathbf{A}}) \end{bmatrix} = 0, \tag{10}$$

where $j = 1, \cdots, q$, $\mathbf{W}$ and $\mathbf{V}_j$ are weight matrices, $\hat{\mathbf{M}} = (\hat{M}_1, \ldots, \hat{M}_n)'$, $\mathbf{Z}(s) = \{\mathbf{Z}_1(s), \ldots, \mathbf{Z}_n(s)\}'$, $\hat{\mathbf{A}}$ 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 $\mathbf{V}_1, \ldots, \mathbf{V}_q$ are introduced to improve efficiency and convergence of the estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. 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 $\mathbf{D} = diag(A_{11}, \ldots, A_{mm})$. 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 $\boldsymbol{\beta}$. To specify $\mathbf{V}_j$ $(j = 1, \ldots, q)$, one could assume $\mathbf{V}_j = \mathbf{A}^{-1}(\partial \mathbf{A}/\partial \alpha_j)\mathbf{A}^{-1}$. Under this specification, the second set of estimating equations in (10) resembles the score equations of the variance components $\boldsymbol{\alpha}$ if the 'response' $\hat{\mathbf{M}}$ followed a multivariate normal distribution $N(0, \mathbf{A})$ (Cressie, 1993, p483).

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

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

where $\boldsymbol{\Omega}$ 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 $N(0, \boldsymbol{\Omega}^{-1})$ on $\boldsymbol{\Theta}$, and results in stabilized variance component estimates of $\boldsymbol{\alpha}$ 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, $\boldsymbol{\Omega} = \omega \mathbf{I}$, where $0 < \omega < 1$, to ensure numerical stability. Note as the sample size m goes to ∞, we have $\frac{1}{m}\boldsymbol{\Omega}\boldsymbol{\Theta} \to 0$. Therefore $\mathbf{G}_m(\boldsymbol{\Theta})$ and $\mathbf{G}_m^*(\boldsymbol{\Theta})$ are asymptotically equivalent, and therefore the large sample results of the original and penalized estimating equations are equivalent.

4.3 Asymptotic Properties and Variance Estimation

We study in this section 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 $\mathbf{G}_m(\boldsymbol{\Theta}) = 0$ exist and are consistent for the true values of $\mathbf{\Theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)$ and that $n^{1/2}\{\hat{\mathbf{\Theta}} - \mathbf{\Theta}_0\}$ is asymptotic normal with mean zero and a covariance matrix which can be easily estimated using a sandwich estimator. The results are formally stated in Proposition 1.

Proposition 1 *Assume the true $\mathbf{\Theta}_0$ is an interior point of an compact set, say, $\mathcal{B} \times \mathcal{A} \in R^{r+q}$, where r is the dimension of $\boldsymbol{\beta}$ and q is the dimension of $\boldsymbol{\alpha}$. Under the regularity conditions 1-5 in the Appendix, when m is sufficiently large, the estimating equation $G_m(\mathbf{\Theta}) = 0$ has a unique solution in a neighborhood of $\mathbf{\Theta}_0$ with probability tending to 1 and the resulting estimator $\hat{\mathbf{\Theta}}$ is consistent for $\mathbf{\Theta}_0$. Furthermore, $\sqrt{m}\{\mathbf{\Sigma}^{(2)}\}^{-1/2}\mathbf{\Sigma}\{(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})' - (\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)'\} \xrightarrow{d} N\{0, \mathbf{I}\}$, where $\mathbf{I}$ is an identity matrix whose dimension is equal to that of $\mathbf{\Theta}_0$, and*

$$\begin{aligned} \mathbf{\Sigma} &= \frac{1}{m}\sum_{u \geq v} E\left\{\frac{\partial}{\partial \mathbf{\Theta}} U_{u,v}(\mathbf{\Theta})\right\} \\ \mathbf{\Sigma}^{(2)} &= \frac{1}{m^2}\sum_{u_1 \geq v_1}\sum_{u_2 \geq v_2} E\{\mathbf{U}_{u_1,v_1}(\mathbf{\Theta}_0)\mathbf{U}_{u_2,v_2}(\mathbf{\Theta}_0)\}. \end{aligned}$$

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

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

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

Although each $\mathrm{E}\left\{\hat{\mathbf{U}}_{u_1,v_2}(\mathbf{\Theta}_0)\hat{\mathbf{U}}'_{u_2,v_2}(\mathbf{\Theta}_0)\right\}$ could be evaluated numerically, the total number of these calculations would be prohibitive, especially when the sample size m is large. To numerically approximate $\widehat{\mathbf{\Sigma}}^{(2)}$, we explore the resampling techniques of Carlstein (1986) and Sherman (1996). Specifically, under the assumption that asymptotically

$$m \times \mathrm{E}\left\{\mathbf{G}_m\mathbf{G}'_m\right\} \to \mathbf{\Sigma}_\infty,$$

we can estimate $\mathbf{\Sigma}_\infty$ by averaging K randomly chosen subsets of size m_j $(j = 1, \cdots, K)$ from the m subjects as

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

where $\widehat{\mathbf{G}}_{m_j}$ is obtained by substituting $\mathbf{\Theta}$ with $\widehat{\mathbf{\Theta}}$ 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 $\widehat{\mathbf{\Sigma}}_\infty$ and $\widehat{\mathbf{\Sigma}}$, the covariance of $\widehat{\boldsymbol{\Theta}}$ can be estimated by $\widehat{\boldsymbol{\Sigma}}^{-1}[1/m \times \widehat{\boldsymbol{\Sigma}}_{\infty}](\widehat{\boldsymbol{\Sigma}}^{-1})'$. For the covariance estimate of the penalized estimator obtained by solving $\mathbf{G}_m^*(\Theta) = 0$, $\widehat{\boldsymbol{\Sigma}}$ is replaced by $\widehat{\boldsymbol{\Sigma}} - \frac{1}{m}\boldsymbol{\Omega}$. 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, and then transformed the T_{ij}^* back to the original survival time scale to obtain T_{ij} using the equation (2) and the above marginal Cox model.

We set the true value $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$, $\alpha_1 = 0.5$ and $\alpha_2 = 2.5$. We varied α_3 in (6) to be 0.5 and 1. Censoring times c_{ij} were generated as independent uniform random variables on $[0, 1]$ and $[0, 2]$, resulting in 70% and 50% censoring, respectively. For each set of parameters, we considered the number of subjects (m) to be 100 and 200. We also considered $m = 400$ with $\alpha_3 = 0.5$ and 70% censoring. In our calculations, we set the penalty parameter to be $\omega = 0.1$. 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 data sets were generated for each configuration, and averages of the point estimates and their standard errors were calculated, along with the coverage rates of the corresponding 95% confidence intervals. The results are summarized in Table 1. These results show that our estimator performed well in finite samples. The finite sample biases of the regression coefficient estimates $\boldsymbol{\beta}$ were negligible, and the standard error estimates agreed well with their empirical counterparts, though 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 $\boldsymbol{\beta}$. The 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 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 $m = 200, \alpha_3 = 0.5$ 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 above with $m = 100$ and censoring proportion=70%, we generated the survival data with the spatial dependence specified by the 'spherical' correlation

$$\rho(d) = 0.5(1 - \frac{3d}{4} + \frac{d^3}{8})I(d \leq 2).$$

but assumed the Matèrn 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 $0.9950, 0.5232$ and 0.5028 respectively, 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 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 to better understand the natural history of asthma and its associated risk factors and to 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 ages at onset of asthma could be identified. Residential addresses were recorded and geocoded. The geographic distance was calculated in the unit of kilometer. 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 followup 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, aging from infancy to 6 years old. Asthma is a disease strongly affected environmental triggers. Since the children had similar backgrounds and living environment and were exposed with similar unmeasured similar physical and social environments, their ages at onset of asthma were likely to be subject to spatial correlation.

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

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

We assumed the Matèrn 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 standard error estimates (11). For checking the robustness of the method, we also varied the smoothness parameter α_3 in (6) to be 0.5, 1 and 1.5.

As 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

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

where $\hat{\mathbf{\Theta}}_j$ was the estimate produced from the jackknife replicate with the j-th 'group' deleted and $\hat{\mathbf{\Theta}}$ was 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 similar spirit of a subsampling scheme proposed by Carlstein (1986, 1988), treated each block approximately independent and seemed plausible for this data set, especially in the presence of weak spatial dependence. Loh and Stein (2004) termed this scheme as the *splitting method* and found it work even better than more complicated block-bootstrapping methods (e.g. Kunsch, 1989; Liu and Singh, 1992; Politis and Romano, 1992; Bulhmann and Kunsch, 1995). Other advanced resampling schemes for spatial data are also available, e.g double-subsampling method (Lahiri et al., 1999; Zhu and Morgan, 2004) and linear estimating equation Jackknifing (Lele, 1991), but are subject to much more computational burden compared with the simple jackknife scheme we used.

The results are presented in Table 2, with the large sample standard errors (SE_a) computed using the method described in Section 4.3 and the Jackknife standard errors (SE_j) computed using (13). The estimates of the regression coefficients and their standard errors 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 data set. The standard errors obtained from the large sample approximation and the Jackknife method were reasonably similar. Low respiratory index was highly significantly associated with the age at onset of asthma, e.g. $\widehat{\beta}_L = 0.3121$ ($SE_a = 0.0440, SE_j = 0.0357$) when $\alpha_3 = 0.5$; $\widehat{\beta}_L = 0.3118$ ($SE_a = 0.0430, SE_j = 0.0369$) when $\alpha_3 = 1.0$; $\widehat{\beta}_L = 0.3124$ ($SE_a = 0.0432, SE_j = 0.0349$) when $\alpha_3 = 1.5$, indicating that a child with a poor respiratory functioning was more likely to develop asthma, after controlling for maternal asthma, maternal cotinine levels and accounting for the spatial variation. No significant association was found between ages 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 choices 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 is small. The parameter α_2 measures global spatial decay of dependence with the spatial distance (measured in kilometers). For example, when $\alpha_3 = 0.5$, i.e., under the exponential model, $\alpha_2 = 2.2977$ means the correlation decays by $1 - \exp(-2.2977 \times 1) \doteq 90\%$ for every one kilometer increase in distance. As pointed out by a reviewer, the value of α_2 should be interpreted with caution as its interpretation depends on the unit of distance.

7 Discussion

We have proposed in this paper a semiparametric normal transformation model for spatial survival data. Although statistical methods for clustered survival data and non-censored spatial data have been well developed, little literature is available for modeling censored spatial survival data. However, direct extensions of models for clustered survival data to censored spatial survival data are difficult to be used to construct 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 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 paper is a 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 pair-wise survival times. The proposed method is computationally easy and is shown to yield consistent and asymptotically normal estimators and yield the 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 $\boldsymbol{\alpha}$ 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 $\boldsymbol{\alpha}$ as it characterizes the underlying spatial dependence, which is sometimes of practical interest. Such a quasi-likelihood type estimating equation for $\boldsymbol{\alpha}$ however would involve third and fourth order moments of the martingale residuals $M_{uv}(s)$, whose computation can be difficult. It would be of future research interest to investigate the efficiency loss of the proposed estimator of $\boldsymbol{\alpha}$ 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 of the regression coefficient estimator $\boldsymbol{\beta}$ and the spatial correlation parameter $\boldsymbol{\alpha}$ 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 maximum likelihood estimators and compare the efficiency and robustness of the spatial semiparametric estimating equation based estimators in this paper with the nonparametric maximum likelihood estimators. It is likely that the nonparametric maximum likelihood estimators of the regression coefficients might be sensitive to the misspecification of the spatial correlation structure, while the spatial semiparametric estimating equation based estimators are robust to such misspecifications. On the other hand, if the semiparametric normal transformation model is a true model, the spatial semiparametric estimating equation based estimators might be less efficient than the nonparametric maximum likelihood estimators. More future research is needed.

We have focused in this paper 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 which specify

$$\log T_i = -\boldsymbol{\beta}'\mathbf{Z}_i + \epsilon_i, i = 1, \ldots, m$$

where ϵ_i follows an unspecified distribution. This model is equal to, marginally, $T_i \sim S_0(t\exp(\boldsymbol{\beta}'\mathbf{Z}_i))$, where $S_0(t)$ is an unspecified survival function. Then we define the normal transformation as $T_i^* = \Phi^{-1}\{1 - S_0(T_i \exp(\boldsymbol{\beta}'\mathbf{Z}_i))\}$. Hence T_i^* follows the standard normal distribution marginally. We can then conveniently impose a spatial structure on the underlying random fields of $\mathbf{T}^* = \{T_i^*, i = 1, \ldots, m\}$ 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 as 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 $S_0(\cdot)$. Rank-based procedure along the line of Jin et al. (2003) may need to be adopted. We will pursue this idea in a separate paper.

References

Abramowitz M., and Stegun I.A. (Editors) (1965), *Handbook of Mathematical Functions,* New York: Dover Publications.

Banerjee, S. and Carlin, B.P (2003), "Semiparametric Spatio-temporal Frailty Modeling," *Environmetrics*, 14 (5), 523-535.

Banerjee, S., Carlin, B.P. and Gelfand, A. E (2004), *Hierarchical Modeling and Analysis for Spatial Data,* Boca Raton: Chapman and 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 Statististical Association,* 88, 9-25.

Bulhmann, 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 and Hall Ltd.

Carlstein, E. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence, " *The Annals of Statististics* 14, 1171-1179.

Carlstein, E. (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 mMpping," *Biometrics,* 43, 671-681.

Cressie, N. (1993), *Statistics for Spatial Data*, New York : Wiley.

Foutz, R.V. (1977), "On the Unique Consistent Solution to the Likelihood Equations," *Journal of the American Statististical Association,* 72, 147-148.

Guyon, X. (1995), *Random Fields on a Network: Modeling, Statistics, and Applications,* New York: Springer.

Kalbfleisch, J. D. and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data.* New York: Wiley.

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,* 1875-1894.

Heagerty, P.J. and Lele, S.R. (1998), "A Composite Likelihood Approach to Binary Spatial Data," *Journal of the American Statististical Association,* 93, 1099-1111.

Hougaard, P. (1986), "A Class of Multivariate Failure Time Distributions," *Biometrika*, 73, 671-678

Hougaard, P. (2000) *Analysis of Multivariate Survival Data,* New York: Springer.

Jin, Z., Lin, D., 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 (9 vols. plus Supplement)*, 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," *ASA Proceedings of the Section on Survey Research Methods,* 763-768.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000), *Continuous Multivariate Distributions. Volume 1: Models and applications,* New York: Wiley.

Lahiri, S. N., Kaiser, M. S., Cressie, N. and Hsu, N. (1999), "Prediction of Spatial Cumulative Distribution Functions Using Subsampling (C/R: p97-110)," *Journal of the American Statististical Association,* 94, 86-97.

Lele, S. (1991), "Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes," *Journal of the Royal Statistical Society, Series B,* 53, 253-267.

Lindsay, B. G. (1988) "Composite Likelihood Methods," *Statistical Inference from Stochastic Processes,* 221-239, Prabhu, N. U. (ed.) American Mathematical Society (Providence).

Liu, R. and Singh, K. (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," *Exploring the Limits of Bootstrap. LePage, Raoul (ed.) and Billard, Lynne (ed.)* 225-248. New York: Wiley.

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 Statististics,* 23, 182-98.

Oakes, D. (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statististical Association,* 84, 487-93.

Parner, E. (1998), " Asymptotic theory for the correlated Gamma-frailty model, " *The Annals of Statististics* 26, 183-214.

Parner, E. (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 Statististics,* 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,Series B,* 58, 509-523.

Stein, M.L. (1999) *Interpolation of Spatial Data: Some Theory of Kriging.* New York: Springer.

Sungur, E. A. (1990), "Dependence Information in Parameterized Copulas," *Communications in Statistics - 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 Statististical 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 Statististical Association,* 84, 1065-1073.

Winkler, G. (1995), *Image analysis, random fields, and dynamic Monte Carlo methods: a mathematical introduction.* New York: Springer.

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

Appendix: Technical Details

A.1: A first order expansion of the Martingale Covariance Rate function

Following Moran (1983) and Kotz et al. (2000, eq (45.89)), some algebra show that when θ is sufficiently small, one 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 respectively, and $o(\theta)$ holds uniformly with respect to (z_1, z_2) in any bounded set. Then

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

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

and

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

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 equations (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

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

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,

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

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

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

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

we have

$$
\begin{aligned}
&\int_0^{\tau_1}\int_0^{\tau_2} \theta^{-1}|A_0(t_1,t_2;\theta) - \theta(e^{t_1}\phi(x_1)-x_1)(e^{t_2}\phi(x_2)-x_2)|dt_1dt_2 \\
\le\; &\int_{\epsilon_1}^{\tau_1}\int_{\epsilon_2}^{\tau_2} \theta^{-1}|A_0(t_1,t_2;\theta) - \theta(e^{t_1}\phi(x_1)-x_1)(e^{t_2}\phi(x_2)-x_2)|dt_1dt_2 \\
&+\int_{\epsilon_1}^{\tau_1}\int_0^{\epsilon_2} \theta^{-1}|A_0(t_1,t_2;\theta)|dt_1dt_2 + \int_{\epsilon_1}^{\tau_1} |e^{t_1}\phi(x_1)-x_1)|dt_1 \int_0^{\epsilon_2} |e^{t_2}\phi(x_2)-x_2)|dt_2 \\
&+\int_0^{\epsilon_1}\int_0^{\epsilon_2} \theta^{-1}|A_0(t_1,t_2;\theta)|dt1dt2 + \int_0^{\epsilon_1} |e^{t_1}\phi(x_1)-x_1)|dt_1 \int_0^{\epsilon_2} |e^{t_2}\phi(x_2)-x_2)|dt_1dt_2 \\
&+\int_0^{\epsilon_1}\int_{\epsilon_2}^{\tau_2} \theta^{-1}|A_0(t_1,t_2;\theta)|dt_1dt_2 + \int_0^{\epsilon_1} |e^{t_1}\phi(x_1)-x_1)|dt_1 \int_{\epsilon_2}^{\tau_2} |e^{t_2}\phi(x_2)-x_2)|dt_2.
\end{aligned}
$$

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 $e^{t_k}\phi(x_k) - x_k$ $(k = 1, 2)$ are integrable over any finite interval $[0, \tau_k]$. Therefore, using ϵ-δ-type arguments one can show all the above components converge to 0 as $\theta \to 0$. Hence

$$
\int_0^{\tau_1}\int_0^{\tau_2} A_0(t_1,t_2;\theta)dt_1dt_2 = \theta\int_0^{\tau_1}\{e^{t_1}\phi(x_1)-x_1\}dt_1\int_0^{\tau_2}\{e^{t_2}\phi(x_2)-x_2\}dt_2 + o(\theta).
$$

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, one shall be 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 decay 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

$$
\sum_{n\ge1} n^a|D_{m_n}|^{-1} < \infty,
$$

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 most useful when considering the asymptotics of kriging. Since 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 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 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 $k = 0, 1, \ldots, 3$.

   We assume all the covariates 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 $k = 0, 1,$

   $$\sup_{(\boldsymbol{\beta},t)\in\mathcal{B}\times[0,\tau]} \left\|m^{-1}\sum_{i\geq j}\mathbf{Z}_i(t)w_{11}^{(i,j)}(t)Y_i(t)e^{\boldsymbol{\beta}Z_i(t)}\otimes\mathbf{Z}_i^k(t) - sw_{11}^{(k)}(\boldsymbol{\beta}, t)\right\| \xrightarrow{p} 0,$$

   $$\sup_{(\boldsymbol{\beta},t)\in\mathcal{B}\times[0,\tau]} \left\|m^{-1}\sum_{i\geq j}\mathbf{Z}_i(t)w_{12}^{(i,j)}(t)Y_j(t)e^{\boldsymbol{\beta}Z_j(t)}\otimes\mathbf{Z}_j^k(t) - sw_{12}^{(k)}(\boldsymbol{\beta}, t)\right\| \xrightarrow{p} 0,$$

   $$\sup_{(\boldsymbol{\beta},t)\in\mathcal{B}\times[0,\tau]} \left\|m^{-1}\sum_{i\geq j}\mathbf{Z}_j(t)w_{21}^{(i,j)}(t)Y_i(t)e^{\boldsymbol{\beta}Z_i(t)}\otimes\mathbf{Z}_i^k(t) - sw_{21}^{(k)}(\boldsymbol{\beta}, t)\right\| \xrightarrow{p} 0,$$

   $$\sup_{(\boldsymbol{\beta},t)\in\mathcal{B}\times[0,\tau]} \left\|m^{-1}\sum_{i\geq j}\mathbf{Z}_j(t)w_{22}^{(i,j)}(t)Y_j(t)e^{\boldsymbol{\beta}Z_j(t)}\otimes\mathbf{Z}_j^k(t) - sw_{22}^{(k)}(\boldsymbol{\beta}, t)\right\| \xrightarrow{p} 0.$$

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

2. (Boundness) Assume the covariate processes $Z_i(\cdot)$ and weights $w_{ij}^{(uv)}(\cdot), \mathbf{v}_{ij}$ are uniformly bounded. In addition, $\Lambda_0(\tau) < \infty$ for $\tau < \infty$.

3. (Differentiability) Assume that the covariance function A_0 is at least twice differentiable.

4. (Positive Definiteness of the Information) Assume matrix $\mathbf{\Sigma} = (\mathbf{\Sigma}_{ij})_{2\times 2}$, has positive eigenvalues, where

$$
\begin{aligned}
\mathbf{\Sigma}_{11} &= \sum_{p=1}^{2}\sum_{q=1}^{2}\int_0^\tau \left\{ sw_{pq}^{(1)}(\boldsymbol{\beta}_0,t) - sw_{pq}^{(0)}(\boldsymbol{\beta}_0,t)\otimes \frac{s^{(1)}(\boldsymbol{\beta}_0,t)}{s^{(0)}(\boldsymbol{\beta}_0,t)}\right\} d\Lambda_0(t),\\
\mathbf{\Sigma}_{12} &= 0,\\
\mathbf{\Sigma}_{21} &= \int_0^\tau \lim_{m\to\infty} m^{-1}\sum_{i\geq j}\mathbf{v}_{ij}\otimes\left[\left\{\mathbf{Z}_i(t)-\frac{s^{(1)}(\boldsymbol{\beta}_0,t)}{s^{(0)}(\boldsymbol{\beta}_0,t)}\right\}Y_i(t)e^{\boldsymbol{\beta}_0'\mathbf{Z}_i(t)}\{M_j(X_j\wedge\tau)+A_{ij}^{(100)}\}\right.\\
&\quad \left.+\int_0^\tau\left\{\mathbf{Z}_j(t)-\frac{s^{(1)}(\boldsymbol{\beta}_0,t)}{s^{(0)}(\boldsymbol{\beta}_0,t)}\right\}Y_j(t)e^{\boldsymbol{\beta}_0'\mathbf{Z}_j(t)}\{M_i(X_i\wedge\tau)+A_{ij}^{(010)}\}\right] d\Lambda_0(t),\\
\mathbf{\Sigma}_{22} &= \lim_{m\to\infty} m^{-1}\sum_{i>j}\mathbf{v}_{ij}\otimes\mathbf{A}_{ij}^{(001)},
\end{aligned}
$$

for $i=j$

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

for $i\neq j$

$$
\begin{aligned}
A_{ij}^{(100)} &= \int_0^{\Lambda_i(\tau\wedge X_i)}\int_0^{\Lambda_j(\tau\wedge X_j)} A_0^{(100)}\{s_1,s_2;\theta_{ij}(\boldsymbol{\alpha}_0)\}ds_1ds_2\\
A_{ij}^{(010)} &= \int_0^{\Lambda_i(\tau\wedge X_i)}\int_0^{\Lambda_j(\tau\wedge X_j)} A_0^{(010)}\{s_1,s_2;\theta_{ij}(\boldsymbol{\alpha}_0)\}ds_1ds_2
\end{aligned}
$$

and

$$\mathbf{A}_{ij}^{(001)} = \int_0^{\Lambda_i(\tau\wedge X_i)}\int_0^{\Lambda_j(\tau\wedge X_j)} A_0^{(010)}\{s_1,s_2;\theta_{ij}(\boldsymbol{\alpha}_0)\}\frac{\partial}{\partial\boldsymbol{\alpha}}\theta_{ij}(\boldsymbol{\alpha}_0)ds_1ds_2.$$

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 the limits above are the probabilistic limits (provided existence) when $m\to\infty$.

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

A.3: Sketchy Proof of Proposition 1

We provide in this section 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 Foutz (1977), namely (1) asymptotic unbiasedness of the estimating equation, i.e. $G_m(\mathbf{\Theta}_0) \xrightarrow{p} 0$; (2) existence, continuity and uniform convergence of the partial derivatives of the estimating equations in a neighborhood of the true parameters, i.e. $(\partial/\partial\mathbf{\Theta})G_m(\mathbf{\Theta})$ converges uniformly in a neighborhood of $\mathbf{\Theta}_0$; and (3) the negative definiteness of the the partial derivatives of the estimating equations at the true values, i.e. $(\partial/\partial\mathbf{\Theta})G_m(\mathbf{\Theta}_0)$ converges in probability to a matrix with strictly negative eigenvalues.

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

$$\hat{\mathbf{U}}_1^{(1)} = m^{-1}\sum_{u\geq v}\int_0^\tau \mathbf{Z}_u(t)w_{11}^{(uv)}(t)d\hat{M}_u(t), \quad \hat{\mathbf{U}}_2^{(1)} = m^{-1}\textstyle\sum_{u\geq v}\int_0^\tau \mathbf{Z}_u(t)w_{12}^{(uv)}(t)d\hat{M}_v(t),$$

$$\hat{\mathbf{U}}_3^{(1)} = m^{-1}\sum_{u\geq v}\int_0^\tau \mathbf{Z}_v(t)w_{21}^{(uv)}(t)d\hat{M}_u(t), \quad \hat{\mathbf{U}}_4^{(1)} = m^{-1}\textstyle\sum_{u\geq v}\int_0^\tau \mathbf{Z}_v(t)w_{22}^{(uv)}(t)d\hat{M}_v(t),$$

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

$$\begin{aligned}
\hat{\mathbf{U}}_1^{(1)} &= m^{-1}\sum_{u\geq v}\int_0^\tau \mathbf{Z}_u(t)w_{11}^{(uv)}(t)dM_u(t) - m^{-1}\sum_{u\geq v}\int \mathbf{Z}_u(t)Y_u(t)e^{\boldsymbol{\beta}'\mathbf{Z}_u(t)}w_{11}^{(uv)}(t)\frac{\sum_{i=1}^m dM_i(t)}{\sum_{i=1}^m Y_i(t)e^{\boldsymbol{\beta}_0'\mathbf{Z}_i(t)}} \\
&= \frac{1}{m}\sum_{i=1}^m\int_0^\tau\left\{\mathbf{Z}_i(t)\{\sum_{j\leq i}w_{11}^{(ij)}(t)\} - \frac{sw_{11}^{(0)}(t)}{s^{(0)}(t)}\right\}dM_i(t) &\text{(A. 2)} \\
&\quad -\frac{1}{m}\sum_{i=1}^m\int_0^\tau\frac{1}{s^{(0)}(t)}\left\{m^{-1}\sum_{u\geq v}\mathbf{Z}_u(t)Y_u(t)e^{\boldsymbol{\beta}_0'\mathbf{Z}_u(t)}w_{11}^{(uv)}(t) - sw_{11}^{(0)}(\boldsymbol{\beta}_0,t)\right\}dM_i &\text{(A. 3)} \\
&\quad -\frac{1}{m}\sum_{i=1}^m\int sw_{11}^{(0)}(t)\left[(m^{-1}\sum_{i=1}^m Y_u(t)e^{\boldsymbol{\beta}_0'\mathbf{Z}_u(t)})^{-1} - \{s^{(0)}(t)\}^{-1}\right]dM_i(t) &\text{(A. 4)} \\
&\quad -\frac{1}{m}\sum_{i=1}^m\int_0^\tau\{m^{-1}\sum_{u\geq v}\mathbf{Z}_u(t)Y_u(t)e^{\boldsymbol{\beta}_0'\mathbf{Z}_u(t)}w_{11}^{(uv)}(t) - sw_{11}^{(0)}(t)\} \\
&\quad \times\left[(m^{-1}\sum_{i=1}^m Y_u(t)e^{\boldsymbol{\beta}_0'\mathbf{Z}_u(t)})^{-1} - \{s^{(0)}(t)\}^{-1}\right]dM_i(t) &\text{(A. 5)}
\end{aligned}$$

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, one can show that $\hat{\mathbf{U}}_2^{(1)}, \hat{\mathbf{U}}_3^{(1)}, \hat{\mathbf{U}}_4^{(1)}$ all converge to 0 in probability. One can also show that $\hat{U}^{(2)}$ is asymptotically equivalent to

$$\mathbf{U}^{(2)} = \frac{1}{m}\sum_{u \geq v} \mathbf{v}_{uv} \left\{ M_u(\tau)M_v(\tau) - \int_0^\tau \int_0^\tau Y_u(s)Y_v(t)A_0\{\Lambda_u(s), \Lambda_v(t)\} d\Lambda_u(s) d\Lambda_v(t) \right\},$$

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

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

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

in probability. The same arguments also apply to $\partial \hat{U}_2^{(1)}/\partial \boldsymbol{\beta}$, $\partial \hat{U}_3^{(1)}/\partial \boldsymbol{\beta}$, and $\partial \hat{U}_4^{(1)}/\partial \boldsymbol{\beta}$, which converge uniformly in a neighborhood of $\mathbf{\Theta}_0$. Hence, in particular, at $\mathbf{\Theta}_0$, the $(1,1)$th block of $\partial G_m(\mathbf{\Theta})/\partial \mathbf{\Theta}$ converges to $-\mathbf{\Sigma}_{11}$. Similarly, we can show that other blocks of $\partial G_m(\mathbf{\Theta})/\partial \mathbf{\Theta}$ converges uniformly at $\mathbf{\Theta}_0$ to $-\mathbf{\Sigma}$, 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 $\mathbf{\Theta}_0$, there exists a unique sequence of $\hat{\mathbf{\Theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})'$ such that $G_m(\hat{\mathbf{\Theta}}) = 0$ with probability going to 1 and $\hat{\mathbf{\Theta}} \xrightarrow{p} \mathbf{\Theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\alpha}_0')'$.

We now consider the asymptotic normality of $\hat{\mathbf{\Theta}}$. A Taylor expansion of $G_m(\hat{\mathbf{\Theta}})$ at the true value $\mathbf{\Theta}_0$ gives

$$-\sqrt{m}\left\{ \frac{\partial}{\partial \mathbf{\Theta}} G_m(\mathbf{\Theta}^*) \right\} (\hat{\mathbf{\Theta}} - \mathbf{\Theta}_0) = \sqrt{m} G_m(\mathbf{\Theta}_0),$$

where $\mathbf{\Theta}^*$ is on the segment between $\mathbf{\Theta}_0$ and $\hat{\mathbf{\Theta}}$. With condition 5 and the assumed spatial dependence, a central limit theorem (Guyon, chap. 3, 1995) applies to the sequence of $G_m(\mathbf{\Theta}_0)$ such that

$$\sqrt{m}\{\mathbf{\Sigma}^{(2)}\}^{-1/2} G_m(\mathbf{\Theta}_0) \to N(0, \mathbf{I})$$

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

$$\sqrt{m}\{\mathbf{\Sigma}^{(2)}\}^{-1/2} \mathbf{\Sigma}(\hat{\mathbf{\Theta}} - \mathbf{\Theta}_0) \to N(0, \mathbf{I})$$

in distribution.

Table 1: Simulation results based on 500 runs. 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 $\beta_1 = 1, \beta_2 = \beta_3 = 0.5, \alpha_1 = 0.5, \alpha_2 = 2.5$. Both the empirical ($SE_e$) and estimated standard errors (SE_a) are reported, along with the 95% coverage probabilities.

Sample Size	α_3	censoring	Parameter	Estimate	SE_e	SE_a	cov prob
100	0.5	70%	β_1	0.9909	0.2491	0.2456	91.5%
			β_2	0.5068	0.2138	0.1933	99.0%
			β_3	0.5044	0.2149	0.1916	92.6%
			α_1	0.4789	0.1827	0.1915	89.2%
			α_2	2.0555	0.9275	0.7994	73.0%
	0.5	50%	β_1	0.9920	0.1971	0.2033	92.9%
			β_2	0.5134	0.1731	0.1628	92.5%
			β_3	0.4831	0.1702	0.1548	90.4%
			α_1	0.4656	0.1520	0.1533	90.8%
			α_2	2.1292	0.9958	0.8916	79.0%
	1	70%	β_1	0.9836	0.2511	0.2467	90.3%
			β_2	0.5112	0.2127	0.1897	89.1%
			β_3	0.5113	0.2066	0.1936	91.5%
			α_1	0.4767	0.1814	0.1935	90.6%
			α_2	2.3043	0.9685	0.7941	71.9%
	1	50%	β_1	1.007	0.2052	0.2114	91.2%
			β_2	0.5139	0.1845	0.1659	90.6%
			β_3	0.4986	0.1699	0.1566	91.2%
			α_1	0.4796	0.1539	0.1507	88.0%
			α_2	2.358	1.0239	0.8176	74.3%
200	0.5	70%	β_1	0.9869	0.1556	0.1702	94.7%
			β_2	0.4940	0.1341	0.1312	92.8%
			β_3	0.4882	0.1421	0.1323	92.8%
			α_1	0.4902	0.1400	0.1397	92.2%
			α_2	2.3575	1.0620	0.9496	80.4%
	0.5	50%	β_1	0.9819	0.1300	0.1441	95.1%
			β_2	0.4951	0.1133	0.1100	93.8%
			β_3	0.4792	0.1223	0.1100	90.2%
			α_1	0.4990	0.1011	0.1094	92.2%
			α_2	2.4218	0.9966	0.8051	82.4%
	1	70%	β_1	0.9993	0.1740	0.1688	91.9%
			β_2	0.4909	0.1387	0.1245	92.3%
			β_3	0.4960	0.1303	0.1296	92.5%
			α_1	0.5118	0.1395	0.1380	90.0%
			α_2	2.6356	1.1288	0.9811	80.1%
	1	50%	β_1	0.9838	0.1437	0.1395	92.1%
			β_2	0.4830	0.1204	0.1066	90.2%
			β_3	0.4803	0.1174	0.1076	90.2%
			α_1	0.5136	0.1113	0.1042	92.7%
			α_2	2.5033	1.0106	0.7608	84.1%
400	0.5	70%	β_1	0.9616	0.1055	0.1285	94.0%
			β_2	0.4960	0.1040	0.1032	96.0%
			β_3	0.5083	0.0970	0.0990	94.0%
			α_1	0.5077	0.1093	0.0963	94.2%
			α_2	2.4439	1.2866	1.0205	90.4%

Table 2: Results of analysis of the East Boston Asthma Study under the normal transformation model assuming the Matèrn correlation and the marginal Cox model. Estimates were calculated by the spatial semiparametric estimating equation method and the large sample standard errors (SE_a) were computed using the method described in Section 4.3 and the Jackknife standard errors (SE_j) were computed using the formulation (13) in Section 6.

	$\alpha_3 = 0.5$			$\alpha_3 = 1$			$\alpha_3 = 1.5$		
Parameters	Estimate	SE_a	SE_j	Estimate	SE_a	SE_j	Estimate	SE_a	SE_j
β_L	0.3121	0.0440	0.0357	0.3118	0.0430	0.0369	0.3124	0.0432	0.0349
β_M	0.2662	0.3314	0.3222	0.2644	0.3289	0.3309	0.2676	0.3283	0.3340
β_C	0.0294	0.1394	0.1235	0.02521	0.1270	0.1063	0.0277	0.1288	0.1083
α_1	1.68E-3	9.8E-3	0.0127	0.74E-3	5.0E-3	7.1E-3	0.72E-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

Figure 1: The cross ratio at $(t_1, t_2) = (0.5, 0.5)$ as a function of the correlation coefficient of the transformed survival times θ_{ij}. The individual survival times marginally follow a unit exponential distribution.

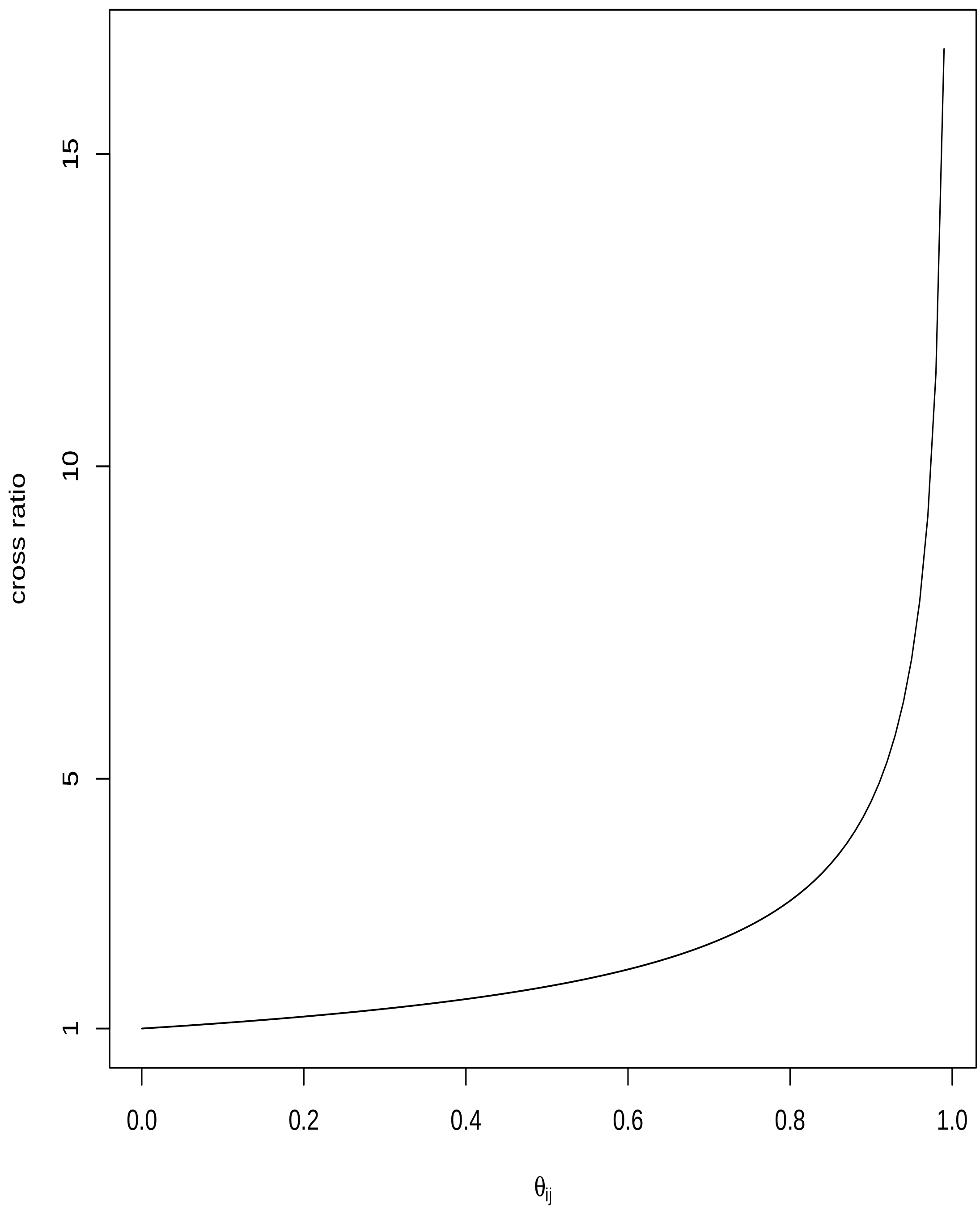

Figure 2: The empirical density plots of the model parameter estimates from the simulation study when $m = 200, \alpha_3 = 0.5$ and censoring proportion =50%