Summer 2022

Semiparametric estimation with clustered right censored data via multivariate gaussian random fields

Fathima Zahra Sainul Abdeen

SEMIPARAMETRIC ESTIMATION WITH CLUSTERED RIGHT CENSORED DATA VIA MULTIVARIATE GAUSSIAN RANDOM FIELDS

by

FATHIMA ZAHRA SAINUL ABDEEN

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS WITH STATISTICS EMPHASIS

2022

Approved by:

Dr. Akim Adekpedjou, Advisor
Dr. V. A. Samaranayake
Dr. Gayla Olbricht
Dr. Martin Bohner
Dr. Maciej Zawodniok

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following two articles, formatted in the style used by the Missouri University of Science and Technology.

Paper I, Pages 14-69, is under review in Metrika.

Paper II, Pages 70-101, is intended for submission to Canadian Journal of Statistics.

ABSTRACT

Consider a fixed number of clustered areas identified by their geographical coordinate that are monitored for the occurrences of an event such as pandemic, epidemic, migration to name a few. Data collected on units at all areas include time varying covariates and other environmental factors that may affect event occurrences. The event times in every area can be independent. They can also be correlated with correlation between two units induced by an unobservable frailty. In both cases, the collected data is considered pairwise to account for spatial correlation between all pair of areas. The pairwise right censored data is probit-transformed yielding a multivariate Gaussian random field preserving the spatial correlation function. The data is analyzed using counting process and geostatistical formulation that led to a class of weighted pairwise semiparametric estimating functions. In the independence case, estimators of models unknowns are shown to be consistent and asymptotically normally distributed under infill-type spatial statistics asymptotic. Detailed small sample numerical studies that are in agreement with the theoretical results are provided in the independence case. In the dependence case, the estimators are shown to be inefficiency when the dependence is ignored. The foregoing procedures are applied to Leukemia survival data in Northeast England.

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my advisor, Dr. Akim Adekpedjou, for his unconditional guidance and support that I received throughout the duration of my stay at Missouri University of Science and Technology. I was fortunate to have had an advisor who is so passionate, motivated and committed to research. Without you, this dissertation would not have been possible. I would also like to give special thanks to Dr. Sophie D-Niang for her valuable inputs over the course of this research work.

I am grateful to Dr. V.A. Samaranayake, Dr. Gayla Olbricht, Dr. Martin Bohner and Dr. Maciej Zawodniok for serving in my committee. Their valuable inputs and comments helped improve my dissertation. I extend my sincere gratitude to all the faculty and the staff of the Department of Mathematics and Statistics, Missouri University of Science and Technology. The financial assistance of the Department of Mathematics and Statistics during the course of my graduate studies is gratefully acknowledged.

In addition, I would also like to thank my father, mother and my younger siblings for motivating me and supporting me to pursue my PhD in the United States. I also want to thank my children Aleina, Shayaan and Yara for all the emotional support. Lastly, I would like to thank my husband for his support and encouragement. I could not have completed this task without his assistance and enthusiasm.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

LIST OF TABLES

Table Page

1. INTRODUCTION

In this section of the dissertation, a detailed description of essential mathematical preliminaries is provided for the reader to get a better understanding of the concepts used.

1.1. MATHEMATICAL PRELIMINARIES FOR SURVIVAL ANALYSIS

The pioneering work by Aalen (1978) on the theory of counting processes has been the key to the development of statistical tools for analyzing data in reliability and survival analysis. A detailed discussion of these topics can be found in Andersen et al. (2012), Chung et al. (1990) and Fleming and Harrington (2011).

Let $(\Omega, \mathscr{F}, P)$ be a complete probability space and $T = [0, \tau] \subset \mathbb{R}$ be an interval of time.

Definition 1.1.1 *A filtration $\boldsymbol{F} = \{\mathscr{F}_t, t \in T\}$ on $(\Omega, \mathscr{F}, P)$ is an increasing family of σ-algebras, that is, $\forall t \leq s$, $\mathscr{F}_t \subseteq \mathscr{F}_s \subseteq \mathscr{F}$.*

Note here that in the case of a stochastic process, $\mathscr{F}_t$ could be taken to be all information generated by the process up to time t, and is called the natural history of the process. From now on, the natural filtration associated with the probability space $(\Omega, \mathscr{F}, P)$ will be denoted by $\mathbf{F}$.

Definition 1.1.2 *A stochastic process $\boldsymbol{X} = \{X_t, t \geq 0\}$ is called cadlag if its simple paths $\{\boldsymbol{X}(t, w) : t \in T\}$ are right continuous with left hand limits for almost all w. Furthermore, the set of all cadlag functions is called the Skorohod space.*

Definition 1.1.3 *A counting process is a stochastic process $\{N(t) : t \geq 0\}$ adapted to a filtration $\boldsymbol{F}$ with $N(0) = 0$ and $N(t) < \infty$ almost surely (a.s), and whose paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with jumps of size +1.*

Definition 1.1.4 *A stochastic process* $\boldsymbol{X} = \{X_t, t \geq 0\}$ *is:*

1. *Integrable if* $\sup_{t\in T} E(X(t)) < \infty$*,*

2. *Square integrable if* $\sup_{t\in T} E(X(t)^2) < \infty$*,*

3. *Bounded if there exists a finite constant* Γ *such that* $P\left\{\sup_{t\in T} |X(t)| < \Gamma\right\} = 1$.

From now on, cadlag stochastic processes will only be considered.

Definition 1.1.5 *A collection* $\boldsymbol{M} = \{M_t, t \geq 0\}$ *is an* $\boldsymbol{F}$*-martingale if* $\boldsymbol{M}$ *is* $\boldsymbol{F}$*-adapted and satisfies:*

1. *Integrability:* $E(|M_t|) < \infty$ *for all* $t \in T$*,*

2. *Martingale property:* $E(M_t|\mathscr{F}_s) = M_s$ *a.s* $\forall s < t$.

A sub martingale is obtained if (2) in previous definition is replaced by $E(M_t|\mathscr{F}_s) \geq M_s$ $a.s$ $\forall s < t$. On the other hand, a super martingale is obtained by replacing (2) in previous definition by $E(M_t|\mathscr{F}_s) \leq M_s$ $a.s$ $\forall s < t$.
Now, the notion of a predictable process is discussed.

Definition 1.1.6 *The* σ*-algebra generated by all the sets of the form:*

1. $[0] \times A, A \in \mathscr{F}_0$ *and,*

2. $(a, b] \times A, 0 \leq a < b < \infty, A \in \mathscr{F}_a$,

is called the predictable σ*-algebra for* $\boldsymbol{F}$*, where* $\mathscr{F}_0$ *is the information at time 0.*

Lemma 1.1.1 *Let* $\boldsymbol{F}$ *be a filtration, and* $\boldsymbol{X}$ *a left-continuous real-valued process adapted to* $\boldsymbol{F}$*. Then* $\boldsymbol{X}$ *is predictable.*

Proposition 1.1.1 *Let* $\boldsymbol{X}$ *be an* $\mathscr{F}_t$*-predictable process. Then, for any* $t > 0$*,* $X(t)$ *is* $\mathscr{F}_t$*-measurable.*

An important theorem that allows decomposing a submartingale is discussed next.

Theorem 1.1.1 ***Doob-Meyer Decomposition*** *Let* $\boldsymbol{M} = \{M_t, t \geq 0\}$ *be a right continuous, nonnegative submartingale with respect to filtration* $\boldsymbol{F}$. *Then, there exists a right-continuous martingale* $\mathscr{M}(t)$ *and an increasing right-continuous predictable process* $A(t)$ *such that* $M(t) = \mathscr{M}(t) + A(t)$ *a.s.*

Note that, if M is a martingale with $E(M^2(t)) < \infty$ for $t > 0$, Jensen's inequality indicates that $M^2(t)$ is a submartingale.

Corollary 1.1.1 *Let* $\boldsymbol{M}$ *be a cadlag martingale with respect to* $\boldsymbol{F}$. *Then, there exists a unique increasing right-continuous predictable process denoted by* $\langle \boldsymbol{M}, \boldsymbol{M} \rangle (t)$ *called the predictable quadratic variation process of* $\boldsymbol{M}$, *such that* $\langle \boldsymbol{M}, \boldsymbol{M} \rangle (0) = 0$ *a.s,* $E \langle \boldsymbol{M}, \boldsymbol{M} \rangle (t) < \infty$ *for all t and* $\left\{\boldsymbol{M}^2(t) - \langle \boldsymbol{M}, \boldsymbol{M} \rangle (t) : t \geq 0\right\}$ *is a right continuous martingale.*

Notion of stochastic integration is presented next. A detailed discussion can be found in Chung et al. (1990).

Theorem 1.1.2 *Suppose* M *is a finite variation local square integrable martingale,* H *a predictable process and* $\int_0^t H^2 d\langle M \rangle$ *locally integrable. Then,* $\int_0^t H dM$ *is a local square integrable martingale and its quadratic variation process is given by*

$$\left\langle \int H dM \right\rangle (t) = \int_0^t H^2 d\langle M \rangle.$$

The above theorem can be further generalized to a vector of martingales **M** and **M'** and matrices **H** and **K** of predictable processes. In that case, the predictable covariation process is given by

$$\left\langle \int \mathbf{H} d\mathbf{M}, \int \mathbf{K} d\mathbf{M'} \right\rangle = \int_0^t \mathbf{H} d \langle \mathbf{M}, \mathbf{M'} \rangle \mathbf{K}^t$$

where A^t denotes the transpose of a matrix A.

Definition 1.1.7 *Suppose a filtration* $\boldsymbol{F}$ *on* $(\Omega, \mathscr{F}, P)$ *is given. A multivariate counting process* $\boldsymbol{N} = (N_1, \ldots, N_k)$ *is a vector of k* $\boldsymbol{F}$*- adapted cadlag processes for which:*

1. $N_i = 0 \;\; \forall i = 1, 2, \ldots, k$*,*

2. Their jumps are of size one and no two components can jump at the same time,

3. Their paths are nondecreasing and piecewise constant.

Note that because the components of the counting process $\mathbf{N}$ are adapted, cadlag, locally bounded and nondecreasing, they are local submartingales. So, by the Doob-Meyer decomposition, there exists a compensator of N_i, say Λ_i. Λ_i is referred to as the cumulative intensity process of the counting process.
The following proposition makes the important connection among counting processes, martingales and stochastic integration which is crucial in this work.

Proposition 1.1.2 *Let* $\boldsymbol{N}$ *be a multivariate counting process and let* $\boldsymbol{\Lambda} = \int \lambda$ *be its associated vector of compensator processes such that each component of* $\boldsymbol{\Lambda}$ *is absolutely continuous. Let* $\boldsymbol{M} = \boldsymbol{N} - \boldsymbol{\Lambda}$ *be the resulting vector of local martingales. If* $\boldsymbol{H}$ *is a vector of locally bounded and predictable processes, then* $\int \boldsymbol{H} d\boldsymbol{M}$ *are vectors of local square integrable martingales with a quadratic variation process given by*

$$\left\langle \int \boldsymbol{H} d\boldsymbol{M} \right\rangle = \int \boldsymbol{H} diag\left\{\boldsymbol{\lambda}\right\} \boldsymbol{H}^t ds,$$

where $diag\left\{\boldsymbol{\lambda}\right\}$ *is the diagonal matrix of associated intensity processes.*

The idea of constructing likelihood with counting process data was first introduced by Jacod (1975). Considering counting process data, the likelihood function can be written in a product integral form, which is a continuous version of the simple product Π.

Let $\Delta N_i(t) = N_i(t) - N_i(t-)$ be the jump process, and let the intensity process depends on some p-dimensional parameter θ. Then, the likelihood in $[0, t]$ can be written as

$$L(\theta, t) = \prod_{i=1}^{n} \prod_{v \in [0,t]} \left\{\lambda_i(v, \theta)^{\Delta N_i(v)} \times (1 - \lambda_i(v, \theta))^{1 - \Delta N_i(v)}\right\}, \tag{1.1}$$

where $N_i(t)$ is the counting process for each individual i in the study and $\lambda_i(t,\theta)$ is the hazard rate at time t which is a function of θ for a parametric model. Simplifying (1.1) using Taylor expansion and noting $1-\lambda_i(v,\theta)dv \approx \exp(-\lambda_i(v,\theta))dv$, we obtain

$$L(\theta,t) \propto \prod_{i=1}^{n}\left[\prod_{v\in[0,t]}\left\{\lambda_i(v,\theta)^{\Delta N_i(v)}\right\}\times \exp\left\{-\int_0^t \lambda_i(v,\theta)dv\right\}\right]. \tag{1.2}$$

Next, by taking the logarithm of (1.2), the log-likelihood process is obtained given by

$$l(\theta,t) = \sum_{i=1}^{n}\left\{\int_0^t \log[\lambda_i(v,\theta)]dN_i(v) - \int_0^t \lambda_i(v,\theta)dv\right\}. \tag{1.3}$$

The score process $U_\theta(\theta,t)$ is obtained by taking the gradient of (1.3) with respect to θ.

$$\begin{aligned} U_\theta(\theta,t) &= \sum_{i=1}^{n}\left\{\int_0^t \frac{\partial}{\partial\theta}\log[\lambda_i(v,\theta)]dN_i(v) - \int_0^t \frac{\partial}{\partial\theta}\lambda_i(v,\theta)dv\right\} \\ &= \sum_{i=1}^{n}\left\{\int_0^t \frac{\partial}{\partial\theta}\log[\lambda_i(v,\theta)]dM_i(v)\right\}. \end{aligned}$$

A result which is key to obtaining asymptotic properties of the estimators is presented next.

Theorem 1.1.3 ***Rebolledo's Martingale Central Limit Theorem*** *For each $n = 1,2,\ldots$, let $\boldsymbol{M}^{(n)} = \left(M_1^{(n)}, M_2^{(n)}, \ldots, M_k^{(n)}\right)$ be vectors of local square-integrable martingales where each may be defined on different sample spaces with respect to different filtration. For $\epsilon > 0$, let $\boldsymbol{M}_\epsilon^{(n)}$ be a vector of local square integrable martingales such that $\left|M_h^{(n)} - M_{\epsilon h}^{(n)}\right|$ is a local square integrable martingale and $\left|\Delta M_h^{(n)} - \Delta M_{\epsilon h}^{(n)}\right| \le \epsilon$. Let $\left\langle \boldsymbol{M}^{(n)}\right\rangle, n = 1,2,\ldots$ be the $k\times k$ matrix processes with elements $\left\langle M_h^{(n)}, M_{h'}^{(n)}\right\rangle$. Assume the following conditions for $T_0 \subseteq T$:*

1. *There exists a matrix of deterministic functions $\boldsymbol{V}(t)$ such that $\left\langle \boldsymbol{M}^{(n)}\right\rangle(t) \xrightarrow{p} \boldsymbol{V}(t), \forall t \in T_0$, as $n \longrightarrow \infty$,*

2. $\left\langle \boldsymbol{M}_{\epsilon h}^{(n)}(t)\right\rangle \xrightarrow{p} 0, \forall t \in T_0$, *$h$ and $\epsilon > 0$ as $n \longrightarrow \infty$.*

Then

$$\left(M^{(n)}(t_1), \ldots, M^{(n)}(t_k)\right) \xrightarrow{d} \left(M^{(\infty)}(t_1), \ldots, M^{(\infty)}(t_k)\right), \forall t_1, \ldots, t_k \in T_0.$$

Moreover, if T_0 is dense in T and contains τ if $\tau \in T$, then the same conditions imply that

$$\boldsymbol{M}^{(n)} \xrightarrow{d} \boldsymbol{M}^{(\infty)} \;\; in \;\; D(T)^k \;\; as \;\; n \to \infty,$$

where $\boldsymbol{M}^{(\infty)}$ is a vector of continuous Gaussian martingales.

1.2. MATHEMATICAL PRELIMINARIES FOR SPATIAL STATISTICS

This subsection contains essential preliminaries on spatial statistics.

Definition 1.2.1 *Let:*

- $S \subset \mathbb{R}^d$ *be a spatial set,*
- $(\Omega, \mathscr{F}, P)$ *be a probability space,*
- (E, ϵ) *be s measurable set.*

A random field X, also called a spatial process, is a family $X = \{X_s, s \in S\}$ of random variables indexed by $s \in S$ from $(\Omega, \mathscr{F}, P)$ to (E, ϵ), where S denotes the spatial set of sites and E denotes the state space of the process.

Definition 1.2.2 *A Gaussian random field X on S is a process such that, for all finite subset ξ of S and all sequence of reals $a = (a_s, s \in \xi)$, the random variable $\sum_{s\in\xi} a_s X_s$ has a Gaussian distribution.*

Definition 1.2.3 *A spatial process $X = \{X_s, s \in S\}$ is said to be second order, if for all s in S, we have $E[X_s^2] < +\infty$, in this case, one can consider the mean function*

$$\begin{aligned} m : S &\rightarrow \mathbb{R} \\ s &\mapsto m(s) = E[X_s] \end{aligned}$$

and the covariance function

$$\begin{aligned} c : S \times S &\rightarrow \mathbb{R} \\ (s,t) &\mapsto c(s,t) = Cov(X_s, X_t). \end{aligned}$$

Definition 1.2.4 *A second-order random field X on S is said to be stationary if it has a constant mean function and its covariance function is invariant by translation, i.e.*

$$\begin{aligned} \forall s \in S : m(s) = m \\ \forall (s,t) \in S^2,\ \forall h \in S : c(s+h, t+h) = c(s,t) \end{aligned}$$

Definition 1.2.5 *If X is stationary, the function*

$$\begin{aligned} C : S &\rightarrow \mathbb{R} \\ h &\mapsto C(h) = c(0,h) \end{aligned}$$

is called a stationary covariance function.

Definition 1.2.6 *The stationary correlation function of a stationary random field X is*

$$\begin{aligned} \rho : S &\rightarrow \mathbb{R} \\ h &\mapsto \rho(h) = \frac{C(h)}{C(0)}. \end{aligned}$$

Proposition 1.2.1 *Let C be the stationary function of second-order spatial process. Then:*

1. $C(h) = C(-h)$ *(even function),*

2. $\forall h \in S$: $|C(h)| \leq C(0)$ *(bounded function),*

3. *If C is continuous at the origin, then it is uniformly continuous on S.*

Proposition 1.2.2 *Let C be the stationary function of second-order spatial process. Then:* $\forall n \geq 1, \forall a \in \mathbb{R}^n, \forall (s_1, ..., s_n) \in S^n : \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j C(s_i - s_j) \geq 0.$

Proposition 1.2.3 *Let C be the stationary function of a second-order spatial process. Then:*

1. *If A is a linear function from* $\mathbb{R}^d$ *to* $\mathbb{R}^d$*, the random field* $X^A = \{X_{As}, s \in S\}$ *is stationary with covariance function* $C^A(s) = C(As)$.

2. *If* $C_1, ..., C_n$ *are stationary functions, then:*

   - $\forall (\alpha_1, \alpha_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ *the function* $C(h) = \alpha_1 C_1(h) + \alpha_2 C_2(h)$ *is a stationary covariance function,*

   - $C(h) = C_1(h)C_2(h)$ *is a stationary covariance function,*

   - $\lim_{n \to +\infty} C_n(h) = C(h)$ *exists for all h, then C is also a stationary covariance function.*

Proposition 1.2.4 *A covariance function is positive semidefinite if* $\forall n \geq 1$, $\forall (s_1, ..., s_n) \in S^n$ *and* $\forall a = (a_1, ..., a_n) \in \mathbb{R}^n$, $\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j c(s_i, s_j) \geq 0$.

Proposition 1.2.5 *The covariance function is positive definite if* $\forall n \geq 1$, $\forall (s_1, ..., s_n) \in S^n$ *where* $s_1, ..., s_n$ *are distincts,* $\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j c(s_i, s_j) = 0 \Longleftrightarrow (a_i = 0, \forall i = 1, ..., n)$.

Definition 1.2.7 *A spatial process X is said to be strictly stationary if* $\forall k \in \mathbb{N}$, $\forall (t_1, ..., t_k) \in S^k$ *and* $\forall h \in S$ *the distribution of the random vector* $(X_{t_1+h}, ..., X_{t_k+h})$ *is independent of h.*

Definition 1.2.8 *A second-order spatial process X has an isotropic covariance function if $Cov(X_s, X_t)$ depends only on $||t-s||$, i.e. if there exists a function C_0 from $\mathbb{R}^+$ to $\mathbb{R}$ such that $c(s,t) = C_0(||s-t||)$ for all $(s,t) \in S^2$. Here $||\cdot||$ denotes the euclidean norm on $\mathbb{R}^d$.*

Definition 1.2.9 *A spatial process X is said to be intrinsically stationary or intrinsic if the processes*

$$\Delta X^h = \left\{\Delta X_S^h = X_{s+h} - X_s; s \in S\right\}$$

are stationary, for all $h \in S$.

Definition 1.2.10 *A spatial process X is said to be intrinsic if its increments are of order two and such that:*

$$\forall (s,h) \in S^2 : E(X_{s+h} - X_s) = 0,$$

$$\forall s \in S : Var(X_{s+h} - X_s) = 2\gamma(h).$$

The function γ is called the semi-variogram function of X.

Definition 1.2.11 *The semi-variogram γ of a spatial process X is said to be isotropic if there exists a function γ_0 such that, $\gamma(h) = \gamma_0(||h||)$ for all $h \in S$.*

Proposition 1.2.6 *If X is a second order stationary process with covariance function C, then X is intrinsic with semi-variogram $\gamma(h) = C(0) - C(h)$.*

Proposition 1.2.7 *The semi-variogram function γ of an intrinsic process X satisfies the following:*

1. *$\gamma(h) = \gamma(-h)$ (even function) and $\gamma(0) = 0$,*
2. *If A is a linear map on $\mathbb{R}^d$, then the function $h \mapsto \gamma(Ah)$ is also a semi-variogram function,*
3. *If γ is continuous at 0, then γ is continuous at every s where γ is locally bounded,*

4. *If γ is bounded in the neighborhood of 0, then there exists positive reals a and b such that, for all $x \in S$: $\gamma(x) \leq a||x||^2 + b$.*

Proposition 1.2.8 *The semi-variogram γ of an intrinsic process X is conditionally negative definite, i.e. for all $n \in \mathbb{N}^*$, for all $a \in \mathbb{R}^n$ such that $\sum_{i=1}^n a_i = 0$ and for all $(s_1, ..., s_n) \in S^n$, we have: $\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(s_i - s_j) \leq 0$.*

Theorem 1.2.1 *A function γ defined on $\mathbb{R}^d$ is a semi-variogram if, and only if, it is conditionally negative definite.*

Proposition 1.2.9 *Suppose X is an intrinsic process with bounded semi-variogram, such that $\lim_{||h|| \to +\infty} \gamma(h) = \gamma(+\infty) < +\infty$. Then X is second order stationary and $\gamma(+\infty) = C(0) = Var(X_s)$.*

We seek a spatial correlation that is a function of distance between spatial locations, so called isotropic spatial covariance function.

Definition 1.2.12 *The spherical covariance function between subjects i and j located at geographical location i and geographical location j is given by*

$$C_{sph}(d_{ij}) = \sigma^2 \left(1 - \frac{3}{2}\frac{|dij|}{a} + \frac{1}{2}\frac{|d_{ij}|^3}{a^3}\right)$$

where d_{ij} is the distance between the two locations, and a is the range.

The spherical covariance decreases until it disappears when range is reached. The parameter σ^2 is the maximum value of the covariance attained at the origin.

Definition 1.2.13 *The exponential covariance function between subjects i and j located at geographical location i and geographical location j is given by*

$$C_{exp}(d_{ij}) = \sigma^2 \exp\left(-\frac{|d_{ij}|}{a}\right) with \quad a > 0$$

where d_{ij} is the distance between the two locations.

As the distance between locations increases, the spatial covariance falls off exponentially. How quickly the covariance falls off is determined by the parameter a.

Definition 1.2.14 *Matérn family is a class of isotropic covariance functions which specifies the covariance function as* $\sigma^2 M(\boldsymbol{h}|\nu, a)$ *where* $\sigma^2 > 0$ *is the marginal variance and*

$$M(\boldsymbol{h}|\nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)}(a||\boldsymbol{h}||)^{\nu}\boldsymbol{K}_{\nu}(a||\boldsymbol{h}||)$$

is the spatial correlation at distance $||\boldsymbol{h}||$ *and* $\boldsymbol{h} \in \mathbb{R}^d$. *Here* $\boldsymbol{K}_\nu$ *is the modified Bessel function of the second kind and* $a > 0$ *is a spatial scale parameter, whose inverse,* $1/a$, *is sometimes referred to as a correlation length.*

Theorem 1.2.2 *A continuous function* γ *defined on* $\mathbb{R}^d$ *such that* $\gamma(0) = 0$ *is a semi-variogram if, and only if, for all a > 0, the function* $h \mapsto e^{-a\gamma(h)}$ *is a covariance function.*

Definition 1.2.15 *When the limit* $\lim_{||h||\to+\infty} \gamma(h) = \gamma(+\infty) < +\infty$ *exists, its value* $\gamma(+\infty)$ *is called sill.*

Definition 1.2.16 *The range (resp. practical range) is the distance where (resp. 95% of) the value of the sill is reached.*

Definition 1.2.17 *A semi-variogram has a nugget effect component when* $\lim_{||h||\to+0} \gamma(h) = \tau > 0$.

Graphical representation of sill, range and nugget effect can be found in Figure 1.1.

Definition 1.2.18 *The increasing domain asymptotic is a sampling structure in spatial statistics where new observations are added at the boundary points of an area.*

Definition 1.2.19 *The infill asymptotic consists of a sampling structure where new observations are added in between existing locations.*

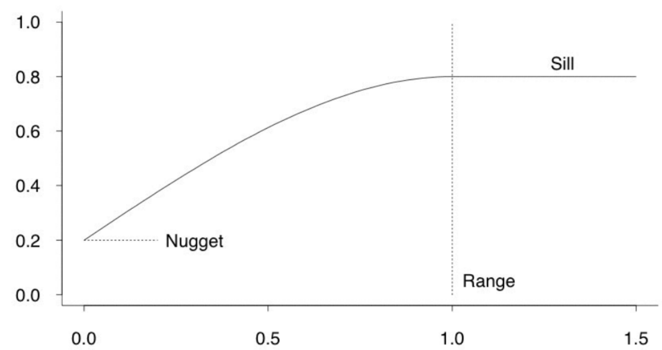

Figure 1.1. Variogram: nugget, sill and range.

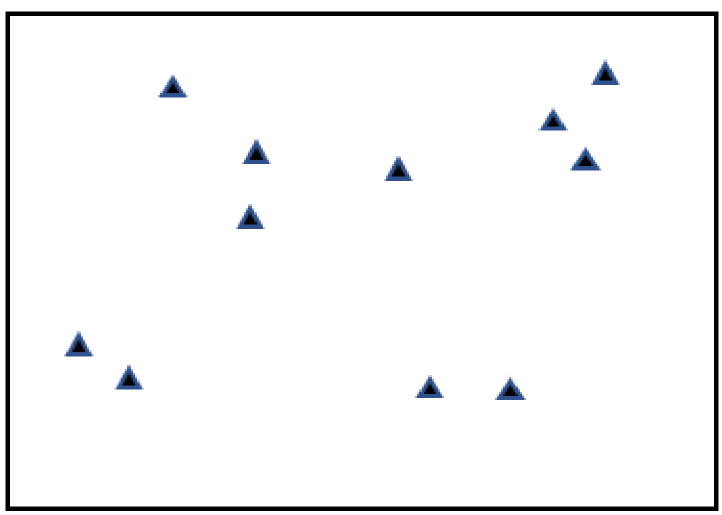

Figure 1.2. Spatial locations of individuals.

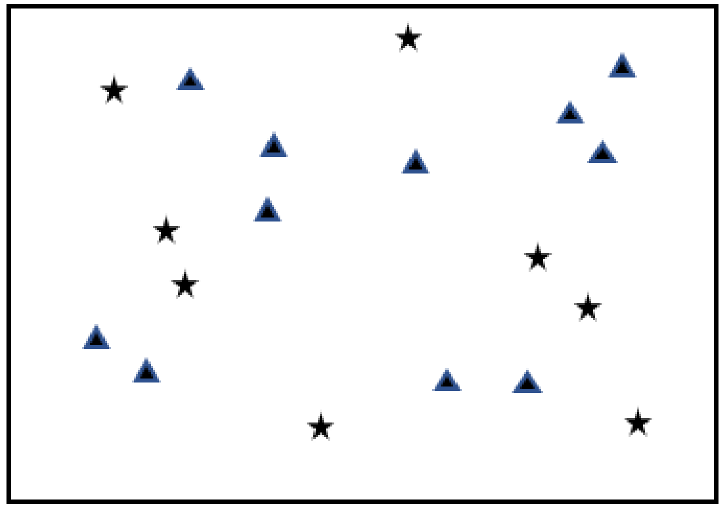

Figure 1.3. Infill asymptotic.

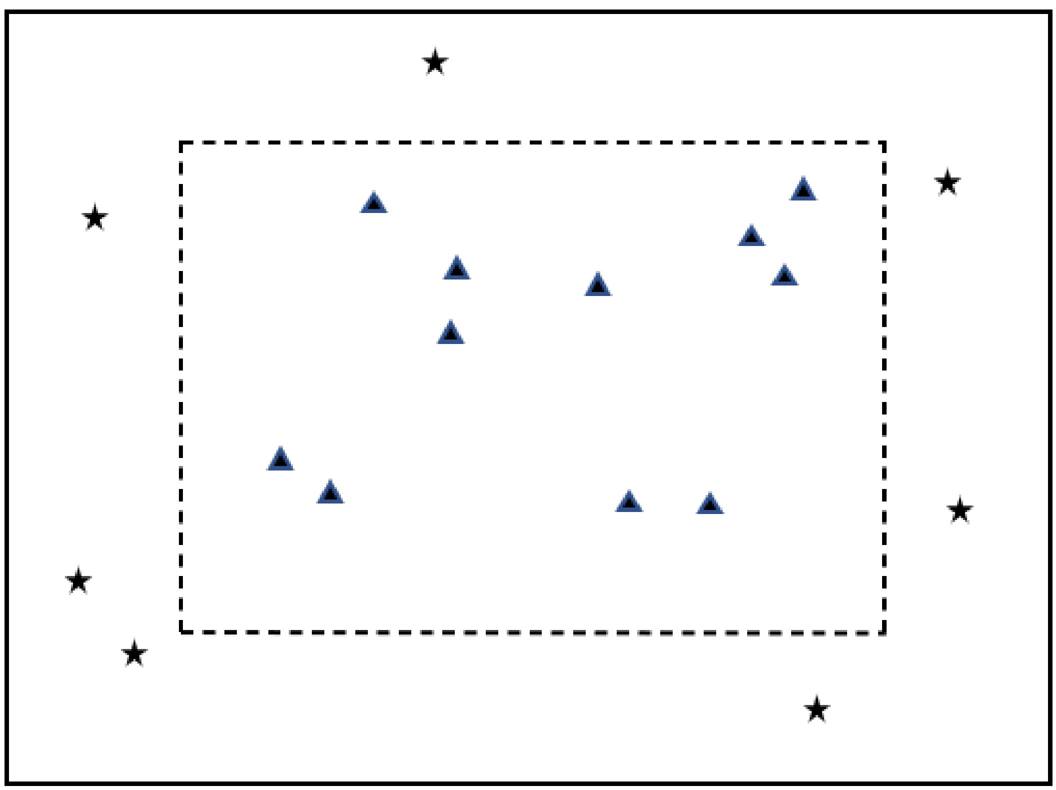

Figure 1.4. Increasing domain asymptotic

PAPER

I. MODELING SPATIALLY CLUSTERED RIGHT CENSORED DATA VIA MULTIVARIATE GAUSSIAN RANDOM FIELDS

Fathima Zahra Saunul Abdeen
Department of Mathematics & Statistics
Missouri University of Science and Technology
Rolla, Missouri 65409–0050
Email: fs2y7@mst.edu

Akim Adekpedjou
Department of Mathematics & Statistics
Missouri University of Science and Technology
Rolla, Missouri 65409–0050
Email: akima@mst.edu

S. Dabo-Niang
Laboratoire Paul Painvelé UMR CNRS 8524
INRIA-MODAL
University of Lille
Email: sophie.dabo@univ-lille.fr

ABSTRACT

Consider a fixed number of clustered areas identified by their geographical coordinates that are monitored for the occurrences of an event such as pandemic, epidemic, migration to name a few. Data collected on units at all areas include time varying covariates and environmental factors. The collected data is considered pairwise to account for spatial correlation between all pair of areas. The pairwise right censored data is probit-transformed, yielding a multivariate Gaussian random field preserving the spatial correlation function.

The data is analyzed using counting process and geostatistical formulation that led to a class of weighted pairwise semiparametric estimating functions. Estimators of models' unknowns are shown to be consistent and asymptotically normally distributed under infill-type spatial statistics asymptotic. Detailed small sample numerical studies that are in agreement with theoretical results are provided. The foregoing procedures are applied to Leukemia survival data in Northeast England.

Keywords: Spatial correlation; Gaussian random fields; Composite likelihood; Estimating function; Infill asymptotic; Mixing; Clustered failure times

1. INTRODUCTION

Right censored data are encountered in various settings such as biomedical, reliability, actuarial science, sociology, politics, and public health to name a few. They are part of a class of data called survival or failure time data which include, among others, the left and right censored, left and right truncation, and interval censored data. Research with these types of data is well documented. This dissertation pertains to another aspect of failure time data, namely one where spatial modeling is incorporated via geostatistical locations of units of interest. Consider the situation where these units, located at areas described by their longitude and latitude in a two dimensional surface are monitored for the occurrence of some event such as onset of disease, epidemic, claims filed as a result of property losses, cancer, or migration of individuals from one area to another to seek better living conditions. There exist nuisance parameters such as environmental factors, social and physical environments, population density, or weather conditions beyond the control of the investigators that can have substantial impact on the occurrence of events between two areas via their spatial coordinates. Two concrete examples of such data are given in biomedical studies. Many more can be found in the book by Goldstein (1995).

Example 1: *East Boston Asthma Study: cf. Li and Ryan (2002).*
A total of 753 subjects are enrolled in a Community Health Clinic in the east Boston area. Questionnaire data pertaining to residential addresses, demographic variables, asthma status, geographic coordinates, and other environmental factors were collected during regularly scheduled visits. Geocoding the dataset allows linkage with various community-level covariates to individuals in the east Boston area from U.S. census data at the census block level. Because children residing in nearby census blocks were often exposed to unmeasured similar physical and social environments, the investigators suspected there might exist spatial correlation across different communities. The goal of the study was to identify significant risk factors associated with age at asthma onset while accounting for the possible spatial correlation among the locations.

Example 2: *Leukemia Survival Data: Henderson et al. (2002) and Gorst (1995).*
1043 adults were diagnosed with leukemia between 1982 and 1998, in Northeast England, which is comprised of 24 administrative districts boxed in 100km^2. The data is a high-quality database that holds records of incidence and subsequent survival status of all leukemia cases in the region. Recorded also was the background variation in population or environmental characteristics, which could enable further epidemiological studies. Past studies, while informal, have suggested that there could be district-to-district variation in survival rates above and beyond what might be expected to occur by chance alone.

In the first example, residents of east Boston are mainly relatively low income with similar social and economical backgrounds who are often exposed to similar physical and social environments. One child per geographical area is considered in the modeling in Li and Lin (2006), whereas in Li and Ryan (2002) many units were considered per region. In both modeling approaches, spatial correlation was considered among geographical areas. The different geographical locations in east Boston are spatially correlated since adjacent neighborhoods usually have a lot in common and the potential for spatial dependence exists. Hence, correct inference on the association of the main covariates with the event-specific survival times relies on careful consideration of the underlying spatial correlation. In the Leukemia survival data, clustered survival data was considered and it is of interest when investigating how clustering and spatial aspects affect one region versus another, or how spatial traits can help with the identification of regions with high risk of leukemia. The Leukemia data fits more closely with the problem at hand here and can serve as a comparison between the two methodologies. In both cases, environmental factors in a given location may affect nearby locations thereby inducing the so-called *spatial correlation*, that is a correlation between the geographical locations of two units.

Modeling failure time data when spatial correlation is present has emerged as an area of active research, especially with right censored data. The models of interest are part of multivariate survival models that contain a parameter modeling the association between event times T_i and T_j, $i \neq j$ of two independent units. Such models include bivariate frailty, copulas, marginal models, cluster models, and spatially correlation-type models via the covariation process using a martingale representation. With right censored data, the references are of Li and Ryan (2002), Henderson et al. (2002), Banerjee et al. (2003), Banerjee and Dey (2005), Li and Lin (2006), Diva et al. (2007), Diva et al. (2008), Paik and Ying (2012), Pan et al. (2014), Hunt (1978), Bronnenberg (2005), Engen (2007) and Paik and Ying (2012). However, interest in spatial correlation dates back to the pioneering work of Krige, and recently Matheron (1962). Frailty, cluster, marginal, and copula models do not properly account for spatial correlation that is inherent with these data. As a consequence, sophisticated techniques of geostatistics, coupled with modern failure time data analysis are needed. In recognition of that, Li and Lin (2006), with right censored data assumed a Cox model for failure time and applied a probit-type transformation of the failure times yielding a multivariate Gaussian random field. Furthermore, they imposed a spatial structure on the associated random fields that properly captured the spatial patterns among regions.

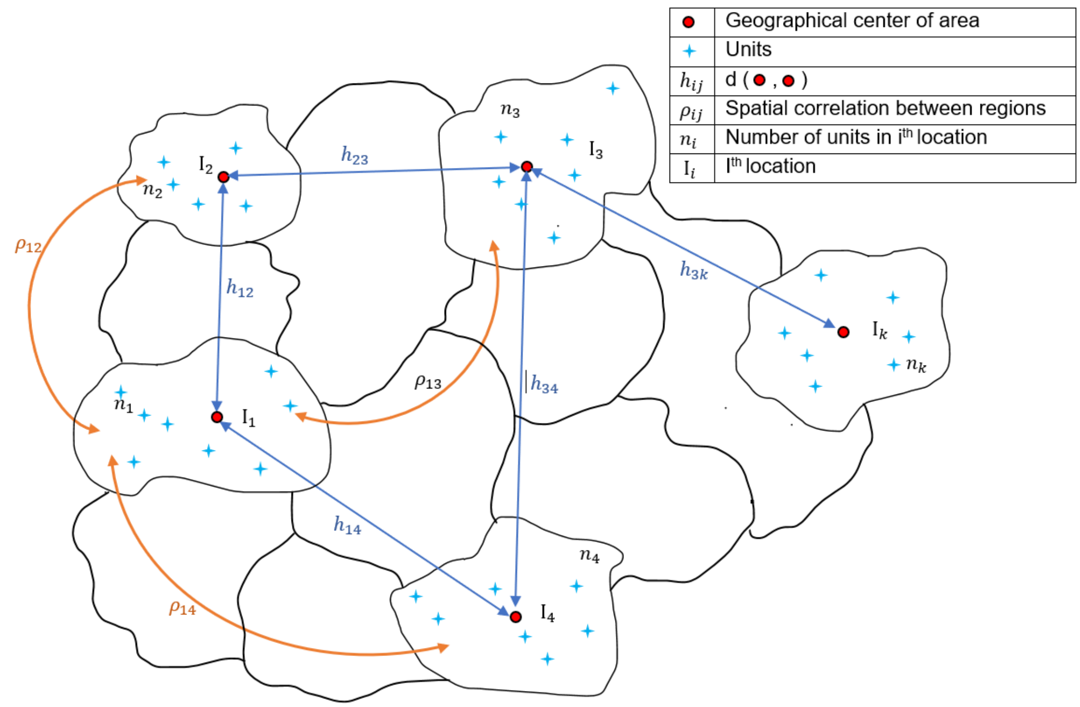

Figure 1. Pictorial representation of the setting

This dissertation is concerned with the development of models for estimating the regression parameters with clustered right censored data that account for spatial patterns between various locations. This is important in the sense that if the spatial impact leads to drastic consequences, local authorities could take necessary preventive actions to reduce damage. It is therefore of considerable importance to develop models for estimating the distribution function of time to event while accounting for spatial correlation. Multiple units per location are considered in order to reflect the real life situation and the Leukemia data will be used for illustration since it fits more closely with the setting with a pictorial representation given in Figure 1.

Henderson et al. (2002) modeled spatial association via a mean random frailty per region wherein individual frailty Z_i within a region j with mean frailty μ_j was assumed to follow a gamma distribution with parameters depending on μ_j. The vector $(\mu_1, ..., \mu_k)$ is assumed to follow a multivariate normal distribution whose variance-covariance matrix is a function of the distance between regions. This study incorporated spatial association in the failure times via a probit transformation leading to a multivariate Gaussian random field with the spatial correlation matrix being a function of the distance between locations. The two modeling approaches are applied to the same data and it is shown that our approach is preferable in terms of better statistical results. Henderson et al. (2002) did not provide large sample properties; this study provides all parameters involved in the models for the purpose of making inference and doing further investigations tailored to a specific area.

Though some work has been done on incorporating spatial correlation in modeling, very few of the works model many units per geostatistical location while accounting for spatial patterns. The aim of this dissertation is to develop statistical models for spatially correlated right censored data for multiple units per location where regression parameters have a region and/or area level interpretation and in which spatial correlation is properly incorporated. Ideas in the work of Li and Lin (2006) are borrowed by transforming the set of failure times using a probit-type function allowing the vector of right censored data times to follow a multivariate Gaussian random field (MGRF).

This part of the dissertation proceeds as follows. In Section 2, stochastic process machinery was developed for this type of data and our model choices was motivated. Section 3 deals with some preliminary results that will set the stage for the estimation procedures in Section 4. In Section 4, weighted estimating score processes were proposed and their asymptotically unbiasedness was shown. Section 5 is on the existence of solutions and the infill asymptotic results of the estimators. Section 6 presents the results of the numerical studies, which indicate good approximation to the true parameters, and an illustrative application with the Leukemia dataset. This part of the dissertation then concludes with a summary and future directions.

2. SPATIALLY CORRELATED RIGHT CENSORED DATA AND MODELS

The first critical step in the modeling is to identify a suitable dependence model between spatial locations. As noted earlier, a geostatistical formulation will be the focus, that relies on the fitting of covariance and cross-covariance structures for Gaussian random fields for mathematical and computational convenience. This approach also facilitates incorporation of the spatial correlation parameters in the modeling via the covariation process between two locations resulting from the martingale modeling.

To facilitate reading of the dissertation, the following notation on locations and number of units per location will be adopted throughout. There is a total of k locations with each being described by its longitude and latitude in a two dimensional coordinate with $\mathrm{I}_i = (l_{i1}, l_{i2})$. If no confusion arises, we will just write location i. The locations will be denoted by i and j, so that $i, j \in \{1, ..., k\} := \mathcal{L}$. Each location i has n_i units. Units are denoted by the letters r or s. For instance, in location i, we have $r = 1, ..., n_i$ so that $r \in \{1, ..., n_i\} := \mathcal{L}_i$. Likewise, $s \in \{1, ..., n_j\} = \mathcal{L}_j$. For convenience, the compact notation $(i, r) \in \mathcal{L} \times \mathcal{L}_i$ may be adapted similarly for (j, s).

2.1. PAIRWISE RIGHT CENSORED DATA

Consider k geographical locations described by two dimensional coordinates $\{\mathrm{I}_i = (l_{i1}, l_{i2}); i = 1, ..., k\}$ where l_{i1} and l_{i2} denote longitude and latitude of the i^{th} geographical location respectively. Let n_i be the number of subjects in the i^{th} geographical location. Each unit is observed until failure or censoring, whichever occurs first. At time t, for the r^{th} $(r = 1, 2, ..., n_i)$ unit in the i^{th} $(i = 1, 2, ..., k)$ geographical location, failure or censoring time is recorded by $W_i^{(r)}$ and $C_i^{(r)}$ respectively. Let $\delta_i^{(r)} = \mathrm{I}(W_i^{(r)} \leq C_i^{(r)})$, $T_i^{(r)} = W_i^{(r)} \wedge C_i^{(r)}$ be the usual notation with right censored data. The variable $\delta_i^{(r)}$ indicates that either censoring or failure has occurred for unit r in location i. For $(i, r) \in \mathcal{L} \times \mathcal{L}_i$, a p-dimensional vector $\mathbf{x}_i^{(r)}(t)$ of possibly time varying covariates is recorded at time

t. Location i is assumed to be spatially correlated with j, $i \neq j$ and denote the spatial correlation between the two by $\rho_{ij} := \rho(\|\mathbf{l}_i - \mathbf{l}_j\|)$, where $\|\mathbf{l}_i - \mathbf{l}_j\|$ is the Euclidean distance between $\mathbf{l}_i$ and $\mathbf{l}_j$. The total observables entities per location at time t are therefore,

$$\mathbf{O}(\mathrm{I}_i) = \cup_{r=1}^{n_i} \mathbf{O}^{(r)}(\mathrm{I}_i) = \cup_{r=1}^{n_i} \{\mathbf{x}_i^{(r)}(t), T_i^{(r)}, \delta_i^{(r)}\}. \tag{1}$$

In the present setting of spatially correlated events, the random observables in (1) will be taken pairwise for the purpose of accounting for the spatial correlation ρ_{ij}. Consequently, the spatially correlated right censored data on $\mathcal{L}$, on which estimation is conducted is given by

$$\boldsymbol{O} = \left\{ \left[\left(\mathbf{O}(\mathrm{I}_i), \mathbf{O}(\mathrm{I}_j) \right) ; \rho_{ij} \right] ; i \neq j ; (i, j) \in \{1, ..., k\}^2 = \mathcal{L} \times \mathcal{L} \right\}. \tag{2}$$

2.2. STOCHASTIC PROCESS MODELING

With a view towards the multivariate Gaussian random field (MGRF), the stochastic processes needed in the sequel are introduced. For $(i, r) \in \mathcal{L} \times \mathcal{L}_i$, define the counting and at-risk process by $\mathrm{N}_i^{(r)}(t) = \delta_i^{(r)} \mathrm{I}(T_i^{(r)} \leq t)$ and $\mathrm{Y}_i^{(r)}(t) = \mathrm{I}(T_i^{(r)} \geq t)$ respectively. Note that $\mathrm{N}_i^{(r)}(t)$ indicates if an event has occurred by time t, whereas $\mathrm{Y}_i^{(r)}(t)$ indicates if unit (i, r) is at risk at time t. $\mathrm{Y}_i^{(r)}(\cdot)$ may be modified to allow left truncation or other general at-risk processes. It is further assumed that the study ends at a time τ with $\tau \geq \max_{r,i} \mathrm{T}_i^{(r)}$. So that the interval $[0, \tau] = \mathcal{T}$ is the observation time zone. The entire history at all geostatistical locations at the end of the study is contained in the σ-field $\mathcal{F} = \bigvee_{i=1}^{k} \bigvee_{r=1}^{n_i} \mathcal{F}_{i,\tau}^{(r)}$ with

$$\mathcal{F}_{i,\tau}^{(r)} = \sigma \left(N_i^{(r)}(t), Y_i^{(r)}(t), t \in \mathcal{T} \right).$$

To proceed with the modeling, it is assumed that the instantaneous hazard function is different from location to location. If $\lambda_i(t)$ is the instantaneous failure rate in $(t, t + dt)$ for all units in location i, from stochastic integration theory, the compensator process of

$N_i^{(r)}(t)$ is $A_i^{(r)}(t)$ given by $A_i^{(r)}(t) = \int_0^t Y_i^{(r)}(u)\lambda_i(u)du$ so that for each (i,r), the process

$$\left\{ M_i^{(r)}(t) = N_i^{(r)}(t) - \int_0^t Y_i^{(r)}(u)\lambda_i^{(r)}(u)du : t \in \mathcal{T} \right\}$$

is a zero-mean square-integrable martingale with respect to the filtration $\mathcal{F}_{i,t}^{(r)}$.

The choice of $\lambda_i(t)$ is crucial in obtaining the MGRF. Many choices are possible, such as additive, multiplicative, additive-multiplicative, or accelerated hazard-type models. The Cox model was chosen because it is easier to apply a logit transformation in the pursuit of the MGRF. However, an accelerated model can also be used, but a rank-based estimation approach would have to be employed to estimate the unknown in the model.

To that end, for $(i,r) \in \mathcal{L} \times \mathcal{L}_i$, let $\mathbf{x}_i^{(r)}(t)$ be a p-dimensional vector of covariates. The Cox model is postulated with different baseline per location, but the same regression parameter $\boldsymbol{\beta}$ for all locations given by

$$\lambda_i^{(r)}(t) = \lambda_{0i}(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(t)),$$

where $\mathbf{a}'$ denotes the transpose of the vector $\mathbf{a}$, and $\boldsymbol{\beta}$ is a p-dimensional vector of regression parameters. The baseline hazard per location is $\lambda_{0i}(t)$ and $\{\lambda_{0i}(t) : i = 1, ..., k\}$ is the set of unspecified baseline hazard functions to be estimated.

Remark 1 *The choice of same $\boldsymbol{\beta}$ coefficient for all locations is motivated by modeling the same event for all units in all locations. However, the baseline hazard is chosen to be different among the locations (see also Spiekerman and Lin (1998) and Lin (1994)). The case of competing failures can be considered also, cf. Wei et al. (1989).*

2.3. MULTIVARIATE GAUSSIAN RANDOM FIELD

Gaussian Random Fields (GRF) and their multivariate counterpart MGRF play a dominant role in spatial modeling, especially in geostatistics. Estimation of parameters are facilitated if the models proposed can lead to the construction of MGRF. Since counting processes and martingales have been the cornerstone of modeling failure time data via the hazard function, it turns out that, making the event times normally distributed will lead to the construction of MGRF. The motivation behind this approach is threefold: (i) the marginal distribution of the event times follow a model that accounts for covariates, (ii) prediction of event occurrences at a new location is faster with GRF using existing software packages and kriging techniques, and (iii) the approach facilitates construction of pairwise composite likelihood process, estimation of parameters, as well as large sample properties via estimating functions. With a view towards the MGRF construction, for $(i, r) \in \mathcal{L} \times \mathcal{L}_i$, let $\mathbf{x}_i^{(r)} := \mathbf{x}_i^{(r)}(t)$ if ambiguity does not arise, $\Lambda(t|\mathbf{x}_i^{(r)})$ is the cumulative hazard function, and $\bar{F}_i^{(r)}(t|\mathbf{x}_i^{(r)}) = \exp\left[-\Lambda(t|\mathbf{x}_i^{(r)})\right]$ the survivor function. Then $\bar{F}_i^{(r)}(T_i^{(r)}|\mathbf{x}_i^{(r)})$ follows a uniform distribution on $(0, 1)$ and $\Lambda_i^{(r)}(T_i^{(r)}|\mathbf{x}_i^{(r)})$ follows a unit exponential distribution $\mathrm{EXP}(1)$. Those facts are well known. If $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, the probit transformation of a variable U in $(0, 1)$ is $\Phi^{-1}(U)$. Hence,

$$\tilde{\mathrm{T}}_i^{(r)} := \Phi^{-1}\left[1 - e^{-\Lambda_{0i}(T_i^{(r)})\exp(\boldsymbol{\beta}'\mathbf{x}_i^{(r)})}\right]$$

is the probit transformation of the failure time $T_i^{(r)}$, which follows a standard normal distribution $\mathrm{N}(0, 1)$. Note here that for each location i, its vector of failure times

$$\tilde{\mathbf{T}}_i = (\tilde{T}_i^{(1)}, \tilde{T}_i^{(2)}, ..., \tilde{T}_i^{(n_i)})$$

form an n_i multivariate normal distribution. Consequently, a MGRF can be constructed with $\tilde{\mathbf{T}}$ given by

$$\tilde{\mathbf{T}} = \left(\tilde{\mathbf{T}}_1, \tilde{\mathbf{T}}_2, ..., \tilde{\mathbf{T}}_k\right)_{(n_1,...,n_k)},$$

by imposing a spatial structure induced by a $(\sum_{i=1}^{k} n_i)\times(\sum_{i=1}^{k} n_i)$ spatial correlation matrix $\mathbf{\Xi}$ with block matrices $\mathbf{J}_{n_i\times n_i} = (1)_{n_i\times n_i}$, $i = 1, ..., k$ as diagonal elements, and the off diagonal elements $(n_i, n_j)\mathrm{I}\{i \neq j\}$ depends on the spatial correlation ρ_{ij} between two locations. The matrix $\mathbf{\Xi}$ takes the form

$$\mathbf{\Xi} = \begin{bmatrix} \sigma_{11}^2\mathbf{J}_{n_1\times n_1} & \rho_{12}\mathbf{J}_{n_1\times n_2} & \cdots & \rho_{1j}\mathbf{J}_{n_1\times n_j} & \cdots & \rho_{1k}\mathbf{J}_{n_1\times n_k} \\ \rho_{21}\mathbf{J}_{n_2\times n_1} & \sigma_{22}^2\mathbf{J}_{n_2\times n_2} & \cdots & \rho_{2j}\mathbf{J}_{n_2\times n_j} & \cdots & \rho_{2k}\mathbf{J}_{n_2\times n_k} \\ \vdots & \vdots & \vdots & \vdots & & \\ \rho_{i1}\mathbf{J}_{n_i\times n_1} & \rho_{i2}\mathbf{J}_{n_2\times n_2} & \cdots & \rho_{ij}\mathbf{J}_{n_i\times n_j} & \cdots & \rho_{ik}\mathbf{J}_{n_i\times n_k} \\ \vdots & \vdots & \vdots & \vdots & & \\ \rho_{k1}\mathbf{J}_{n_k\times n_1} & \rho_{k2}\mathbf{J}_{n_k\times n_2} & \cdots & \rho_{kj}\mathbf{J}_{n_k\times n_j} & \cdots & \sigma_{kk}^2\mathbf{J}_{n_k\times n_k} \end{bmatrix}_{(\sum_{i=1}^{k} n_i)\times(\sum_{i=1}^{k} n_i)}$$

In the above matrix, the diagonal elements are the variance covariance matrices of the failure times within a given region. Since in this case, $\|h\| = 0$, those elements reduce to the marginal variances $\sigma_{ii}^2\mathrm{I}\{ii \in \{11, 22, ..., kk\}\}$ in the Matérn spatial correlation function that is considered here, and will be introduced in the next subsection.

2.4. CHOICE OF SPATIAL CORRELATION MODEL

As indicated earlier, the critical part in identifying significant risk factors that trigger event occurrences is to identify the best spatial correlation function. Henderson et al. (2002) proposed a multivariate gamma frailty model incorporating spatial dependence between locations as was done in Banerjee et al. (2003). Diva et al. (2007) extended the work of Banerjee and Carlin (2003) by generalizing their Multivariate Conditional Autoregressive Models. A pairwise joint distribution that depends on the distance between locations has been investigated by Paik and Ying (2012). Copula models on the other hand have been proposed by Lawless and Yilmaz (2011) and Yilmaz and Lawless (2011). The problem with the use of frailty or copula is that the former models within cluster correlation using frailties or random effects, and the latter models joint distribution of two failure times, and consequently do not really model spatial correlation. This study seeks spatial correlation that is a function of distance between spatial locations, so called *isotropic spatial covariance* functions. They have received a great deal of attention recently, specifically the Matérn family Matérn (1986); Guttorp and Gneiting (2006); Gneiting et al. (2010) given by

$$C(\|\mathbf{h}\|) = \sigma^2 M(\mathbf{h}; \nu, a) = \sigma^2 \left(\frac{2^{1-\nu}}{\Gamma(\nu)} (a\|\mathbf{h}\|)^{\nu} K_{\nu} (a\|\mathbf{h}\|) \right), \tag{3}$$

where σ^2 is the marginal variance or *sill*, that is the variance if $\|\mathbf{h}\| = \|\mathbf{l}_i - \mathbf{l}_j\| = 0$. $\nu > 0$ is a smoothing parameter that controls the differentiability of a Gaussian process with this covariance; and $a > 0$ is a *range* parameter that measures the correlation decay as the separation between two locations increases. $K_\nu(\cdot)$ and $\Gamma(\cdot)$ are the Bessel and gamma functions respectively. When $\nu = 0.5$ and $+\infty$, the exponential and Gaussian covariance are recovered and given by

$$C(\mathbf{h}) = \sigma^2 \exp(-a\|\mathbf{h}\|), \tag{4}$$

$$C(\mathbf{h}) = \sigma^2 \exp(-a^2\|\mathbf{h}\|^2), \tag{5}$$

respectively. More details on sill and range can be found in Section 3 of Handcock and Stein (1993) or Section 1 of Gneiting et al. (2010). The Matérn family turns out to be a good choice because of its flexibility in modeling various types of spatial correlation structure in many fields and possesses a good interpretability of the parameters. The importance of this family is also highlighted in Stein (1999), page 14. Note that in (3), if $i = j$, $\|\mathbf{h}\| = 0$ we get $C(\mathbf{h}) = \sigma^2$ the marginal variance which corresponds to the case of no spatial correlation.

In what follows, it is assumed that the spatial correlation function depends on the q-dimensional parameter $\boldsymbol{\delta} = (\delta_1, ..., \delta_q)$ each describing various elements of the family. A Matérn-type family for spatial correlation on the transformed failure times is assumed, translating into $\boldsymbol{\delta} = (\sigma^2, \text{range}, \text{sill})$, that is $q = 3$. The transformation leads to a MGRF where the marginal failure times follow the postulated Cox model with a population level interpretation for the regression parameter $\boldsymbol{\beta}$, and facilitates estimation of the spatial as well as regression parameters. Thus

$$\boldsymbol{\Xi} = \left[\boldsymbol{\Xi}_{ij}\right]_{i,j=1,...,k} = \left[\mathrm{Cov}(\tilde{\mathbf{T}}_i, \tilde{\mathbf{T}}_j)\right]_{i,j=1,...,k} = (\rho_{ij}(\boldsymbol{\delta}))_{i,j=1,...,k}.$$

For compactness, the notation $\rho_{ij}(\boldsymbol{\delta})$ will be used for $\rho(\mathbf{l}_i, \mathbf{l}_j; \boldsymbol{\delta})$.

3. ESTIMATION-PRELIMINARY

The unknowns arise from two models, the spatial correlation and the Cox models. The Cox model with its unknown infinite dimensional baseline parameters $\lambda_{0i}(t), i = 1, ..., k$ belongs to a class C of hazards on $\Re^+$. The regression coefficient $\boldsymbol{\beta}$ is in $\Re^p$, whereas the q-dimensional Matérn spatial correlation $\boldsymbol{\delta}$ is in $\Re^q$. For the Matérn family, $q = 3$, the theory for an unknown q is found. So, the model parameter of main interest is

$$\boldsymbol{\theta} = [(\lambda_{01}(t), ..., \lambda_{0k}(t)); (\beta_1, ..., \beta_p); (\delta_1, ..., \delta_q)] \in \boldsymbol{\Theta},$$

where $\boldsymbol{\Theta} \subset C^k \times \Re^p \times \Re^q_+$. The observables $\boldsymbol{O}$ in (2) will be used for making inference on $\boldsymbol{\theta}$.

Remark 2 *These models have $(k + p + q)$ unknowns, which raises the question of identifiability. Let $p_{\boldsymbol{\theta}}(\cdot)$ be the probability model on $\boldsymbol{O}$. The issues of identifiability will not arise, that is the Kullback-Leibler information will be positive for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ under the assumptions that, under $p_{\boldsymbol{\theta}}(\cdot)$: (i) no two regions have the same longitude and latitude; (ii) for every* *region i, $\sum_{r=1}^{n_i} Y_i^{(r)}(t) > 0$, that is at least one failure occurs per region; (iii) for every i, $n_i \geq 2$ and (iv) for $A \in \mathcal{G}_i = \sigma\left(\sum_{r=1}^{n_i} Y_i^{(r)}(t), t \in \mathcal{T}\right)$, and $B \in \mathcal{G}_j$ (likewise defined), $P(A \cap B) > 0$. The last assumption ensures estimation of the spatial correlation parameter, hence a uniquely defined spatial correlation function.*

3.1. AALEN-BRESLOW ESTIMATOR OF $\lambda_{0i}(t)$ AND ITS PROPERTIES

Following the notation in Section 2.3, and as indicated earlier, for each (i, r),

$$\left\{M_i^{(r)}(t) = N_i^{(r)}(t) - \int_0^t Y_i^{(r)}(u)\lambda_i^{(r)}(u)du : t \in \mathcal{T}\right\}$$

is a zero-mean martingale with respect to the filtration $\mathcal{F}_{i,t}^{(r)}$. It then follows, via method of moments, that an Aalen-Breslow estimator for $\Lambda_{0i}(\cdot) = \int_0^{\cdot} \lambda_{0i}(u)du$, for $i \in \mathcal{L}$ is given by

$$\hat{\Lambda}_{0i}(t) = \int_0^t \frac{\sum_{r=1}^{n_i} dN_i^{(r)}(u)}{\sum_{r=1}^{n_i} Y_i^{(r)}(u)\exp\left(\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)\right)}, \tag{6}$$

with the k dimensional vector of baseline hazard being $\hat{\boldsymbol{\Lambda}}_0(t) = \left(\hat{\Lambda}_{01}(t), ..., \hat{\Lambda}_{0k}(t)\right)$. Observe that $\hat{\boldsymbol{\Lambda}}_0(t)$ is not yet an estimator because it still depends on the unknown regression parameter $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$. The expression in (6) will later be substituted for $\lambda_{0i}(t)$ to estimate $\boldsymbol{\beta}$ and to obtain the in-probability limits of the score matrix.

In order to facilitate understanding of the asymptotic properties of the parameters in these models, it is important to go through some properties of $\hat{\boldsymbol{\Lambda}}_0(t)$, in particular $\hat{\lambda}_{0i}(t)$, properly standardized. The $\hat{\lambda}_{0i}(t)$ in the next theorem is one where $\boldsymbol{\beta}$ is replaced by its estimator $\hat{\boldsymbol{\beta}}$ and pertains to the consistency of $\hat{\lambda}_{0i}(t)$ as $n_i \to \infty$, for each $i \in \mathcal{L}$. As will be discussed later in the large sample properties section, the requirement $\wedge_{i=1}^{n} n_i \to \infty$ suffices to satisfy the infill asymptotic property under which the large sample properties are obtained. This consistency result will be needed when showing the consistency of $\hat{\boldsymbol{\beta}}$, which requires the existence of an in-probability limit for the score variance. The following theorem can be shown easily using various stochastic processes arguments (cf. Wei et al. (1989)).

Theorem 1 *For $i = 1, ..., k$ and as $n_i \to \infty$, the estimator $\hat{\lambda}_{0i}(t)$ is consistent for $\lambda_{0i}(t)$, that is*

$$\sup_{t\in[0,\tau]} |\hat{\lambda}_{0i}(t) - \lambda_{0i}(t)| \xrightarrow{p} 0.$$

Proof: It suffices to show that, for $i = 1, ..., k$ and as $n_i \to \infty$,

$$\sup_{t\in[0,\tau]} \left|\hat{\Lambda}_{0i}(t) - \Lambda_{0i}(t)\right| \xrightarrow{p} 0.$$

By triangular inequality we obtain

$$\begin{aligned} \sup_{t\in[0,\tau]} \left|\hat{\Lambda}_{0i}(t|\hat{\boldsymbol{\beta}}) - \Lambda_{0i}(t)\right| \leq & \sup_{t\in[0,\tau]} \left|\hat{\Lambda}_{0i}(t|\hat{\boldsymbol{\beta}}) - \hat{\Lambda}_{0i}(t|\boldsymbol{\beta}_0)\right| + \sup_{t\in[0,\tau]} \left|\hat{\Lambda}_{0i}(t|\boldsymbol{\beta}_0) - \Lambda^*_{0i}(t)\right| \\ & + \sup_{t\in[0,\tau]} \left|\Lambda^*_{0i}(t) - \Lambda_{0i}(t)\right| \end{aligned} \tag{7}$$

where

$$\Lambda^*_{0i}(t) = \int_0^t \lambda_{0i}(u) I\left\{\sum_{r=1}^{n_i} Y_i^{(r)}(u) > 0\right\} du.$$

It suffices to show that each term in the right-hand side of (7) is asymptotically negligible. By Taylor expansion of first term of (7) around $\boldsymbol{\beta}_0$ we obtain

$$\begin{aligned} \frac{\left|\hat{\Lambda}_{0i}(t|\hat{\boldsymbol{\beta}}) - \hat{\Lambda}_{0i}(t|\boldsymbol{\beta}_0)\right|}{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0} &= \hat{\Lambda}'_{0i}(t|\boldsymbol{\beta}^*) \\ \left|\hat{\Lambda}_{0i}(t|\hat{\boldsymbol{\beta}}) - \hat{\Lambda}_{0i}(t|\boldsymbol{\beta}_0)\right| &= \left|-(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \int_0^t \frac{S_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)^{\otimes 2}} \cdot \sum_{r=1}^{n_i} dN_i^{(r)}(u)\right| \\ &= \left|-(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \int_0^t \frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} \cdot \sum_{r=1}^{n_i} dN_i^{(r)}(u)\right| \end{aligned} \tag{8}$$

where $\boldsymbol{\beta}^* \in (\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0)$ such that $\boldsymbol{\beta}^* \xrightarrow{p} \boldsymbol{\beta}_0$. Now we consider the term $\int_0^t \frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} \cdot \sum_{r=1}^{n_i} dN_i^{(r)}(u)$ of (8)

$$\begin{aligned}
\int_0^t \frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} \cdot \sum_{r=1}^{n_i} dN_i^{(r)}(u) &= \int_0^t \frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} \left[\sum_{r=1}^{n_i} dM_i^{(r)}(u) \right. \\
&\quad \left. + \sum_{r=1}^{n_i} Y_i^{(r)}(u) \lambda_i^{(r)}(u) du \right] \\
&= \int_0^t \frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} \cdot \sum_{r=1}^{n_i} dM_i^{(r)}(u) \\
&\quad + \int_0^t E_i^{(1)}(u,\boldsymbol{\beta}^*) \lambda_{0i}(u) du. \qquad (9)
\end{aligned}$$

By regularity condition III the second term on the right side of (9) converges to finite value of $\int_0^t e_i^{(1)}(u,\boldsymbol{\beta}^*)\lambda_{0i}(u)du$. For the first term on the right side of (9) we observe that

$$\begin{aligned}
\left| \int_0^t \frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} \cdot \sum_{r=1}^{n_i} dM_i^{(r)}(u) \right| &\le \left| \int_0^t \left[\frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} - \frac{e_i^{(1)}(u,\boldsymbol{\beta}^*)}{s_i^{(0)}(u,\boldsymbol{\beta}^*)} \right] \sum_{r=1}^{n_i} dM_i^{(r)}(u) \right| \\
&\quad + \left| \int_0^t \frac{e_i^{(1)}(u,\boldsymbol{\beta}^*)}{s_i^{(0)}(u,\boldsymbol{\beta}^*)} \sum_{r=1}^{n_i} dM_i^{(r)}(u) \right|. \qquad (10)
\end{aligned}$$

By regularity condition III and Andersen and Gill (1982) theorem I.2 the two terms on the right side of (10) are asymptotically negligible. Therefore

$$\sup_{t\in[0,\tau]} \left| \int_0^t \frac{E_i^{(1)}(u,\boldsymbol{\beta}^*)}{S_i^{(0)}(u,\boldsymbol{\beta}^*)} \cdot \sum_{r=1}^{n_i} dM_i^{(r)}(u) \right| = o_p(1).$$

By regularity condition III we obtain

$$\int_0^t E_i^{(1)}(u,\boldsymbol{\beta}^*)\lambda_{0i}(u)du \xrightarrow{p} \int_0^t e_i^{(1)}(u,\boldsymbol{\beta}^*)\lambda_{0i}(u)du$$

and by consistency of $\hat{\boldsymbol{\beta}}$, the first term of (7) is negligible. The second term of (7) turns out to be a local martingale namely

$$
\begin{aligned}
\hat{\Lambda}_{0i}(t|\boldsymbol{\beta}_0) - \Lambda^*_{0i}(t) &= \int_0^t \frac{\sum_{r=1}^{n_i} dN_i^{(r)}(u)}{\sum_{r=1}^{n_i} Y_i^{(r)}(u)\exp\left(\boldsymbol{\beta}'\mathbf{X}_i^{(r)}(u)\right)} \\
&\quad - \int_0^t \lambda_{0i}(u) I\left\{\sum_{r=1}^{n_i} Y_i^{(r)}(u) > 0\right\} du \\
&= \int_0^t \frac{1}{\sum_{r=1}^{n_i} Y_i^{(r)}(u)\exp\left(\boldsymbol{\beta}'\mathbf{X}_i^{(r)}(u)\right)} \\
&\quad \times \left[\sum_{r=1}^{n_i} dN_i^{(r)}(u) - \sum_{r=1}^{n_i} Y_i^{(r)}(u)\lambda_{0i}(u)\exp\left(\boldsymbol{\beta}'\mathbf{X}_i^{(r)}(u)\right) du\right] \\
&= \int_0^t \frac{1}{\sum_{r=1}^{n_i} Y_i^{(r)}(u)\exp\left(\boldsymbol{\beta}'\mathbf{X}_i^{(r)}(u)\right)} \\
&\quad \times \left[\sum_{r=1}^{n_i} dN_i^{(r)}(u) - \sum_{r=1}^{n_i} Y_i^{(r)}(u)\lambda_i^{(r)}(u)\right] \\
&= \frac{\sum_{r=1}^{n_i} dM_i^{(r)}(u)}{\sum_{r=1}^{n_i} Y_i^{(r)}(u)\exp\left(\boldsymbol{\beta}'\mathbf{X}_i^{(r)}(u)\right)}.
\end{aligned}
$$

Now by triangular inequality

$$
\begin{aligned}
\left|\hat{\Lambda}_{0i}(t|\boldsymbol{\beta}_0) - \Lambda^*_{0i}(t)\right| &\leq \left|\frac{1}{n}\int_0^t \frac{ndM_i^{(r)}(u)}{S_i^{(0)}(\boldsymbol{\beta}_0, u)} - \int_0^t \frac{dM_i^{(r)}(u)}{s_i^{(0)}(\boldsymbol{\beta}_0, u)}\right| \\
&\quad + \left|\frac{1}{n}\int_0^t \frac{dM_i^{(r)}(u)}{s_i^{(0)}(\boldsymbol{\beta}_0, u)}\right| \\
&\leq \left|\frac{1}{n}\int_0^t \left[\frac{n}{S_i^{(0)}(\boldsymbol{\beta}_0, u)} - \frac{1}{s_i^{(0)}(\boldsymbol{\beta}_0, u)}\right] dM_i^{(r)}(u)\right| \\
&\quad + \left|\frac{1}{n}\int_0^t \frac{dM_i^{(r)}(u)}{s_i^{(0)}(\boldsymbol{\beta}_0, u)}\right| \qquad (11)
\end{aligned}
$$

Considering the first term of (11) we obtain

$$\begin{aligned}
\sup_{t\in[0,\tau]}\left|\frac{1}{n}\int_0^t\left[\frac{n}{S_i^{(0)}(\boldsymbol{\beta}_0,u)}-\frac{1}{s_i^{(0)}(\boldsymbol{\beta}_0,u)}\right]dM_i^{(r)}(u)\right| \quad\leq\quad & \left\{\sup_{t\in[0,\tau]}\left[\frac{n}{S_i^{(0)}(\boldsymbol{\beta}_0,t)}\right.\right.\\
& \left.\left.-\ \frac{1}{s_i^{(0)}(\boldsymbol{\beta}_0,t)}\right]\right\}\times\left\{\frac{N_i^{(r)}(\tau)}{n}\right.\\
& \left.+\ \frac{1}{n}\int_0^t Y_i^{(r)}(u)\lambda_i^{(r)}(u)du\right\}.
\end{aligned}$$

By regularity condition III and VII

$$\sup_{t\in[0,\tau]}\left|\frac{1}{n}\int_0^t\left[\frac{n}{S_i^{(0)}(\boldsymbol{\beta}_0,u)}-\frac{1}{s_i^{(0)}(\boldsymbol{\beta}_0,u)}\right]dM_i^{(r)}(u)\right|=o_p(1).$$

Now the second term on the right hand side of (11) converges weakly to a zero mean Gaussian process by Andersen and Gill (1982) theorem I.2. Hence it is negligible and as a consequence

$$\sup_{t\in[0,\tau]}\left|\hat{\Lambda}_{0i}(t|\boldsymbol{\beta}_0)-\Lambda_{0i}^*(t)\right|=o_p(1).$$

Using Markov's inequality, for any $\epsilon\in(0,1)$, for the last term on the right hand side of (7),

$$\begin{aligned}
P\left\{\sup_{t\in[0,\tau]}\left|\Lambda_{0i}^*(t)-\Lambda_{0i}(t)\right|>\epsilon\right\} \quad&\leq\quad E\left\{\sup_{t\in[0,\tau]}\left|\Lambda_{0i}^*(t)-\Lambda_{0i}(t)\right|\right\}/\epsilon\\
&\leq\quad E\left\{\sup_{t\in[0,\tau]}\left|\int_0^t\lambda_{0i}(u)I\left\{\sum_{r=1}^{n_i}Y_i^{(r)}(u)>0\right\}du\right.\right.\\
&\qquad\left.\left.-\int_0^t\lambda_{0i}(u)du\right|\right\}/\epsilon\\
&\leq\quad E\left\{\sup_{t\in[0,\tau]}\int_0^t I\left\{\sum_{r=1}^{n_i}Y_i^{(r)}(u)=0\right\}\lambda_{0i}(u)du\right\}/\epsilon\\
&\leq\quad \int_0^t P\left(\sum_{r=1}^{n_i}Y_i^{(r)}(u)=0\right)\lambda_{0i}(u)du/\epsilon\\
&\leq\quad \int_0^t\left[P\left(Y_i^{(1)}(u)\right)=0\right]^{n_i}\lambda_{0i}(u)du/\epsilon.
\end{aligned}$$

By the dominated convergence theorem, since $P\left(Y_i^{(1)}(u)\right) = 0$ is bounded. Then

$$P\left\{\sup_{t\in[0,\tau]}\left|\Lambda_{0i}^*(t) - \Lambda_{0i}(t)\right| > \epsilon\right\} \leq \left\{P\left(Y_i^{(1)}(t^*) = 0\right)\right\}^{n_i}\Lambda_{0i}(t^*)/\epsilon$$

Since $P\left(Y_i^{(1)}(t^*) = 0\right) < 1$ and $\Lambda_{0i}(t^*) < \infty$ as $n_i \to \infty$ it follows that

$$\sup_{t\in[0,\tau]}\left|\Lambda_{0i}^*(t) - \Lambda_{0i}(t)\right| = o_p(1).$$

This completes the proof.

Remark 3 *An important result worth pointing out in connection with Theorem 1 is the convergence under the infill asymptotic of the random field* $\widehat{\boldsymbol{W}}(t)$ *given by*

$$\widehat{\boldsymbol{W}}(t) = \sqrt{n}\left(\hat{\Lambda}_{01}(t) - \Lambda_{01}(t), ..., \hat{\Lambda}_{0k}(t) - \Lambda_{0k}(t)\right)$$

to the multivariate Gaussian random field $\mathcal{W}(t) = (\mathcal{W}_1(t), ..., \mathcal{W}_k(t))$ *on the space of continuous functions on* $D^k[0,\tau]$ *equipped with the metric* $d(\boldsymbol{f},\boldsymbol{g}) = \max_{i\in\{1,...,k\}}\{f_i(t) - g_i(t)\}$. *Such a result can be used for making simultaneous inference on* $\Lambda_{0i}(t)$ *at some fixed time points and constructing confidence bands for all the baseline or a subset of them depending on interest, or testing equality of the baseline hazards at two different locations* i *and* j. *The latter and former could be important for epidemiologists and authorities since the results can be used to assess severity of a certain disease or pandemic at various times of the calendar year or having an idea about which locations among the ones under investigation have higher failure rates.*

3.2. JOINT MODELING

For a pair of units $(r, s) \in (\mathcal{L}_i, \mathcal{L}_j)$ and $t \in [0, \tau]$, counting, at-risk, and compensator processes are defined as before by

$$[N_i^{(r)}(t), Y_i^{(r)}(t), A_i^{(r)}(t)] \quad \text{and} \quad [N_j^{(s)}(t), Y_j^{(s)}(t), A_j^{(s)}(t)],$$

respectively. Then, $\{M_i^{(r)}(t) : t \in [0, \tau]\}$ and $\{M_j^{(s)}(t) : t \in [0, \tau]\}$ are each a zero-mean martingale with respect to the filtration $\mathcal{F}_{i,t}^{(r)}$ and $\mathcal{F}_{j,t}^{(s)}$ respectively. With a view toward joint modeling, for $(t_1, t_2) \in [0, \tau]^2$, we introduce the joint counting process $N_{ij}^{(r,s)}(\cdot, \cdot)$ by $N_{ij}^{(r,s)}(t_1, t_2) = \mathrm{I}\{T_i^{(r)} \geq t_1, T_j^{(s)} \geq t_2\}$. The covariance function $\mathrm{cov}(M_i^{(r)}(t_1), M_j^{(s)}(t_2))$ is defined by

$$E(M_i^{(r)}(t_1)M_j^{(s)}(t_2)|T_i^{(r)} > t_1, T_j^{(s)} > t_2) = A_{i,j}^{(r,s)}(t_1, t_2) = \langle M_i^{(r)}(t_1), M_j^{(s)}(t_2)\rangle.$$

Using stochastic integration theory,

$$E\left(M_i^{(r)}(t_1)M_j^{(s)}(t_2) - \int_0^{t_1}\int_0^{t_2} Y_i^{(r)}(u_1)Y_j^{(s)}(u_2)A_{i,j}^{(r,s)}(du_1, du_2)\right) = 0.$$

The spatial correlation between two locations implies that the covariance function depends on the spatial parameter $\boldsymbol{\delta}$ via the spatial correlation ρ_{ij} by virtue of the transformation leading to the construction of the MGRF. Let $G(\cdot, \cdot; \rho_{ij})$ be the bivariate survival function of the transformed failure times $\tilde{T}_i$ and $\tilde{T}_j$. Then, the original bivariate survivor function $\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij})$ for $(r, s) \in (\mathcal{L}_i, \mathcal{L}_j)$ is given by

$$\begin{aligned} \bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}) &= \mathrm{P}\left(T_i^{(r)} > t_1, T_j^{(s)} > t_2; \rho_{ij}\right) \\ &= G\left[\Phi^{-1}(F_i^{(r)})(t_1), \Phi^{-1}(F_j^{(s)}(t_2)); \rho_{ij}\right], \end{aligned} \tag{12}$$

with $F_i^{(r)}(t_1)$ and $F_j^{(s)}(t_2)$ being the marginal distribution functions of $T_i^{(r)}$ and $T_j^{(s)}$ respectively. Following Prentice and Cai (1992), $A_{ij}^{(r,s)}(t_1, t_2)$, the joint compensator is given by

$$A_{ij}^{(r,s)}(dt_1, dt_2; \rho_{ij}) = A_0[\Lambda_i^{(r)}(t_1), \Lambda_j^{(s)}(t_2); \rho_{ij}]\Lambda_i^{(r)}(dt_1)\Lambda_j^{(s)}(dt_2),$$

with the baseline joint compensator $A_0[\cdot, \cdot; \rho_{ij}]$ given by

$$\begin{aligned} A_0(t_1, t_2; \rho_{ij}) \quad = \quad & \frac{\partial^2}{\partial t_1 \partial t_2}\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}) + \bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}) \\ & + \frac{\partial}{\partial t_1}\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}) + \frac{\partial}{\partial t_2}\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}). \end{aligned}$$

Remark 4 *The covariance function $A_{ij}^{(r,s)}(dt_1, dt_2; \rho_{ij})$ in conjunction with $\Lambda_{0i}(dt_1)$ and $\Lambda_{0j}(dt_2)$ determines the joint distribution of T_i and T_j given the covariates $\boldsymbol{x}_i^{(r)}$ and $\boldsymbol{x}_j^{(s)}$. The original bivariate survivor function of T_i and T_j given in (12) can be taken to be of the Clayton family (cf. Clayton (1978)) or the Frank family model (cf. Genest (1987)). For the Clayton model for instance, the joint survivor function takes the form*

$$\bar{F}(t_i, t_j; \rho_{ij}) = \left(e^{t_i\rho_{ij}} + e^{t_j\rho_{ij}} - 1\right)^{-\frac{1}{\rho_{ij}}}.$$

4. ESTIMATION

This section is dedicated to discuss theory on estimating regression and spatial parameters.

4.1. WEIGHTED ESTIMATING FUNCTIONS

Estimation of parameters with spatially correlated random censorship data poses challenges because: (i) the high dimension of the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\delta}; \boldsymbol{\Lambda}_0)$, and (ii) the full likelihood L($\boldsymbol{\theta}$|Data) is intractable. Since it is quite difficult to apply direct maximum likelihood method in the spirit of Jacod (1976), the *pairwise likelihood* approach is adapted as an alternative. The main reference is Lindsay (1988). See also Varin et al. (2011) for an overview of composite likelihood applications in various fields. Lindsay et al. (2011) also discusses issues and strategies for the selection of composite likelihood.

The idea is to form pairwise likelihoods, a product of likelihoods for data in two spatial locations that can be the basis of an unbiased estimating function, and then be used for parameter estimation. It is a special case of a more general class of pseudo likelihoods called *composite likelihoods* which allows addition of likelihoods in a situation where the components do not represent independent replicates. The technique has good theoretical properties and behaves well in many applications concerning spatial statistics Hjort and Omre (1994); Heagerty and Lele (1998); Lele and Taper (2002); Varin and Vidoni (2005); Varin et al. (2005). Moreover, it is robust to model mis-specification, is computationally advantageous when dealing with data that has a complex structure, and the estimated parameter is the same as in the complete model Lindsay et al. (2011). In the present setting, for estimating $\boldsymbol{\beta}$, and as will be seen later, the covariation of the vector $\mathbf{M}(t) = (M_1(t), ..., M_k(t))$ depends on the spatial correlation parameter $\boldsymbol{\delta}$ and the unequal number of units per geographical site. The spatial dependency between locations may be severe or moderate. As indicated in Liang and Zeger (1986), there may be a loss in efficiency of the estimator of $\boldsymbol{\beta}$ when accounting for spatial correlation, especially when it is severe. In the aim of increasing efficiency, the idea of Liang and Zeger (1986) is followed by proposing generalized estimating equations that include weights in the estimating functions. The weights are chosen in general to balance out severe versus moderate spatial dependence.

With a view toward estimating $\boldsymbol{\beta}$ that accounts for pairwise spatial correlation between two locations $(i, j) \in \mathcal{L}^2$, more notation is introduced in the sequel. If $\mathbf{a} = (a_1, a_2)$ is a 1×2 row vector and its transpose denoted by $\mathbf{a}'$ is a 2×1 column vector. For $(r, s) \in (\mathcal{L}_i, \mathcal{L}_j)$, define $\mathbf{H}_{ij}^{(r,s)}(t) = (H_i^{(r)}(t), H_j^{(s)}(t))$, $\mathbf{M}_{ij}^{(r,s)}(t) = (M_i^{(r)}(t), M_j^{(s)}(t))'$. A 2×2 matrix $\mathbf{W}^{ij}(\boldsymbol{\delta}) = (w^{ij}(\boldsymbol{\delta}))$ is further defined whose elements are function of the spatial correlation $\boldsymbol{\delta}$ and the number of units in locations i and j by

$$\mathbf{W}^{ij}(\boldsymbol{\delta}_0) = \begin{pmatrix} w_{11}^{ij}(\boldsymbol{\delta}_0) & w_{12}^{ij}(\boldsymbol{\delta}_0) \\ w_{21}^{ij}(\boldsymbol{\delta}_0) & w_{22}^{ij}(\boldsymbol{\delta}_0) \end{pmatrix}. \tag{13}$$

Then, given $\boldsymbol{\delta}_0$, the *pairwise estimating* equation is defined for $\boldsymbol{\beta}$ between two locations at time t by

$$\mathrm{U}^{[ij]}(t, \boldsymbol{\beta}|\boldsymbol{\delta}_0) = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \int_0^t \mathbf{H}_{ij}^{(r,s)}(u)\mathbf{W}^{ij}(\boldsymbol{\delta}_0)\mathbf{M}_{ij}^{(r,s)}(u)du. \tag{14}$$

At time $t \in \mathcal{T}$, the generalized estimating equation for $\boldsymbol{\beta}$ over all pairs is

$$\mathrm{U}(t, \boldsymbol{\beta}|\boldsymbol{\delta}_0) = \sum_{i \leq j} \mathrm{U}^{[ij]}(t, \boldsymbol{\beta}|\boldsymbol{\delta}_0). \tag{15}$$

The weights matrix given in (13) adapts to dependencies between locations, especially when dependency is strong and censoring within a location is light, and help improve efficiency of the estimates under such scenarios. Two remarks are worth mentioning here.

Remark 5 *Replacing $\boldsymbol{W}^{ij}(\boldsymbol{\delta}_0)$ by the identity matrix in (14), results in the case of no spatial correlation between locations. Observe also that if $\mathbf{W}^{ij}(\boldsymbol{\delta}_0)$ is replaced by the variance covariance matrix of $(M_1(t), M_2(t), ..., M_k(t))$ with $M_i(t) = \sum_{r=1}^{n_i} M_i^{(r)}(t)$, the weights actually depend on the regression coefficient $\boldsymbol{\beta}$ as well as the spatial correlation parameter $\boldsymbol{\delta}$ since its compensator $\boldsymbol{A}(t)$ depends on both. Then (14) can be re-expressed as a function of $\boldsymbol{\beta}$ alone by first replacing $\boldsymbol{\delta}$ in (14) by $\hat{\boldsymbol{\delta}}$, a $\sqrt{n} = \sqrt{\sum_i n_i}$-consistent estimator of $\boldsymbol{\delta}_0$ that satisfies $\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) = O_p(1)$. In that case, (15) will take the form $U(t; \boldsymbol{\beta}, \hat{\boldsymbol{\delta}})$, and the estimated $\boldsymbol{\beta}$ would still be consistent. For further discussions, cf. Liang and Zeger (1986). This approach will be adopted in the estimation of $\boldsymbol{\beta}$ in the next section.*

Remark 6 *An important property of the estimating function (15) is the robustness of the resulting estimator $\hat{\boldsymbol{\beta}}$ even if the spatial correlation is misspecified and remains asymptotically unbiased even under the misspecification.*

Examining $\mathrm{U}^{[ij]}(\cdot,\cdot|\boldsymbol{\delta}_0)$, it can be written as a sum of four terms each of which is given below

$$
\begin{aligned}
U_1^{ij}(t) &= \sum_{r=1}^{n_i}\int_0^t w_{11}^{ij}H_i^{(r)}(u,\boldsymbol{\beta})M_i^{(r)}(du),\\
U_2^{ij}(t) &= \sum_{r=1}^{n_i}\sum_{s=1}^{n_j}\int_0^t w_{12}^{ij}H_j^{(s)}(u,\boldsymbol{\beta})M_i^{(r)}(du),\\
U_3^{ij}(t) &= \sum_{r=1}^{n_i}\sum_{s=1}^{n_j}\int_0^t w_{21}^{ij}H_i^{(r)}(u,\boldsymbol{\beta})M_j^{(s)}(du),\\
U_4^{ij}(t) &= \sum_{s=1}^{n_j}\int_0^t w_{22}^{ij}H_j^{(s)}(u,\boldsymbol{\beta})M_j^{(s)}(du).
\end{aligned}
\tag{16}
$$

For the purpose of estimating $\boldsymbol{\delta}_0$, note that $\mathcal{E}\{M_i^{(r)}(t_1)M_j^{(s)}(t_2) - A_{ij}^{(r,s)}(dt_1,dt_2;\rho_{ij})\} = 0$. The goal is to find a weighted function of $M_i^{(r)}(t_1)M_j^{(s)}(t_2) - A_{i,j}^{(r,s)}(dt_1,dt_2;\rho_{ij})$ that can serve as an estimating function for $\boldsymbol{\delta}_0$ with the flavor of score function. Define the $(k\times k)$ matrix $\mathbf{A}(t_1,t_2;\rho(\boldsymbol{\delta})) = \left(A_{ij}(t_1,t_2;\rho(\boldsymbol{\delta}))\right)_{i,j=1,\ldots,k}$, with $(i,j)^{th}$ entry given by

$$
A_{ij}(t_1,t_2;\rho(\boldsymbol{\delta}) = \sum_{r=1}^{n_i}\sum_{s=1}^{n_j} A_{ij}^{(r,s)}(t_1,t_2;\rho(\boldsymbol{\delta})).
$$

Let $\nabla_{\delta_l}\mathbf{A}(t_1,t_2;\rho(\boldsymbol{\delta}))$, $l=1,...,q$, be the matrix of elementwise derivatives of $\mathbf{A}(t,\rho(\boldsymbol{\delta}))$ with respect to δ_l. Define

$$
\Pi_l = \mathbf{A}^{-1}\left[\nabla_{\delta_l}\mathbf{A}\right]\mathbf{A}^{-1},
$$

where $\mathbf{A}$ is for $\mathbf{A}(t_1,t_2;\rho(\boldsymbol{\delta}))$ for compactness. Then, for $l=1,...,q$, following Cressie (1993), Page 483, it can be shown that $\mathcal{E}(\mathbf{M}(t))\Pi_l\mathcal{E}(\mathbf{M}(t)) + \mathrm{tr}(\Pi_l\mathbf{A}) = 0$, where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Consequently, a score function can be defined for estimating the l^{th} component of $\boldsymbol{\delta}$ using two locations by

$$
\mathrm{U}_{\delta_l}^{ij}(t_1,t_2) = \mathbf{M}(t)\Pi_l\mathbf{M}'(t) + \mathrm{tr}(\Pi_l\mathbf{A}) := \mathbf{M}(t)\Pi_l\mathbf{M}'(t) + \mathrm{tr}(\mathbf{A}^{-1}\mathbf{A}_{\delta_l}). \tag{17}
$$

The expression in (17) can be viewed as a score process and its sum over all pairwise spatial locations (i, j) can serve as an estimating function for δ_l. So, the estimating function over all pairs of spatial locations for $\boldsymbol{\delta}$ is the $q \times 1$ vector $\mathrm{U}_{\boldsymbol{\delta}}(t_1, t_2; \rho(\boldsymbol{\delta})) = (\mathrm{U}_{\delta_l}(t), l = 1, ..., q)'$ where $\mathrm{U}_{\delta_i}(s, t)$ is given by

$$\mathrm{U}_{\delta_l}(t_1, t_2; \rho(\boldsymbol{\delta})) = \sum_{(i,j), i \leq j} \mathrm{U}^{ij}_{\delta_l}(t_1, t_2; \rho(\boldsymbol{\delta})).$$

4.2. UNBIASED ESTIMATING FUNCTIONS

The unbiased estimating functions concept is one of the requirements for showing the existence of consistent solutions to the equations $\mathrm{U}_{\delta_l}(t_1, t_2) = 0$ and $\mathrm{U}^{[ij]}(\boldsymbol{\beta}; \cdot, \cdot|\boldsymbol{\delta}_0) = \mathbf{0}$, respectively. Two conditions need to be satisfied for the existence and consistency of the estimate: (i) the asymptotic unbiasedness of the two estimating functions, and (ii) the existence of in-probability limit of the information matrix. To show (i) for $\mathrm{U}_{\delta_l}(t)$, the concept of mixing coupled with the multivariate Chebyshev inequality is applied in particular. As for $\mathrm{U}^{[ij]}(\boldsymbol{\beta}; \cdot, \cdot|\boldsymbol{\delta}_0) = \mathbf{0}$, some regularity conditions applied on its derivatives yield the result. Finally, Theorem 2 of Foutz (1976) will be used to show existence and consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$, the sequence of solutions to the aforementioned equations.

In geostatistics, asymptotic properties can be investigated in two different ways: the *Increasing domain asymptotic* or the *Infill asymptotic*. The increasing domain asymptotic is a sampling structure in spatial statistics where new observations are added at the boundary points of an area, whereas the infill asymptotic consists of a sampling structure where new observations are added in between existing locations. The latter is appropriate when the spatial locations are fixed and in a bounded domain and one is interested in adding new observations to each location. This study will use the infill asymptotic since the number of locations is fixed at k. Letting $\min_i\{n_1, ..., n_k\} \to \infty$ satisfies the infill asymptotic criterion. Therefore, in what follows, the statement $n = \min_i n_i \to \infty$ means *Infill Asymptotic*. Readers are refereed to Cressie (1993), Section 7.3.1, page 480 for details.

4.2.1. Regularity Conditions.

I For $i = 1, ..., k$, $\Lambda_{0i}(t) < \infty$.

II For each $(i, r) \in \mathcal{L} \times \mathcal{L}_i$ and $t \in \mathcal{T}$, $\mathbf{x}_i^{(r)}(t)$ is uniformly bounded.

III For $(i, r) \in \mathcal{L} \times \mathcal{L}_i$ and for each i, define $S_i^{(m)}(\boldsymbol{\beta}, t) = \sum_{r=1}^{n_i} Y_i^{(r)}(t)[\mathbf{x}_i^{(r)}(t)]^{\otimes m} e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(t)}$. Let $\mathcal{E}$ denote expectation operator. There exists $\mathcal{B} \subset \Re^p$, a neighborhood of $\boldsymbol{\beta}_0$, and functions $s_i^{(m)}(\boldsymbol{\beta}, t)$ such that $\mathcal{E}(S_i^{(m)}(\boldsymbol{\beta}, t)) = s_i^{(m)}(\boldsymbol{\beta}, t)$, and that, for each $i = 1, ..., k$, $m = 0, 1, 2$

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} \|S_i^{(m)}(\boldsymbol{\beta}, t) - s_i^{(m)}(\boldsymbol{\beta}, t)\| \xrightarrow{p} 0,$$

and the $s_i^{(m)}(\boldsymbol{\beta}, t)$ are uniformly bounded on $[0, \tau] \times \mathcal{B}$ with continuous partial derivatives.

IV Define

$$\hat{\boldsymbol{\Omega}}(\boldsymbol{\theta}) = \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}) = \frac{1}{n^2} \sum_{i \le j} \sum_{i' \le j'} \mathcal{E}\left[\mathbf{V}_{ij}(\hat{\boldsymbol{\theta}})\mathbf{V}'_{i'j'}(\hat{\boldsymbol{\theta}})\right].$$

There exists a positive definite matrix $\boldsymbol{\Omega}(\boldsymbol{\theta}_0)$ such that

$$\sup_{t \in [0, t^*]} \left\|\hat{\boldsymbol{\Omega}}(\boldsymbol{\theta}) - \boldsymbol{\Omega}(\boldsymbol{\theta}_0)\right\| \xrightarrow{p} 0.$$

V Weight matrices conditions

i $\|\mathbf{W}(\boldsymbol{\delta}) - \mathbf{W}(\boldsymbol{\delta}_0)\| \xrightarrow{p} 0$, where $\|\mathbf{a} - \mathbf{b}\| = \sup_{ij} |a_{ij} - b_{ij}|$.

ii For $\cdot\cdot \in \{11, 12, 21, 22\}$, $\boldsymbol{\nabla}_{\boldsymbol{\beta}} w_{\cdot\cdot}^{ij}(\hat{\boldsymbol{\delta}}) \xrightarrow{p} \boldsymbol{\nabla}_{\boldsymbol{\beta}} w_{\cdot\cdot}^{ij}(\boldsymbol{\delta}_0)$ and $\boldsymbol{\nabla}_{\boldsymbol{\beta}} w_{\cdot\cdot}^{ij}(\boldsymbol{\delta})$ are continuous functions of $\boldsymbol{\delta}_0$.

VI For $\cdot\cdot \in \{11, 22\}$, $m = 0, 1$, $n = \sum_{i=1}^{k} n_i$, all of the following three quantities converge in probability to zero.

$$\sup_{(\boldsymbol{\beta},t)\in\mathcal{B}\times[0,\tau]} \left\| \frac{1}{n} \sum_{i\geq j} \sum_{r=1}^{n_i} H_i^{(r)}(t) w_{\cdot\cdot}^{ij} Y_i^{(r)}(t) e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(t)} \otimes \mathbf{x}_i^{(r)m}(t) - s(w_{\cdot\cdot}^{(m)}, \boldsymbol{\beta}; t) \right\|$$
$$\sup_{(\boldsymbol{\beta},t)\in\mathcal{B}\times[0,\tau]} \left\| \frac{1}{n} \sum_{i\geq j} \sum_{s=1}^{n_j} \sum_{r=1}^{n_i} H_i^{(r)}(t) w_{12}^{ij} Y_j^{(s)}(t) e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(t)} \otimes \mathbf{x}_j^{(s)m}(t) - s(w_{12}^{(m)}, \boldsymbol{\beta}; t) \right\|$$
$$\sup_{(\boldsymbol{\beta},t)\in\mathcal{B}\times[0,\tau]} \left\| \frac{1}{n} \sum_{i\geq j} \sum_{s=1}^{n_j} \sum_{r=1}^{n_i} H_j^{(s)}(t) w_{21}^{ij} Y_i^{(r)}(t) e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(t)} \otimes \mathbf{x}_i^{(r)m}(t) - s(w_{21}^{(m)}, \boldsymbol{\beta}; t) \right\|$$

VII Mixing conditions for unbiasedness of the score process for $\boldsymbol{\delta}$

i Let $\mathcal{M}_{-\infty}^{k}$ and $\mathcal{M}_{k}^{\infty}$ be the σ-fields generated by the observables $\{\mathbf{O}(\mathrm{I}_i) : i \leq k\}$ and $\{\mathbf{O}(\mathrm{I}_i) : i > k\}$ respectively. The mixing condition is

$$\sup_{A\in\mathcal{M}_{-\infty}^{k}, B\in\mathcal{M}_{k}^{\infty}} |P(A\cap B) - P(A)\cdot P(B)| \leq C(||i-j||)$$

where $C(\cdot)$ is some mixing function.

ii For $(i, j) \in \mathcal{L}\times\mathcal{L}$, the weight $\boldsymbol{v}_{ij}(\boldsymbol{\delta})$ on the compensators are uniformly bounded with continuous partial derivatives with respect to $\boldsymbol{\delta}$.

VIII Joint compensator condition

i The function $A_0(t_1, t_2; \boldsymbol{\delta})$ exists and has bounded second derivatives in the range of the arguments $(\boldsymbol{\beta}, \boldsymbol{\delta})$ for all $(\boldsymbol{\beta}, \boldsymbol{\delta}) \in \mathcal{B}\times\mathcal{D}$, where $\mathcal{D} \subset \Re^q$. Moreover, $\mathrm{A}_0(t_1, t_2; \boldsymbol{\delta})$ is continuously differentiable as a function of $(\boldsymbol{\beta}, \boldsymbol{\delta})$ and the partial derivative $A_0(dt_1, t_2, \boldsymbol{\delta}) := A_0^{100}$, $A_0(t_1, dt_2, \delta) := A_0^{010}$, $A_0(t_1, t_2, \nabla_\delta) := A_0^{001}$ and $A_0(dt_1, dt_2, \boldsymbol{\delta}) := A_0^{110}$ are bounded on $\mathcal{B}\times\mathcal{D}$ for all values of the arguments.

ii Any linear combination of the joint compensator partial derivatives with respect to any of its arguments converges to a bounded function.

Discussion on Regularity Conditions: Conditions I to III are the regular stability conditions imposed on derivatives of the at-risk process that arise in models that involve the Cox hazard functions. They are the expectations and variances of the covariates $\mathbf{x}_i^{(r)}(t)$ of (i, r). Condition VI is on stability condition of the weight applied to the estimated spatial correlation parameters, whereas VII are stability conditions guaranteeing convergence of the variance-covariance matrix of the joint process, namely the block $\mathbf{\Sigma}_{11}$. Likewise, Condition VI together with VII pertains to infill asymptotic stability of the block $\mathbf{\Sigma}_{21}$ and $\mathbf{\Sigma}_{22}$, with VII only needed for the latter.

The following theorem pertains to the asymptotic unbiasedness of $\mathbf{U}(\boldsymbol{\beta}, t)$.

Theorem 2 *Under Conditions I to V, as $n \to \infty$, $\sup_{t\in[0,\tau]} \frac{1}{n}\boldsymbol{U}(\boldsymbol{\beta}, t|\hat{\boldsymbol{\delta}})| \xrightarrow{p} 0$.*

Proof: Using the fact that $\mathrm{U}^{[ij]}(\boldsymbol{\beta}, t|\hat{\boldsymbol{\delta}})$ is the sum of four terms, by the triangle inequality, the following is true:

$$\begin{aligned} |\mathrm{U}^{[ij]}(\boldsymbol{\beta}, t|\hat{\boldsymbol{\delta}})| \quad &\leq \quad |\mathrm{U}_1^{ij}(\boldsymbol{\beta}, t)| + |\mathrm{U}_2^{ij}(\boldsymbol{\beta}, t)| \\ &\quad + |\mathrm{U}_3^{ij}(\boldsymbol{\beta}, t)| + |\mathrm{U}_4^{ij}(\boldsymbol{\beta}, t)|. \end{aligned} \tag{18}$$

It suffices to show that each term in the RHS of (18) converges to zero in probability. Without loss of generality, this is only shown for $\mathrm{U}_1^{ij}(\boldsymbol{\beta}, t)$ only. Asymptotic negligibility of the remaining terms are obtained in similar manner. Thus,

$$
\begin{aligned}
\frac{1}{n}\mathrm{U}_1^{ij}(\boldsymbol{\beta}, t) \quad = \quad & \frac{1}{n}\sum_{i=1}^{k}\sum_{r=1}^{n_i}\int_0^{\tau}\left\{H_i^{(r)}(t)\left(\sum_{i\le j} w_{11}^{ij}\right) - \frac{s(t, w_{11}^{(0)})}{s_i^{(0)}(t)}\right\} dM_i^{(r)}(t) \\
& -\frac{1}{s_i^{(0)}(t)}\left\{\frac{1}{n}\sum_{i\le i} H_i^{(r)}(t)Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0' \mathbf{x}_i^{(r)}(t)} w_{11}^{ij} - s_i(t, w_{11}^{(0)})\right\} dM_i^{(r)}(t) \\
& -s(w_{11}^{(0)}(t))\left[\left(\frac{1}{n}\sum_{i=1}^{k} Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0' \mathbf{x}_i^{(r)}(t)}\right)^{-1} - \left(s_i^{(0)}(t)\right)^{-1}\right] dM_i^{(r)}(t) \qquad (19) \\
& -\left[\frac{1}{n}\sum_{ß\le j} H_i^{(r)}(t)Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_i^{(r)}(t)} w_{11}^{ij} - s(t, w_{11}^{(0)})\right] \\
& \cdot\left[\left(\frac{1}{n}\sum_{i=1}^{k} Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0' \mathbf{x}_i^{(r)}(t)}\right)^{-1} - \left(s_i^{(0)}(t)\right)^{-1}\right] dM_i^{(r)}(t).
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\mathrm{U}_2^{ij}(\boldsymbol{\beta}, t) \quad = \quad & \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r=1}^{n_i}\sum_{s=1}^{n_j}\int_0^{\tau}\left\{H_i^{(r)}(t)\left(\sum_{i\le j} w_{12}^{ij}\right) - \frac{s(t, w_{12}^{(0)})}{s_j^{(0)}(t)}\right\} dM_j^{(s)}(t) \\
& -\frac{1}{s_j^{(0)}(t)}\left\{\frac{1}{n}\sum_{i\le i} H_i^{(r)}(t)Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0' \mathbf{x}_j^{(s)}(t)} w_{12}^{ij} - s_i(t, w_{12}^{(0)})\right\} dM_j^{(s)}(t) \\
& -s(w_{12}^{(0)}(t))\left[\left(\frac{1}{n}\sum_{j=1}^{k} Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0' \mathbf{x}_j^{(s)}(t)}\right)^{-1} - \left(s_j^{(0)}(t)\right)^{-1}\right] dM_j^{(s)}(t) \qquad (20) \\
& -\left[\frac{1}{n}\sum_{ß\le j} H_i^{(r)}(t)Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_j^{(s)}(t)} w_{12}^{ij} - s(t, w_{12}^{(0)})\right] \\
& \cdot\left[\left(\frac{1}{n}\sum_{j=1}^{k} Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0' \mathbf{x}_j^{(s)}(t)}\right)^{-1} - \left(s_j^{(0)}(t)\right)^{-1}\right] dM_j^{(s)}(t).
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\mathrm{U}_3^{ij}(\boldsymbol{\beta},t) \quad = \quad & \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r=1}^{n_i}\sum_{s=1}^{n_j}\int_0^{\tau}\left\{H_j^{(s)}(t)\left(\sum_{i\le j} w_{21}^{ij}\right)-\frac{s(t,w_{21}^{(0)})}{s_i^{(0)}(t)}\right\}dM_i^{(r)}(t)\\
& -\frac{1}{s_i^{(0)}(t)}\left\{\frac{1}{n}\sum_{i\le i}H_j^{(s)}(t)Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_i^{(r)}(t)}w_{21}^{ij}-s_i(t,w_{21}^{(0)})\right\}dM_i^{(r)}(t)\\
& -s(w_{21}^{(0)}(t))\left[\left(\frac{1}{n}\sum_{i=1}^{k}Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_i^{(r)}(t)}\right)^{-1}-\left(s_i^{(0)}(t)\right)^{-1}\right]dM_i^{(r)}(t) \qquad (21)\\
& -\left[\frac{1}{n}\sum_{ß\le j}H_j^{(s)}(t)Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_i^{(r)}(t)}w_{21}^{ij}-s(t,w_{21}^{(0)})\right]\\
& \cdot\left[\left(\frac{1}{n}\sum_{i=1}^{k}Y_i^{(r)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_i^{(r)}(t)}\right)^{-1}-\left(s_i^{(0)}(t)\right)^{-1}\right]dM_i^{(r)}(t).
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\mathrm{U}_4^{ij}(\boldsymbol{\beta},t) \quad = \quad & \frac{1}{n}\sum_{j=1}^{k}\sum_{s=1}^{n_j}\int_0^{\tau}\left\{H_j^{(s)}(t)\left(\sum_{i\le j} w_{22}^{ij}\right)-\frac{s(t,w_{22}^{(0)})}{s_j^{(0)}(t)}\right\}dM_j^{(s)}(t)\\
& -\frac{1}{s_j^{(0)}(t)}\left\{\frac{1}{n}\sum_{i\le i}H_j^{(s)}(t)Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_j^{(s)}(t)}w_{22}^{ij}-s_j(t,w_{22}^{(0)})\right\}dM_j^{(s)}(t)\\
& -s(w_{22}^{(0)}(t))\left[\left(\frac{1}{n}\sum_{j=1}^{k}Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_j^{(s)}(t)}\right)^{-1}-\left(s_j^{(0)}(t)\right)^{-1}\right]dM_j^{(s)}(t) \qquad (22)\\
& -\left[\frac{1}{n}\sum_{ß\le j}H_j^{(s)}(t)Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_j^{(s)}(t)}w_{22}^{ij}-s(t,w_{22}^{(0)})\right]\\
& \cdot\left[\left(\frac{1}{n}\sum_{j=1}^{k}Y_j^{(s)}(t)e^{\boldsymbol{\beta}_0'\mathbf{x}_j^{(s)}(t)}\right)^{-1}-\left(s_j^{(0)}(t)\right)^{-1}\right]dM_j^{(s)}(t).
\end{aligned}
$$

Each one of the terms on the right hand side of (19), (20),(21) and (22) are $o_p(1)$ when $n \to \infty$ per the regularity conditions I to V. Therefore $n^{-1}\mathrm{U}_1^{ij}(\boldsymbol{\beta},t) = o_p(1)$. Likewise, under conditions I to V, it can be shown that the remaining three terms in $n^{-1}\mathrm{U}^{[ij]}(\boldsymbol{\beta},t)$ are all asymptotically negligible. Hence $\sup_{t\in[0,\tau]} n^{-1}\mathrm{U}^{[ij]}(\boldsymbol{\beta},t|\hat{\boldsymbol{\delta}}) \xrightarrow{p} 0$ as $n \to \infty$, completing the proof of the asymptotic unbiasedness of $n^{-1}\mathrm{U}^{[ij]}(\boldsymbol{\beta},t|\hat{\boldsymbol{\delta}})$.

For the purpose of showing unbiasedness of the score associated with the spatial correlation parameter, let v_{ij} be the weight between two sites. The weight v_{ij} can be taken to be a function of the spatial correlation $\rho_{ij}(\boldsymbol{\delta})$ between locations i and j and will help increase efficiency of $\hat{\boldsymbol{\delta}}$. With no loss of generality, it suffices to consider a score process between two locations of the form

$$\mathrm{U}^{ij}(t_1, t_2; \boldsymbol{\delta}) = \mathrm{v}_{ij}\left[M_i(t_1)M_j(t_2) - A_{ij}(t_1, t_2; \boldsymbol{\delta})\right],$$

where $A_{ij}(t_1, t_2; \boldsymbol{\delta}) = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \mathrm{A}_{ij}^{(r,s)}(t_1, t_2; \boldsymbol{\delta})$ and $\mathrm{M}_i(t_1) = \sum_{r=1}^{n_i} \mathrm{M}_i^{(r)}(t_1)$, $\mathrm{M}_j(t_2) = \sum_{s=1}^{n_j} \mathrm{M}_j^{(s)}(t_2)$. The corresponding estimating function for $\boldsymbol{\delta}$ over all pairs is given by

$$\mathrm{U}(t_1, t_2; \boldsymbol{\delta}) = \sum_{i \leq j} \mathrm{U}^{ij}(t_1, t_2; \boldsymbol{\delta}).$$

The next theorem is on the asymptotic unbiasedness of $\mathrm{U}(t_1, t_2; \boldsymbol{\delta})$.

Theorem 3 *Under Conditions VI and VIII, as $n \longrightarrow \infty$,*

$$U(t_1, t_2; \boldsymbol{\delta}) = \frac{1}{n} \sum_{(i,j), i \leq j} U^{ij}(t_1, t_2; \boldsymbol{\delta}) \xrightarrow{p} 0.$$

Proof: The mixing condition along with Chebyshev inequality and Condition VII is applied. Let $I_n(\iota_n) = \{i, j / \|i - j\| \geq \iota_n\}$. The set $I_n(\iota_n)$ gives the range beyond which the spatial correlation impact is negligible. The cut off point ι_n depends on the number of locations. In what follows, A' denote complement of a set A. For $\epsilon > 0$, via Chebyshev inequality and

Condition VII,

$$
\begin{aligned}
P\left(n^{-1}\sum_{(i,j),i\le j}\mathrm{U}^{ij}(t_1,t_2;\boldsymbol{\delta})>\epsilon\right) &= \mathrm{P}\left(\sum_{(i,j),i\le j}\mathrm{v}_{ij}\mathrm{M}_i(t_1)\mathrm{M}_j(t_2)>n\epsilon\right.\\
&\quad\left.+\sum_{(i,j),i\le j}\mathrm{v}_{ij}A_{ij}(t_1,t_2)\right)\\
&\le \frac{\sum_{(i,j),i\le j}\mathcal{E}(\mathrm{v}_{ij}M_i(t_1)M_j(t_2))}{(n\epsilon+\sum_{(i,j),i\le j}\mathrm{v}_{ij}A_{ij}(t_1,t_2))^2}\\
&\le \frac{\sum_{(i,j)\in I_n(\iota_n)}\mathcal{E}(M_i(t_1)M_j(t_2))}{n^2\epsilon^2}\\
&\quad+\frac{\sum_{(i,j)\in I'_n(\iota_n)}\mathcal{E}(M_i(t_1)M_j(t_2))}{n^2\epsilon^2}.
\end{aligned}
$$

With a proper choice of mixing function, the last inequality in previous display converges to 0 under the mixing condition VII.

5. LARGE SAMPLE PROPERTIES

This section is devoted to the large sample properties of the estimators. Let $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\delta})$, $\mathrm{U}_1(\boldsymbol{\beta},t) = \mathrm{U}^{[ij]}(t,\boldsymbol{\beta}|\boldsymbol{\delta}_0)$ and $\mathrm{U}_2(t_1,t_2;\boldsymbol{\delta}) = \mathrm{U}^{ij}(t_1,t_2;\boldsymbol{\delta})$. Consider the vector of score processes $\mathbf{V}(t) = (\mathrm{U}_1(\boldsymbol{\beta},t),\mathrm{U}_2(t_1,t_2;\boldsymbol{\delta}))'$ such that

$$
\mathbf{V}(\boldsymbol{\theta};t,t_1,t_2)=\mathbf{V}(\boldsymbol{\theta})=\frac{1}{n}\sum_{i\le j}\begin{pmatrix}\mathrm{U}_1(\boldsymbol{\beta})\\ \mathrm{U}_2(\boldsymbol{\delta})\end{pmatrix}=\frac{1}{n}\sum_{i\le j}\begin{pmatrix}\mathrm{U}^{[ij]}(t,\boldsymbol{\beta}|\boldsymbol{\delta}_0)\\ \mathrm{U}^{ij}(t_1,t_2;\boldsymbol{\delta})\end{pmatrix}:=\frac{1}{n}\sum_{i\le j}\mathbf{V}_{ij}(\boldsymbol{\theta}).\tag{23}
$$

The in-probability limit of the variance covariance matrix of $\mathbf{V}(\boldsymbol{\theta};t,t_1,t_2)$ is given by

$$
\boldsymbol{\Sigma}_n=\frac{1}{n}\sum_{i\le j}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{V}_{ij}(\boldsymbol{\theta})\right)=\begin{pmatrix}\nabla_{\boldsymbol{\beta}}\mathrm{U}_1(\boldsymbol{\beta},t) & \nabla_{\boldsymbol{\delta}}\mathrm{U}_1(\boldsymbol{\beta},t)\\ \nabla_{\boldsymbol{\beta}}\mathrm{U}_2(\boldsymbol{\delta},t) & \nabla_{\boldsymbol{\delta}}\mathrm{U}_2(\boldsymbol{\delta};t_1,t_2)\end{pmatrix}\xrightarrow{p}\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}.\tag{24}
$$

The next theorem is on the existence and consistency of the solution to $\mathbf{V}(\boldsymbol{\theta})=\mathbf{0}$.

Theorem 4 *(a) There exists a sequence of solutions $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\delta}}_n$ to the sequence of estimating equations $U_1(\boldsymbol{\beta}, t) = \mathbf{0}$ and $U_2(t_1, t_2; \boldsymbol{\delta}) = \mathbf{0}$.*
(b) Under Conditions I to VIII, and the infill asymptotic, $\hat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}_0$ and $\hat{\boldsymbol{\delta}}_n \xrightarrow{p} \boldsymbol{\delta}_0$.

Before proving the theorem, a discussion on $\hat{\lambda}_{0i}(t)$, $i = 1, ..., k$ is warranted since its consistency is required for the in-probability limit of the score variance. Recall that $\hat{\lambda}_{0i}(t)$ is given by

$$\hat{\lambda}_{0i}(t) = \frac{\sum_{r=1}^{n_i} dN_i^{(r)}(t)}{S_i^{(0)}(\boldsymbol{\beta}, t)},$$

and is a jump process and will possibly loses efficiency for large n. However, any loss of efficiency using it for the limit is minor as compared to using a more complicated smoothed estimator obtained via kernel and proposed in Ramlau-Hansen (1983) given by

$$\hat{\lambda}_{0i}^{K}(t) = \frac{1}{h_n} \int_0^t K\left(\frac{t-u}{h_n}\right) d\hat{\Lambda}_{0i}(u),$$

where $K(\cdot)$ is some kernel function and h_n, a sequence of positive constants. Although $\hat{\lambda}_{0i}^{K}(t)$ is smoother, both are, however consistent for $\lambda_{0i}(t)$, that is $\sup_{t\in[0,\tau]} |\hat{\lambda}_{0i}(t) - \lambda_{0i}(t)| \xrightarrow{p} 0$. This proceeds to using the consistent version $\hat{\lambda}_{0i}(t)$.

Proof: The inverse function theorem of Foutz (1976) is applied. Three conditions need to be satisfied. (i) asymptotic unbiasedness of the estimating functions, (ii) existence and continuity of the partial derivatives matrix and (iii) the negative definiteness of the matrix of partial derivatives at the true parameter value $\boldsymbol{\theta}_0$. Condition (i) has been already shown in Theorem 2 and Theorem 3. It remains to show (ii) and (iii). Consider $\mathrm{U}_1^{ij}(t)$ given by

$$\mathrm{U}_1^{ij}(t) = \frac{1}{n} \sum_{i \leq j} \sum_{r=1}^{n_i} \int_0^{\tau} w_{11}^{ij} H_i^{(r)}(u, \boldsymbol{\beta}) M_i^{(r)}(du).$$

Since $\lambda_{0i}(t)$ is unknown, substitute it by its consistent Breslow estimator. So the version is $\hat{U}_1^{ij}(t)$ given by

$$\begin{aligned}\hat{U}_1^{ij}(t) &= \frac{1}{n}\sum_{i\le j}\sum_{r=1}^{n_i}\int_0^t w_{11}^{ij} H_i^{(r)}(u,\boldsymbol{\beta}) \\ &\cdot\left\{ dN_i^{(r)}(u) - Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\right. \\ &\left.\cdot\left(\frac{\sum_{r=1}^{n_i} dN_i^{(r)}(u)}{S_i^{(0)}(\boldsymbol{\beta},u)}\right)\right\}.\end{aligned}$$

The gradient of $\hat{\mathrm{U}}_1^{ij}(t)$ with respect to $\boldsymbol{\beta}$ is

$$\begin{aligned}\boldsymbol{\nabla}_\beta \hat{U}_1^{ij}(\boldsymbol{\beta},\tau) &= \frac{1}{n}\sum_{i\le j}\sum_{r=1}^{n_i}\int_0^\tau \left[\boldsymbol{\nabla}_\beta H_i^{(r)}(u,\boldsymbol{\beta})\right] \\ &\cdot w_{11}^{ij}\left(dN_i^{(r)}(u) - Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du\right) \\ &+\frac{1}{n}\sum_{i\le j}\sum_{r=1}^{n_i}\int_0^\tau H_i^{(r)}(u,\boldsymbol{\beta})w_{11}^{ij}\left[-Y_i^{(r)}\mathbf{x}_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du\right. \\ &\left.+Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\otimes\frac{S_i^{(1)}(\boldsymbol{\beta},u)}{S_i^{(0)}(\boldsymbol{\beta},u)}\hat{\lambda}_{0i}(u)du\right] \\ \xrightarrow{p}\ & o_p(1) - \int_0^\tau\left[s\left(u;\boldsymbol{\beta},w_{11}^{(0)}\right)\otimes\frac{s_i^{(1)}(\boldsymbol{\beta},u)}{s_i^{(0)}(\boldsymbol{\beta},u)} - s\left(u;\boldsymbol{\beta},w_{11}^{(1)}\right)\right]\lambda_{0i}(u)du.\end{aligned}$$

Likewise, substituting the Breslow estimator in $\hat{\mathrm{U}}_2^{ij}(t)$, $\hat{\mathrm{U}}_3^{ij}(t)$ and $\hat{\mathrm{U}}_4^{ij}(t)$ the gradients of $\hat{U}_2^{ij}(t)$, $\hat{U}_3^{ij}(t)$ and $\hat{U}_4^{ij}(t)$ are

$$\begin{aligned}
\nabla_{\boldsymbol{\beta}}\hat{U}_2^{ij}(\boldsymbol{\beta},\tau) &= \frac{1}{n}\sum_{i\le j}\sum_{s=1}^{n_j}\sum_{r=1}^{n_i}\int_0^{\tau}\left[\nabla_{\boldsymbol{\beta}}H_j^{(s)}(u,\boldsymbol{\beta})\right] \\
&\quad \times w_{12}^{ij}\left(dN_i^{(r)}(u) - Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du\right) \\
&\quad +\frac{1}{n}\sum_{i\le j}\sum_{r=1}^{n_i}\int_0^{\tau}H_j^{(s)}(u,\boldsymbol{\beta})w_{12}^{ij}\Bigg[-Y_i^{(r)}(u)\mathbf{x}_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du \\
&\quad +Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\otimes\frac{S_i^{(1)}(\boldsymbol{\beta},u)}{S_i^{(0)}(\boldsymbol{\beta},u)}\hat{\lambda}_{0i}(u)du\Bigg] \\
&\xrightarrow{p} o_p(1) - \int_0^{\tau}\left[s\left(t;\boldsymbol{\beta},w_{12}^{(0)}\right)\otimes\frac{s_i^{(1)}(\boldsymbol{\beta},u)}{s_i^{(0)}(\boldsymbol{\beta},u)} - s\left(u;\boldsymbol{\beta},w_{12}^{(1)}\right)\right]\lambda_{0i}(u)du.
\end{aligned}$$

$$\begin{aligned}
\nabla_{\boldsymbol{\beta}}\hat{U}_3^{ij}(\boldsymbol{\beta},\tau) &= \frac{1}{n}\sum_{i\le j}\sum_{s=1}^{n_j}\sum_{r=1}^{n_i}\int_0^{\tau}\left[\nabla_{\boldsymbol{\beta}}H_i^{(r)}(u,\boldsymbol{\beta})\right] \\
&\quad \times w_{21}^{ij}\left(dN_j^{(s)}(u) - Y_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0j}(u)du\right) \\
&\quad +\frac{1}{n}\sum_{i\le j}\sum_{r=1}^{n_i}\int_0^{\tau}H_i^{(r)}(u,\boldsymbol{\beta})w_{21}^{ij}\Bigg[-Y_j^{(s)}(u)\mathbf{x}_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0j}(u)du \\
&\quad +Y_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\otimes\frac{S_j^{(1)}(\boldsymbol{\beta},u)}{S_j^{(0)}(\boldsymbol{\beta},u)}\hat{\lambda}_{0j}(u)du\Bigg] \\
&\xrightarrow{p} o_p(1) - \int_0^{\tau}\left[s\left(t;\boldsymbol{\beta},w_{21}^{(0)}\right)\otimes\frac{s_j^{(1)}(\boldsymbol{\beta},u)}{s_j^{(0)}(\boldsymbol{\beta},u)} - s\left(u;\boldsymbol{\beta},w_{21}^{(1)}\right)\right]\lambda_{0j}(u)du.
\end{aligned}$$

$$
\begin{aligned}
\nabla_{\beta}\hat{U}_4^{ij}(\boldsymbol{\beta},\tau) &= \frac{1}{n}\sum_{i\le j}\sum_{s=1}^{n_j}\int_0^{\tau}\left[\nabla_{\beta}H_j^{(s)}(u,\boldsymbol{\beta})\right] \\
&\cdot w_{22}^{ij}\left(dN_j^{(s)}(u) - Y_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0j}(u)du\right) \\
&+\frac{1}{n}\sum_{i\le j}\sum_{s=1}^{n_j}\int_0^{\tau}H_j^{(s)}(u,\boldsymbol{\beta})w_{22}^{ij}\Bigg[-Y_j^{(s)}\mathbf{x}_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0j}(u)du \\
&+Y_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\otimes\frac{S_j^{(1)}(\boldsymbol{\beta},u)}{S_j^{(0)}(\boldsymbol{\beta},u)}\hat{\lambda}_{0j}(u)du\Bigg] \\
&\xrightarrow{p}\; o_p(1) - \int_0^{\tau}\left[s\left(u;\boldsymbol{\beta},w_{22}^{(0)}\right)\otimes\frac{s_j^{(1)}(\boldsymbol{\beta},u)}{s_j^{(0)}(\boldsymbol{\beta},u)} - s\left(u;\boldsymbol{\beta},w_{22}^{(1)}\right)\right]\lambda_{0j}(u)du.
\end{aligned}
$$

Taking the gradient with respect to $\boldsymbol{\beta}$ of all four terms in $\hat{U}^{[ij]}(t)$, and taking their limits according to the regularity conditions, obtain that, at $\boldsymbol{\theta}_0$, the first block of $\boldsymbol{\Sigma}$, namely $\boldsymbol{\Sigma}_{11} = \left(\boldsymbol{\sigma}_{ij}\right)_{i,j=1,\ldots,k}$ with the $(i,j)^{\text{th}}$ element given by

$$
\begin{aligned}
\boldsymbol{\sigma}_{ij} &= -\int_0^{\tau}\Bigg[s\left(u;\boldsymbol{\beta},w_{11}^{(0)}\right)\otimes\frac{s_i^{(1)}(\boldsymbol{\beta},u)}{s_i^{(0)}(\boldsymbol{\beta},u)} - s\left(u;\boldsymbol{\beta},w_{12}^{(0)}\right)\otimes\frac{s_i^{(1)}(\boldsymbol{\beta},u)}{s_i^{(0)}(\boldsymbol{\beta},u)} \\
&-s\left(u;\boldsymbol{\beta},w_{11}^{(1)}\right) - s\left(u;\boldsymbol{\beta},w_{12}^{(1)}\right)\Bigg]\lambda_{0i}(u)du \\
&-\int_0^{\tau}\Bigg[s\left(u;\boldsymbol{\beta},w_{21}^{(0)}\right)\otimes\frac{s_i^{(1)}(\boldsymbol{\beta},u)}{s_i^{(0)}(\boldsymbol{\beta},u)} - s\left(u;\boldsymbol{\beta},w_{22}^{(0)}\right)\otimes\frac{s_j^{(1)}(\boldsymbol{\beta},u)}{s_j^{(0)}(\boldsymbol{\beta},u)} \\
&-s\left(u;\boldsymbol{\beta},w_{22}^{(1)}\right) - s\left(u;\boldsymbol{\beta},w_{21}^{(1)}\right)\Bigg]\lambda_{0j}(u)du.
\end{aligned}
$$

Note that $\boldsymbol{\sigma}_{ij}$ is a $p\times p$ matrix. Obviously $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, a matrix of 0. The $\boldsymbol{\Sigma}_{22}$ block is the gradient of $\mathrm{U}_2(\boldsymbol{\delta};t_1,t_2)$ with respect to $\boldsymbol{\delta}$. To see how it is derived, let $\mathbf{v}^{ij}$ be a $1\times q$ row vector and $\nabla_{\delta}\mathbf{A}^{ij}(t_1,t_2;\boldsymbol{\delta})$ defined by

$$
\mathbf{v}^{ij} = (v_1^{ij},...,v_q^{ij});\quad \nabla_{\delta}\mathbf{A}^{ij}(t_1,t_2;\boldsymbol{\delta}) = \left(\frac{\partial}{\partial\delta_1}A^{ij},...,\frac{\partial}{\partial\delta_q}A^{ij}\right) := \nabla_{\delta}\mathbf{A}^{ij}
$$

respectively. To make notation compact, for $l = 1, ..., q$, let

$$\frac{\partial}{\partial \delta_l} A^{ij}(t_1, t_2; \boldsymbol{\delta}) = \frac{\partial}{\partial \delta_l} A^{ij}((\delta)) = \nabla_{\delta_l} A^{ij}(\delta) := A^{ij}_{\delta_l}.$$

Then, the $(i, j)^{\text{th}}$ element of $\boldsymbol{\Sigma}_{22}$ is the $q \times q$ matrix given by

$$\boldsymbol{\Sigma}^{ij}_{22}(\boldsymbol{\delta}) = \mathbf{v}^{ij} \otimes \boldsymbol{\nabla}_{\delta} \mathbf{A}^{ij} = \begin{pmatrix} v^{ij}_1 A^{ij}_{\delta_1} & v^{ij}_1 A^{ij}_{\delta_2} & \cdots & v^{ij}_1 A^{ij}_{\delta_q} \\ v^{ij}_2 A^{ij}_{\delta_2} & v^{ij}_2 A^{ij}_{\delta_2} & \cdots & v^{ij}_2 A^{ij}_{\delta_q} \\ \vdots & \vdots & \vdots & \vdots \\ v^{ij}_q A^{ij}_{\delta_1} & v^{ij}_q A^{ij}_{\delta_2} & \cdots & v^{ij}_q A^{ij}_{\delta_q} \end{pmatrix}. \tag{25}$$

Recall also that

$$\nabla_{\delta_l} A^{ij}(\boldsymbol{\delta}) = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \frac{\partial}{\partial \delta_l} A^{(r,s)}_{ij}(t_1, t_2, ; \boldsymbol{\delta}).$$

So that, for example, the $(1, 1)$ component of $\boldsymbol{\Sigma}^{ij}_{22}$ is given by

$$\left[\boldsymbol{\Sigma}^{ij}_{22}\right]_{(1,1)} = v^{ij}_1 A^{ij}_{\delta_1} = v^{ij}_1 \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \frac{\partial}{\partial \delta_1} A^{(r,s)}_{ij}(t_1, t_2; \boldsymbol{\delta}).$$

The in-probability limit of $\left[\boldsymbol{\Sigma}^{ij}_{22}\right]_{(1,1)}$ is

$$\left[\boldsymbol{\Sigma}^{ij}_{22}\right]_{(1,1)} = \lim_{n\to\infty} \frac{1}{n} \sum_{i \le j} \mathbf{v}^{ij} \left[\nabla_{\delta_1} \mathbf{A}^{ij}(t_1, t_2; \boldsymbol{\delta})\right]'.$$

By virtue of the previous derivations, a compact notation for $\boldsymbol{\Sigma}_{22}$ is then

$$\boldsymbol{\Sigma}_{22} = \lim_{n\to\infty} \frac{1}{n} \sum_{i \le j} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \mathbf{v}^{ij} \otimes \boldsymbol{\nabla}_{\delta} A^{(r,s)}_{ij}(t_1, t_2; \boldsymbol{\delta}).$$

That limiting matrix is assumed to exist per Condition VII and is negative definite.

Now consider the block $\mathbf{\Sigma}_{21}$. It is easy to show that the $(i,j)^{th}$ element, $i = j = 1, ..., k$ of the gradient of $\mathrm{U}_2(t_1, t_2; \boldsymbol{\delta})$ with respect to $\boldsymbol{\beta}$ is $\mathbf{\Sigma}_{21}^{ij}$

$$\begin{aligned}
\mathbf{\Sigma}_{21}^{ij} &= \int_0^\tau \mathbf{v}_{ij} \otimes \left[\left\{ \sum_{r=1}^{n_i} H_i^{(r)}(u, \boldsymbol{\beta}) - \frac{S_i^{(1)}(\boldsymbol{\beta}, u)}{S_i^{(0)}(\boldsymbol{\beta}, u)} \right\} \right. \\
&\quad \cdot \left(\sum_{r=1}^{n_i} Y_i^{(r)}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \right) \left\{ \sum_{s=1}^{n_j} M_j^{(s)}(\mathbf{x}_j^{(s)}(u)) \right. \\
&\quad \left. \left. + \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} A_{ij}^{(r,s)}(du_1, u_2; \rho(\boldsymbol{\delta})) \right\} \right] \hat{\lambda}_{0i}(u) du \\
&\quad + \int_0^\tau \mathbf{v}_{ij} \otimes \left[\sum_{s=1}^{n_j} H_j^{(s)}(u, \boldsymbol{\beta}) - \left\{ \frac{S_j^{(1)}(\boldsymbol{\beta}, u)}{S_j^{(0)}(\boldsymbol{\beta}, u)} \right\} \right. \\
&\quad \cdot \sum_{s=1}^{n_j} Y_j^{(s)}(u) e^{\boldsymbol{\beta}' \mathbf{x}_j^{(s)}(u)} \left\{ \sum_{r=1}^{n_i} M_i^{(r)}(\mathbf{x}_i^{(r)}(u)) \right. \\
&\quad \left. \left. + \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} A_{ij}^{(r,s)}(u_1, du_2; \rho(\boldsymbol{\delta})) \right\} \right] \hat{\lambda}_{0j}(u) du.
\end{aligned}$$

Note that $\mathbf{\Sigma}_{21}^{ij}$ is a $p \times p$ matrix and the in-probability limit. Assuming the integration operation is interchangeable and limit is given by

$$\mathcal{E}(\mathbf{\Sigma}_{21}^{ij}) = \lim_{n \to \infty} \frac{1}{n} \sum_{(i,j), i \leq j} \mathbf{\Sigma}_{21}^{ij}.$$

Hence, the partial derivative matrix converges to a matrix $\mathbf{\Sigma}$ which is negative definite at the true parameter value $(\boldsymbol{\beta}_0, \boldsymbol{\delta}_0)$. It then follows from the inverse function theorem of Foutz (1976) that there exists a unique sequence $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}})$ such that $\widehat{\mathbf{V}}(t; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}) = \mathbf{0}$ and $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}) \xrightarrow{p} (\boldsymbol{\beta}_0, \boldsymbol{\delta}_0)$ as $n \to \infty$.

The next theorem is on the asymptotic normality of $(\hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\delta}}_n)$ when properly standardized.

Theorem 5 *Under regularity Conditions I and VIII,*

$$\sqrt{n}\{(\hat{\boldsymbol{\beta}}_n{}', \hat{\boldsymbol{\delta}}_n{}')' - (\boldsymbol{\beta}_0', \boldsymbol{\delta}_0')'\} \xrightarrow{d} N_{p+q}(\mathbf{0}_{p+q}, \Phi),$$

where $\mathbf{\Phi}$ *is a* $(p+q)\times(p+q)$ *matrix given by* $\mathbf{\Phi} = \mathbf{\Sigma}^{-1}\mathbf{\Omega}\mathbf{\Sigma}^{-1}$.

Proof: The central limit theorem is applied for random field given in Remarks (3), page 112 of Guyon (1995). Taylor expansion of $\mathrm{V}(t,\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ yields

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \left[\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{V}(t,\boldsymbol{\theta}) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right]^{-1} \sqrt{n}\mathbf{V}(t,\boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}^*$ is between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$, and $\boldsymbol{\theta}^* \xrightarrow{p} \boldsymbol{\beta}_0$ under the infill asymptotic domain setting. Furthermore, note that

$$\left[\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{V}(t,\boldsymbol{\theta}) \mid_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right] \xrightarrow{p} \mathbf{\Sigma}$$

as $n \longrightarrow \infty$. The $\mathbf{\Omega}$ matrix in $\mathbf{\Phi}$ is the variance of the score vector, which under Conditions V and VI is assumed to exist and converges to a positive definite matrix. The expression of $\mathbf{\Phi}$ is obtained by applying the result of multivariate central limit theorem. Finally, the theorem follows upon applying Remark(3), page 112 of Guyon (1995).

6. NUMERICAL ASSESSMENT AND APPLICATION

This section discusses how the simulation study was done together with the illustrative application.

6.1. NUMERICAL ASSESSMENT

This begins with the selection of the different regions that will be used. The package *raster* on Geographic Data Analysis and Modeling contains the geographical coordinates of many countries. The United States was used the country.

6.1.1. Regions. The *raster* package in R contains the data on the geographical coordinates of well defined subdivisions in many countries. This package was used to get the coordinates for states, counties etc... for the United States. Depending on the country, this package also allows users to select location data with several levels of depth. For the United States, the users can specify either Level 1 for statewise locations or Level 2 for the county wise locations. We use Level 2 data from raster for the simulations. The geographical centers of the 3117 contiguous counties, excluding Hawaii and Alaska are our $\mathbf{l}_i$, $i = 1, ..., 3117$. The other alternative is to choose a state and randomly select counties within the selected state. In Figure 2, the state of Missouri is provided as an example with the coordinates for a couple of the counties. For example, the longitude and latitude of the center of Newton county in the state of Missouri is $(-94.34, 36.91)$.

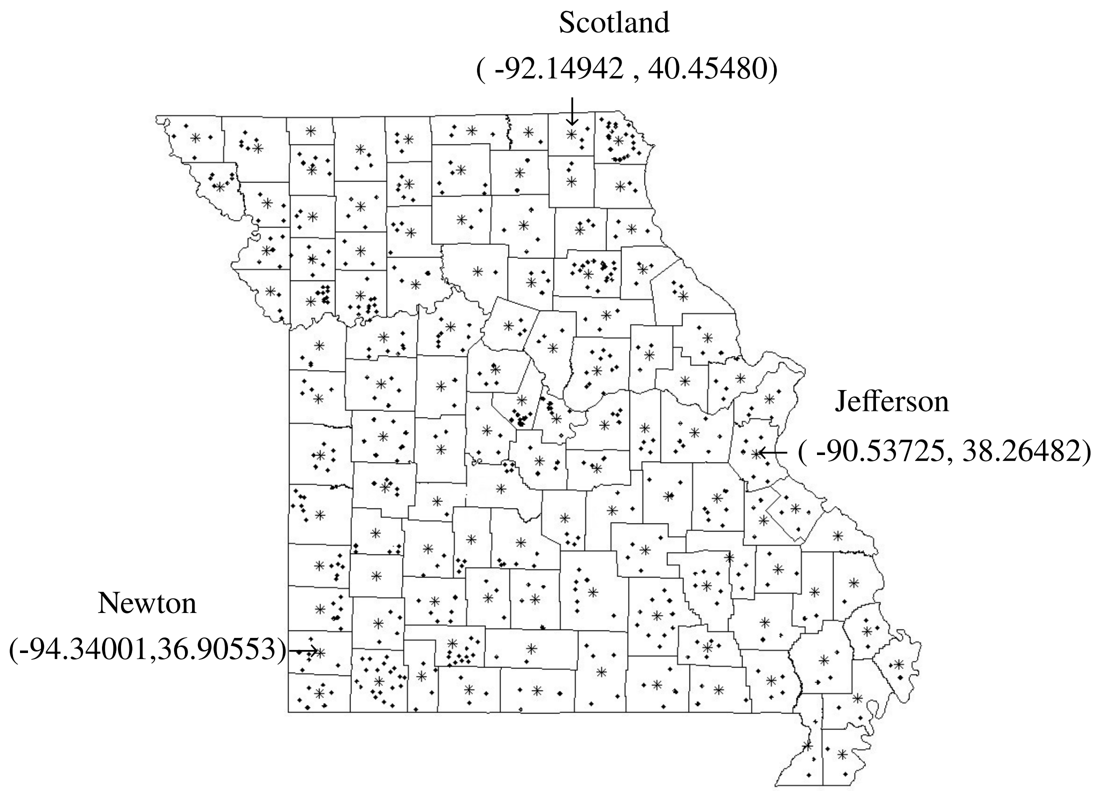

Figure 2. Missouri counties map

6.1.2. Simulation Design. A random sample of $\{n_i : i = 1, .., 3117\}$ people from each county was obtained where n_i is proportional to the county population from the latest census available in R while making sure $n = \sum_{i=1}^{3117} n_i \in \{93510, 155850, 311700\}$. Two covariates were considered $\mathbf{x} = (x_1, x_2)$, where x_1 follows the binomial distribution with parameters $n = n_i$ and $p = 0.5$ and $x_2 \sim \mathrm{N}(0, 0.5)$ resulting in a mixture of categorical and quantitative covariates. The spatial correlation parameter was set at $\boldsymbol{\delta} = (\delta_1, \delta_2) = (\text{range}, \text{sill}) = (0.5, 1.5)$. The regression coefficient vector in the Cox model is $\boldsymbol{\beta} = (\beta_1, \beta_2) = (1, 2)$. As for the proportion of censored observation, less censoring to severe censoring is allowed in order to assess its impact on the spatial correlation. The proportion of censored units was taken to be in $\{5\%, 10\%, 20\%, 25\%\}$, allowing for mild to severe censoring. For the baseline hazard, the Weibull hazard was used given by $\lambda_0(t) = \theta_1\theta_2(\theta_1 t)^{\theta_2 - 1}$. We set $\theta_1 = 1$ since it is the scale parameter and is irrelevant in our simulation. However, the shape parameter θ_2 was taken in $\{0.8, 1.5\}$ to allow for increasing failure over time for $\theta_2 = 2$ and decreasing for $\theta_2 = 0.8$.

6.1.3. Event Times Generation. Under the Cox model with Weibull baseline hazard, failure times were generated via the probit transformation using the following steps:
(i) If $\Phi(\cdot)$ denotes the probit transformation, then solving for the Cox's model, we obtain for $(i, r) \in \mathcal{L} \times \mathcal{L}_i$

$$\Lambda_{0i}(T_i^{(r)}) = -\ln\left(1 - \Phi(\tilde{T}_i^{(r)})\right)\exp[-\boldsymbol{\beta}\mathbf{x}_i^{(r)}(t)].$$

Solving for $T_i^{(r)}$, we obtain

$$T_i^{(r)} = \Lambda_{0i}^{-1}\left[-\ln\left(1 - \Phi(\tilde{T}_i^{(r)})\right)\exp(-\boldsymbol{\beta}\mathbf{x}_i^{(r)}(t))\right],$$

where $\Lambda_{0i}^{-1}(\cdot)$ is the inverse of the Weibull cumulative hazard given by $\Lambda_0^{-1}(t) = t^{\frac{1}{\theta_2}}$.
(ii) For $(i, r) \in \mathcal{L} \times \mathcal{L}_i$, the $T_i^{(r)}$s are generated using the expression

$$T_i^{(r)} = \left[-\ln\left(1 - \Phi(\tilde{T}_i^{(r)})\right)\exp(-\boldsymbol{\beta}\mathbf{x}_i^{(r)}(t))\right]^{\frac{1}{\theta_1}},$$

where $\Phi(\tilde{T}_i^{(r)}) \sim \mathrm{U}(0, 1)$.

6.1.4. Simulated Data. For the purpose of estimating parameters, the study considers two spatial correlation models, namely the exponential and Gaussian model as given in (4), (5) and powered spatial correlation function. For all models, 500 simulation replications were performed with each parameter specification and sample size combination. The results are given in Table 1, Table 2, Table 3, and Table 4. CP stands for censoring percentage.

6.1.5. Comments on The Simulation Results. The results of the simulation study indicate that the estimators of the spatial correlation $\boldsymbol{\delta}$ as well as regression coefficients $\boldsymbol{\beta}$ perform well. One thing to note here is that as the percentage of censoring increases, the biases of the $\boldsymbol{\beta}$ increase regardless of the sample size, whereas the biases of the $\boldsymbol{\delta}$ remain very steady close to each other. This makes sense since the spatial correlation parameters is the correlation between two areas so it is not affected by large samples. However, the bias of $\boldsymbol{\beta}$ will increase because higher censoring translates into less failure times. There is no significant difference in the results between the exponential and Gaussian spatial correlation models. The reason why this is so is both have exponential components so the impact of the large sample will be minor. However, the standard deviations of the estimates of $\boldsymbol{\delta}$ remain without any noticeable pattern with the increasing sample size.

6.2. ILLUSTRATIVE APPLICATION

The foregoing procedures are applied to the Leukemia survival data which was also analyzed in Henderson et al. (2002). The data contains $1,043$ cases of *Acute Myeloid Leukemia* (AML) which were recorded between 1982 and 1998 at 24 administrative districts.

It contains the time $T_i^{(r)}$ for each unit $(i, r) \in \mathcal{L} \times \mathcal{L}_i$, and the censoring indicator $\delta_i^{(r)}$. There were 16% of censored observations. Four covariates were available, that is $\mathbf{x} = (\text{age}, \text{gender}, \text{wbc}, \text{tpi})$, where *wbc* stands for white blood cell count and *tpi* for

Table 1. Numerical results for the Matérn exponential correlation function when $\alpha_2 = 1.5$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.9815984	0.13508195	1.959529	0.19624070	0.4781593	0.10571357	1.490568	0.13263675
	155850	0.9673390	0.05540476	1.927976	0.09517489	0.5249068	0.12352137	1.430255	0.09321591
	311700	0.9698712	0.04587603	1.934979	0.09038914	0.4993314	0.11028005	1.519357	0.10024713
10%	93510	0.9732147	0.13088504	1.940052	0.17449313	0.4847504	0.10627189	1.493118	0.12062017
	155850	0.9667443	0.06144499	1.930267	0.10488670	0.5144859	0.11971700	1.480974	0.11968495
	311700	0.9652277	0.04776958	1.926827	0.09353791	0.5002486	0.12825714	1.388380	0.06354968
20%	93510	0.9697460	0.13133686	1.944975	0.18080929	0.5100953	0.09477460	1.442438	0.08088030
	155850	0.9571049	0.05671763	1.915559	0.10213817	0.4621895	0.09403900	1.531357	0.11815007
	311700	0.9601825	0.04677509	1.913597	0.09314964	0.4843984	0.09691497	1.513268	0.09512435
25%	93510	0.9630415	0.14436885	1.933648	0.18969984	0.5441557	0.10871051	1.524998	0.10970698
	155850	0.9527664	0.05572005	1.902975	0.09731097	0.4634573	0.10042738	1.497794	0.14067757
	311700	0.9508088	0.04542669	1.896139	0.09019429	0.5446168	0.11982288	1.503245	0.11052694

Table 2. Numerical results for the Matérn exponential correlation function when $\alpha_2 = 0.8$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.9783978	0.12970782	1.946787	0.18660250	0.4882348	0.11443885	1.474876	0.09779100
	155850	0.9710274	0.05750256	1.934725	0.09835010	0.4807230	0.11170969	1.481125	0.11471019
	311700	0.9636052	0.04559769	1.922377	0.08916006	0.4655450	0.09277994	1.466099	0.10876257
10%	93510	0.9769663	0.13856558	1.949619	0.18304773	0.4462199	0.11451939	1.438839	0.12150385
	155850	0.9617960	0.05863333	1.917835	0.10512117	0.4269675	0.11660109	1.454666	0.10952768
	311700	0.9638223	0.04267380	1.922667	0.08319622	0.4667544	0.12044353	1.400070	0.10626081
0%	93510	0.9756672	0.13536248	1.939934	0.19446770	0.4463510	0.10309318	1.501104	0.11256934
	155850	0.9606159	0.05433496	1.919771	0.09995991	0.4263876	0.10085830	1.547133	0.10169458
	311700	0.9576884	0.04532230	1.909236	0.08981518	0.4526783	0.10419536	1.529238	0.08352670
25%	93510	0.9598803	0.13346690	1.933900	0.18535909	0.4755711	0.11954677	1.497023	0.09135412
	155850	0.9528170	0.05873037	1.904862	0.10291980	0.4849959	0.11653679	1.494860	0.09271720
	311700	0.9526238	0.04602066	1.900388	0.09120538	0.5250469	0.10591207	1.439092	0.07686285

Table 3. Numerical results for the powered correlation function when $\alpha_2 = 1.5$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.9680262	0.13163930	1.957335	0.18880277	0.4499527	0.0906412	1.501031	0.10911225
	155850	0.9690669	0.05277084	1.940764	0.09057505	0.4717423	0.1037612	1.476357	0.09088143
	311700	0.9656089	0.04355927	1.927449	0.08502461	0.5027827	0.1201323	1.494042	0.08246186
10%	93510	0.9755093	0.12883476	1.935020	0.18185705	0.5051200	0.1165951	1.450969	0.09633239
	155850	0.9632191	0.05576547	1.923382	0.09946338	0.4594674	0.1298239	1.538667	0.12832220
	311700	0.9610232	0.04466272	1.916430	0.08953072	0.4346925	0.1044191	1.457736	0.11169946
20%	93510	0.9741438	0.13358569	1.943066	0.19030131	0.4873341	0.1211949	1.514300	0.15230805
	155850	0.9601557	0.05128941	1.916523	0.09157145	0.4573149	0.1040705	1.573742	0.09072474
	311700	0.9577343	0.04470458	1.909920	0.08776440	0.4691986	0.1214077	1.521292	0.11410709
25%	93510	0.9677951	0.14210420	1.930109	0.17420695	0.4898497	0.1127624	1.457671	0.12063205
	155850	0.9558041	0.05761884	1.908774	0.10170949	0.4882707	0.1081157	1.524964	0.10627590
	311700	0.9530145	0.04792042	1.899907	0.09319083	0.4928822	0.1215053	1.430373	0.09325126

Table 4. Numerical results for the powered correlation function when $\alpha_2 = 0.8$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.9803209	0.13087042	1.969887	0.17617333	0.4779905	0.1184450	1.547133	0.10169458
	155850	0.9703777	0.05345297	1.934411	0.09736861	0.5304549	0.1027453	1.513268	0.09512435
	311700	0.9643776	0.04274039	1.924133	0.08413270	0.4832991	0.1053006	1.520087	0.11196502
10%	93510	0.9784334	0.13113013	1.949669	0.17125655	0.4582216	0.1318242	1.441526	0.09739045
	155850	0.9625114	0.05393592	1.922684	0.09007249	0.5070480	0.1127849	1.424603	0.08090544
	311700	0.9626708	0.04661519	1.921852	0.09134828	0.4313484	0.1125438	1.493118	0.12062017
20%	93510	0.9759235	0.14057294	1.940354	0.18465727	0.5333254	0.1002577	1.511277	0.11847563
	155850	0.9625727	0.05394279	1.915289	0.09652738	0.4789544	0.1132957	1.513660	0.12258979
	311700	0.9593511	0.04602093	1.913568	0.09044250	0.4652688	0.1211030	1.443540	0.11838361
25%	93510	0.9651831	0.13170463	1.915941	0.17906719	0.4301427	0.1213539	1.495397	0.05561716
	155850	0.9506967	0.05534870	1.901424	0.09777947	0.4588298	0.1209582	1.536580	0.13753493
	311700	0.9526172	0.04653209	1.899912	0.09131938	0.4939713	0.1110672	1.523255	0.10580306

Townsend score. The Townsend score is a qualitative value in $[-7, 10]$ describing quality of life in a given area. High values indicate less affluent areas. The factors affecting survival were investigated while accounting for spatial correlation.

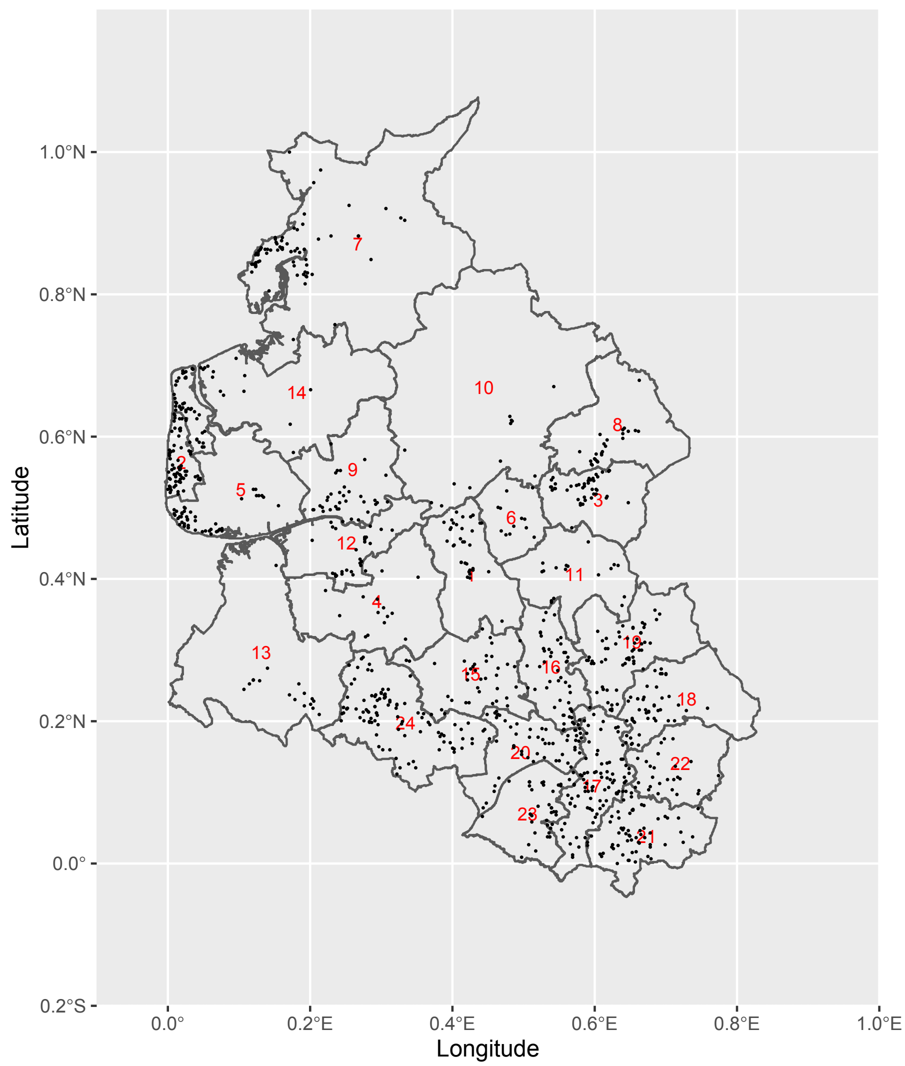

Figure 3. Leukemia data with the 24 districts

Figure 3 shows residential locations of the AML cases during the observation window. Henderson et al. (2002) investigated whether the survival distribution in AML in adults is homogeneous across the region after allowing for known risk factors. In their manuscript, they employ a multivariate frailty that incorporates the effects of covariates,

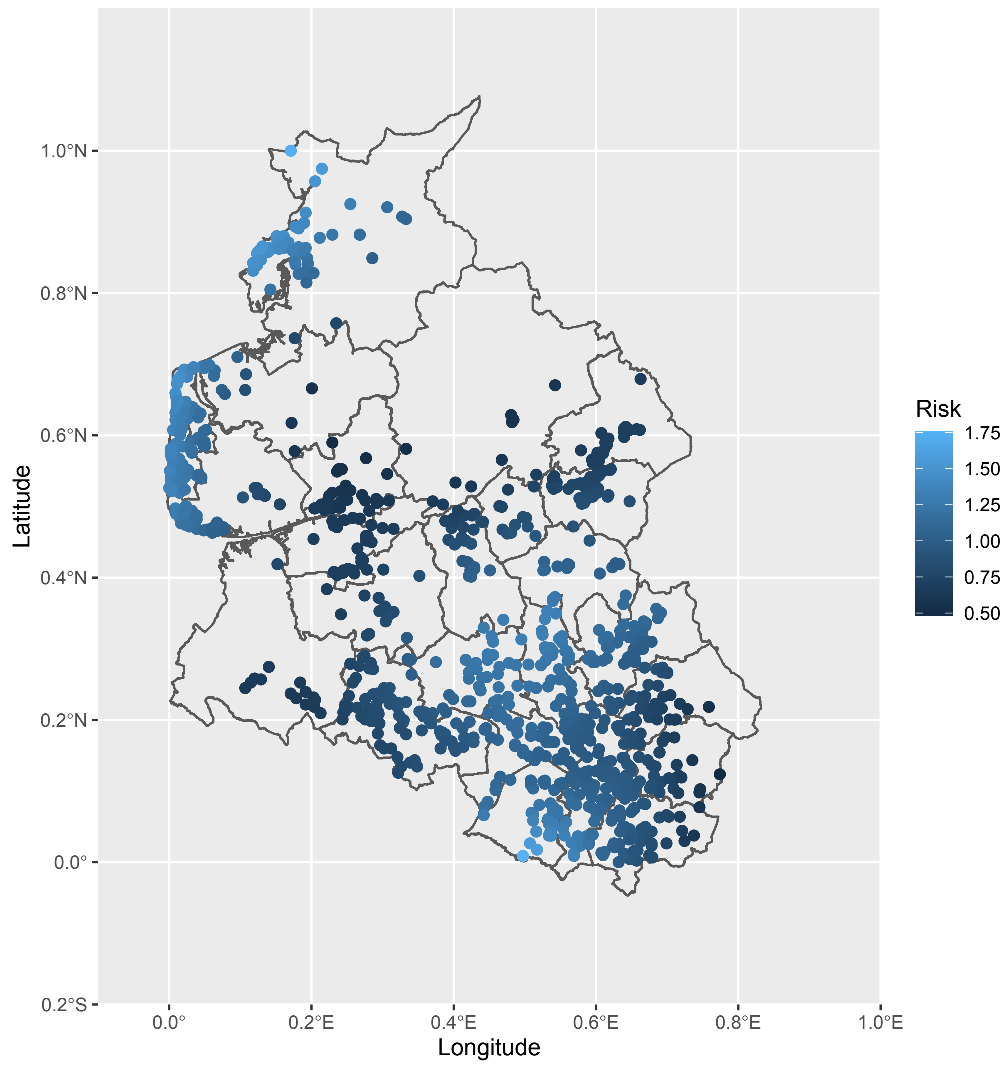

Figure 4. Leukemia data with the 24 districts, units, and risk

individual heterogeneity, and spatial traits. This study's approach and theirs are different. Whereas both use the Cox model as the instantaneous failure rate, their approach in studying spatial variation is done via the use of conditional frailty, where the conditioning random variable for all 24 districts is the vector of mean frailty $\boldsymbol{\mu} = (\mu_1, ..., \mu_{24})$. Specifically, if Z_r is the frailty for unit r in location $\mathbf{I}_j$, and μ_j the mean frailty of all individuals in that location, they postulate that

$$Z_r|\mu_j \sim \Gamma(\xi^{-1}, (\xi\mu_j)^{-1}),$$

with a N$(1,\Xi)$ on μ_j where Ξ measure the spatial variation between districts. Whereas they use a conditional frailty model with variance covariance matrix that is a function of the distance between regions, we embed the spatial correlation in the transformed failure times giving us a multivariate Gaussian random field with variance covariance that is a function of the distance between regions via the Matérn spatial correlation function.

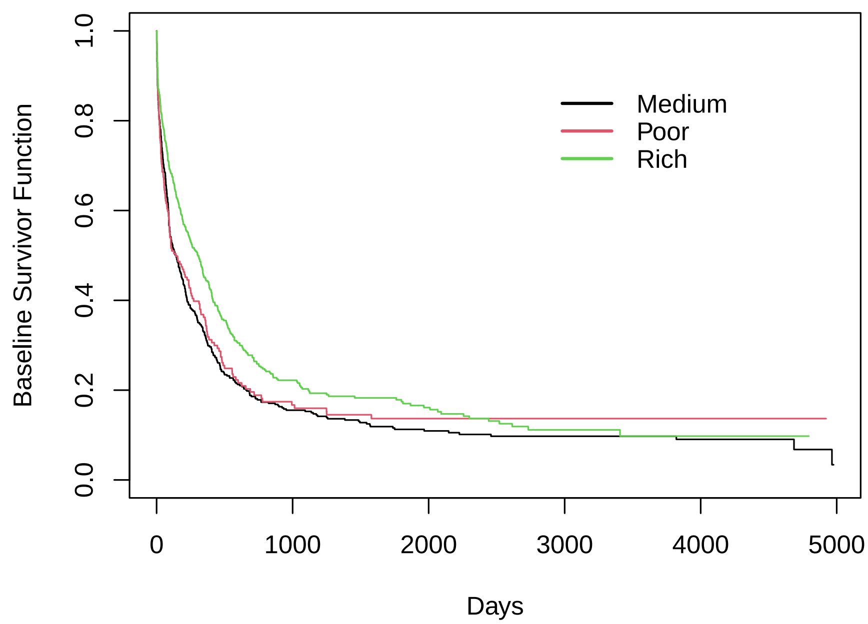

Figure 5. Survival curve of area categories

Before applying our methods, a set of initial data analysis was run. Figure 5 shows the Kaplan-Meier plots by gender. It is clear that survival curve for the female group lies above that of male group. This concurs with the summary statistics in Table5. The variable *tpi* represents the Townsend score. The higher values for *tpi* indicates less affluent areas. All individuals in the study were grouped into 3 categories based on *tpi*. If *tpi* of a person is lower than -1.5, he or she is categorized into *Rich* group. Likewise, if *tpi* of a person falls between -1.5 and -4.5, the person is grouped into *Medium* category. Lastly, if a persons *tpi* is greater than 4.5, that person is categorized into *Poor* group. Figure 5 presents the

Table 5. Summary statistics of Leukemia data by gender

		Female	Male
Survival time	Min	1.0	1.0
	Q1	37.0	45.5
	Q2	182.0	186.0
	Mean	581.5	489.0
	Q3	574.8	490.5
	Max	4922.0	4977.0
Age	Min	14.00	14.0
	Q1	48.00	50.0
	Q2	65.00	65.0
	Mean	60.98	60.5
	Q3	75.00	74.0
	Max	92.00	92.0
wbc	Min	0.00	0.00
	Q1	1.80	1.70
	Q2	7.35	8.10
	Mean	40.42	36.94
	Q3	36.60	41.10
	Max	500.00	500.00

survival curves according to these three areas. From near day 100 to 5000 survival curve of Medium group always lies below than survival curves of other two groups. Moreover, comparing data of Poor and Rich groups, from day 0 to near day 2400, survival curve for Rich group is always above the Poor group. But, interestingly from near day 2400 to 5000 survival curve for poor group is above that of Rich group.

These methods are used to analyze the Leukemia data. Factors that may increase the risk of acute myeloid leukemia include age, gender, prior cancer treatment, environmental factors, blood disorder, genetic disorder, to name a few. Only covariates in the data, were considered assuming that age at onset of acute myeloid leukemia (AML) on adults follows the Cox model. The Matérn model was used to account for the spatial dependence between pair of districts. The hazard function for an r^{th} unit in district i is given by

$$\lambda_i^{(r)}(t) = \lambda_{0i}(t) \exp[\beta_{age} \times \text{age} + \beta_{sex} \times \text{sex} + \beta_{wpc} \times \text{wpc}]$$

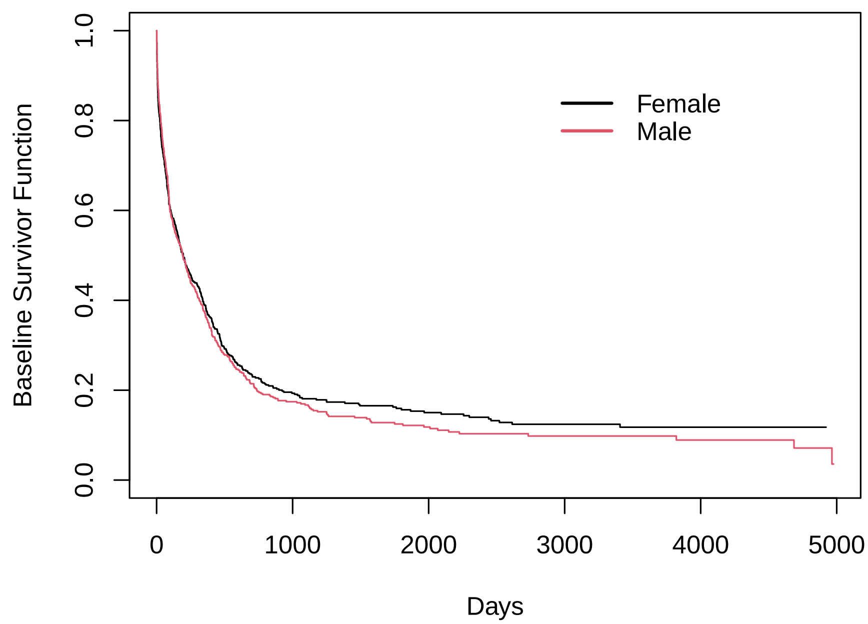

Figure 6. Survival curve of gender

The estimated regression coefficients and the spatial correlation parameters using the estimating functions in Section 5 were obtained. And also the associated standard deviations and confidence intervals were calculated. The results are presented in Table 6. The results of Henderson's approach are also presented in Table 7. The results in both tables show in both models that all regression coefficient are significant concurring with the fact that all the covariates age, sex, and wpc increase the risk of aml. That is more so with our models, and the results concur with our preliminary analysis of the data. The estimated value of the range, which is 1.2418 indicates that the impact of environment vanishes when two units are separated by at least 1.2418 units of distance. The log pseudo marginal likelihoods (LPML) for each model is also given. Despite the fact that this model has more parameters, it has a better LPML, indicating it is the best model between the two. However, this needs to be taken with cautious and deep investigation such as each unit personal geographical location would be needed to arrive at the best model in this situation.

Table 6. Summary of means and standard deviations of regression parameters and spatial parameters for Leukemia data using our models. $LPML = -5991.082$

	Mean	Median	Std.Dev.	95%CI-Low	95% CI-Upp
age	30.65×10^{-3}	27.95×10^{-3}	3.70×10^{-3}	27.95×10^{-3}	34.69×10^{-3}
sex	70.40×10^{-3}	70.61×10^{-3}	0.30×10^{-3}	70.07×10^{-3}	70.61×10^{-3}
wbc	3.00×10^{-3}	3.03×10^{-3}	0.04×10^{-3}	2.95×10^{-3}	3.03×10^{-3}
tpi	34.30×10^{-3}	33.77×10^{-3}	0.72×10^{-3}	33.77×10^{-3}	35.09×10^{-3}
sill	891.25×10^{-3}	913.18×10^{-3}	48.35×10^{-3}	815.88×10^{-3}	913.18×10^{-3}
range	1241.79×10^{-3}	1201.53×10^{-3}	90.02×10^{-3}	1201.53×10^{-3}	1382.70×10^{-3}

Table 7. Summary of means and standard deviations of regression parameters and spatial parameters for Leukemia data using Henderson's model: $LPML = -5925.385$

	Mean	Median	Std.Dev.	95%CI-Low	95% CI-Upp
age	51.95×10^{-3}	52.00×10^{-3}	3.35×10^{-3}	45.07×10^{-3}	58.46×10^{-3}
sex	108.04×10^{-3}	105.01×10^{-3}	108.38×10^{-3}	-101.16×10^{-3}	325.87×10^{-3}
wbc	5.94×10^{-3}	5.94×10^{-3}	0.79×10^{-3}	4.39×10^{-3}	7.53×10^{-3}
tpi	61.37×10^{-3}	61.24×10^{-3}	15.46×10^{-3}	33.07×10^{-3}	93.25×10^{-3}
fv	64.22×10^{-3}	40.42×10^{-3}	80.68×10^{-3}	1.01×10^{-3}	252.93×10^{-3}

7. CONCLUDING REMARKS

The situation where many units clustered in different geographical areas described by their longitude and latitude are monitored for the occurrence of some event. A methodology was developed using a combination of modern survival analysis and geostatistical formulation. Parameters of these models are estimated using unbiased estimating functions and their large sample properties were also examined using infill asymptotic approach that one encounters with spatial data. The methodology can be easily generalized to the case of recurrent events. Another generalization is to consider the geographical coordinate of each unit within a given geographical area. It is also possible to consider both within and between areas spatial correlation. Another important area of interest is to develop models that account for correlation between event time via frailty when the event is allowed to recur. Another possible future direction is using another model for modeling connection between failure covariates and failure times such as the accelerated failure time model. However, other estimating approaches, such as rank-based would need to be applied to the transformed event times.

REFERENCES

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120.

Banerjee, S. and Carlin, B. (2003). Semi-parametric spatiotemporal analysis. *Environmetrics*, 14:523–535.

Banerjee, S. and Dey, D. K. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Anal.*, 11(2):175–191.

Banerjee, S., Wall, M., and Carlin, B. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in minnesota. *Biostatistics*, 4(1):123–142.

Bronnenberg, B. J. (2005). Spatial models in marketing research and practice. *Appl. Stoch. Models Bus. Ind.*, 21(4-5):335–343.

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(1):141–151.

Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

Diva, U., Banerjee, S., and Dey, D. K. (2007). Modelling spatially correlated survival data for individuals with multiple cancers. *Stat. Model.*, 7(2):191–213.

Diva, U., Dey, D. K., and Banerjee, S. (2008). Parametric models for spatially correlated survival data for individuals with multiple cancers. *Stat. Med.*, 27(12):2127–2144.

Engen, S. (2007). Stochastic growth and extinction in a spatial geometric Brownian population model with migration and correlated noise. *Math. Biosci.*, 209(1):240–255.

Foutz, R. V. (1976). On the consistency of locally asymptotically most stringent tests. *Canad. J. Statist.*, 4(2):211–219.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.*, 105(491):1167–1177.

Goldstein, H. (1995). *Multilevel Statistcal Models*. New York. Edward Arnold.

Gorst (1995). North west leukaemia register. *Technical report*.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics. XLIX. On the Matérn correlation family. *Biometrika*, 93(4):989–995.

Guyon, X. (1995). *Random fields on a network*. Probability and its Applications (New York). Springer-Verlag, New York. Modeling, statistics, and applications, Translated from the 1992 French original by Carenne Ludeña.

Handcock, M. S. and Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, 35(4):403–410.

Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *J. Amer. Statist. Assoc.*, 93(443):1099–1111.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *J. Amer. Statist. Assoc.*, 97(460):965–972.

Hjort, N. L. and Omre, H. (1994). Topics in spatial statistics. *Scand. J. Statist.*, 21(4):289–357. With discussion and a reply by the authors.

Hunt, F. Y. (1978). *GENETIC AND SPATIAL VARIATION IN SOME SELECTION-MIGRATION MODELS*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–New York University.

Jacod, J. (1976). Un théorème de représentation pour les martingales discontinues. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(3):225–244.

Lawless, J. F. and Yilmaz, Y. E. (2011). Semiparametric estimation in copula models for bivariate sequential survival times. *Biom. J.*, 53(5):779–796.

Lele, S. and Taper, M. L. (2002). A composite likelihood approach to (co)variance components estimation. *J. Statist. Plann. Inference*, 103(1-2):117–135. C. R. Rao 80th birthday felicitation volume, Part I.

Li, Y. and Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *J. Amer. Statist. Assoc.*, 101(474):591–603.

Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics*, 58(2):287–297.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Stat. Med.*, 13:2233–2247.

Lindsay, B. G. (1988). Composite likelihood methods. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)*, volume 80 of *Contemp. Math.*, pages 221–239. Amer. Math. Soc., Providence, RI.

Lindsay, B. G., Yi, G. Y., and Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statist. Sinica*, 21(1):71–105.

Matérn, B. (1986). *Spatial variation*, volume 36 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, second edition. With a Swedish summary.

Matheron, G. (1962). *Traité de gostatistique appliquée*. Statistics for Biology and Health. Editions Technip, France.

Paik, J. and Ying, Z. (2012). A composite likelihood approach for spatially correlated survival data. *Comput. Statist. Data Anal.*, 56(1):209–216.

Pan, C., Cai, B., Wang, L., and Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data. *Comput. Statist. Data Anal.*, 74:198–208.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, 79(3):495–512.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.*, 11(2):453–466.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.*, 93(443):1164–1175.

Stein, M. L. (1999). *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York. Some theory for Kriging.

Varin, C., Hø st, G., and Skare, O. i. (2005). Pairwise likelihood inference in spatial generalized linear mixed models. *Comput. Statist. Data Anal.*, 49(4):1173–1191.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica*, 21(1):5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.*, 84(408):1065–1073.

Yilmaz, Y. E. and Lawless, J. F. (2011). Likelihood ratio procedures and tests of fit in parametric and semiparametric copula models with censored data. *Lifetime Data Anal.*, 17(3):386–408.

II. MODELING SPATIALLY CORRELATED SURVIVAL DATA WITH FRAILTY

Fathima Zahra Sainul Abdeen
Department of Mathematics & Statistics
Missouri University of Science and Technology
Rolla, Missouri 65409–0050
Email: fs2y7@mst.edu

Akim Adekpedjou
Department of Mathematics & Statistics
Missouri University of Science and Technology
Rolla, Missouri 65409–0050
Email: akima@mst.edu

ABSTRACT

We consider the same setting, namely a fixed number of clustered areas identified by their geographical coordinate that are monitored for the occurrences of an event such as pandemic, epidemic, migration to name a few. Data collected on units at all areas include time varying covariates and environmental factors. We allow for association between event times in every area using an unobservable frailty. The frailty are assumed to be the same per area, and are independent. The collected data is again considered pairwise to account for spatial correlation between all pair of areas, and their frailty unobservables Z_i and Z_j. The pairwise right censored data is again probit-transformed yielding a multivariate Gaussian random field given the values of the frailties. We provide a detailed small sample numerical studies and we show that ignoring correlation between unit in a given area leads to biased estimators.

Keywords: Spatial correlation; Gaussian random fields; Composite likelihood; Estimating function; Frailty; Mixing; Clustered failure times

1. INTRODUCTION

Spatially correlated data arise in various crucial fields such as ecology, clinical trials, and epidemiology to name a few. Therefore, developing statistical models that are capable of accounting for spatial correlation is of utmost importance.

In addition to possible covariates, it may be possible for a survival model to be affected by unobserved random factors called *frailties* via its hazard function. Ignoring so called frailties and solely depending on covariates in modeling survival times can have consequences including ending up having unreliable parameter estimates. Therefore, it is of significant importance to account for frailty variables.

Spatial statistical methods have been described in detail by Cressie (1993) and statistical tools needed for modeling normally distributed data have been developed. However, modeling spatially correlated survival data with frailties has not been considered by Cressie (1993). Sudipto, Banerjee et al. (2003) and Bradley (2005) discussed a few hierarchical methods for modeling survival data which are spatially correlated. They considered spatially arranged clusters according to their frailties and used two different approaches called "geostatistical" and "lattice" to model survival times taking spatially correlated hazards into account. However, the covariance structure they obtained in their study did not correspond to a proper covariance structure. Kosorok et al. (2004) considered a class of semiparametric regression models that are one parameter extension of Cox model. They performed non-parametric likelihood based inferences while assuming hazard given the covariates and random frailty has proportional hazard form multiplied by the corresponding frailty. However, they did not consider spatial dependence of survival times. Petersen (1998) used an additive frailty model for modeling correlated survival time. Even though he did not take spatial correlation as the specific correlation in his study, the frailty model he proposed and the corresponding estimation methods he derived give an easy and flexible approach to model multivariate correlated survival times by clearly distinguishing between dependence parameters and regression parameters with baseline hazards. However, they restricted associations between individuals to be always positive which might not be the case in applications. Moreover, they did not consider how the effect of the choice of frailty distribution can be problematic in modeling. Li and Lin (2006) proposed a new class of semiparametric normal transformation models for right censored spatially correlated survival data. Their model is a flexible one that provides a semiparametric likelihood approach to generate censored spatial survival data that have a spatial correlation structure which allows individual observations to marginally follow the Cox proportional hazard model. However, their model does not account for frailties associated with survival times.

Li and Lin (2006) considered right-censored spatially correlated survival data and performed semi parametric inferences. They assumed that each of the clusters in their study had only one subject. In addition, they did not consider frailties. However, in real world applications, one observation per cluster is a quite rare case and also, unobservable random factors are often thought to interfere with the observations. Therefore, in our study, we extend his idea to a general setting so that, our model is capable of handling many observations per cluster as well as unobservable random factors.

The rest of this part of the dissertation is structured as follows. In Section 2, a semiparametric model is introduced with a normal transformation that can handle spatially correlated survival data including unobservable random factors. Section 3 describes developing estimating equations, that are spatial semiparametric, for spatial correlation parameters and regression coefficients given the corresponding frailties. In Section 4, a simulation study was performed using R software package to assess the performance of the proposed method when finite samples are taken into account.

2. SPATIALLY CORRELATED SURVIVAL DATA WITH FRAILTY

This section gives a discussion on notation, pairwise survival data, Cox model used, Multivariate Gaussian Random fields and spatial correlation models.

2.1. NOTATION

The previous chapter and the current one have similar notations except the inclusion of the unobserved frailty.

In addition, let $\mathbf{Z} = (Z_1, Z_2, ..., Z_k)$ be a k-dimensional vector of independently and identically distributed positive random variables called *frailties* which are unobservable random factors affecting the event occurrences of the subjects at each geographical location. We also assume that each subject in the same geographical location has the same frailty. Therefore, the frailty shared by all subjects in location i is denoted by Z_i.

The censoring times $C_i^{(r)}$ and true survival times $W_i^{(r)}$ are assumed independent of each other, given the independent frailties Z_i and covariates $\mathbf{x}_i^{(r)}(t)$.

The observable data per area given frailty is

$$\mathbf{O}(\mathrm{I}_i|Z_i) = \{\mathbf{x}_i^{(r)}(t), T_i^{(r)}, \delta_i^{(r)}, Z_i\}$$

2.2. THE MODEL

According to our model, hazard function of true survival time $T_i^{(r)}$ is assumed to follow the following shared frailty model given below in (1) marginally, where $\lambda_{0i}(t)$ is the baseline hazard function for the i^{th} geographical location with different baseline per location, but same regression parameter $\boldsymbol{\beta}$ for all locations. As we stated before, it is assumed that all the individuals in a given geographical location i have the same frailty Z_i. Hence $\lambda_i^{(r)}(\cdot)$ has the form

$$\lambda_i^{(r)}(t|Z_i) = Z_i \lambda_{0i}(t) \exp\left(\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(t)\right). \tag{1}$$

In (1), $\boldsymbol{\beta}$ is a p-dimensional regression parameters vector and $\boldsymbol{\beta}'$ stands for its transpose. For $i = 1, ..., k$, the baseline hazard functions $\lambda_{0i}(t)$ are unspecified and need to be estimated.

2.3. MGRF FRAMEWORK WITH FRAILTY

First, let $\Lambda_i(t|\mathbf{x}_i^{(r)}(t), Z_i)$ be the cumulative hazard function for i^{th} geographical location. Then the cumulative survival function $\bar{F}_i^{(r)}(t|\mathbf{x}_i^{(r)}(t), Z_i) = \exp\left[-\Lambda(t|\mathbf{x}_i^{(r)}(t), Z_i)\right]$ follows a uniform distribution on $(0, 1)$. It can also be shown that $\Lambda_i^{(r)}(T_i^{(r)}|\mathbf{x}_i^{(r)}(t), Z_i)$ follows a unit exponential distribution EXP(1). If $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, then the probit transformation of a variable U in $(0, 1)$ is $\Phi^{-1}(U)$. Based on this we probit transform our failure time $T_i^{(r)}$ to obtain $\tilde{T}_i^{(r)}$ as follows.

$$\tilde{\mathrm{T}}_i^{(r)} := \Phi^{-1}\left[1 - e^{-Zi\Lambda_{0i}(T_i^{(r)})\exp(\beta'\mathbf{x}_i^{(r)}(t))}|Z_i\right]$$

This transformed version of event times follows a standard normal distribution $N(0, 1)$. By defining a vector of transformed failure times for each subject in location i as below

$$\tilde{\mathbf{T}}_i = (\tilde{T}_i^{(1)}, \tilde{T}_i^{(2)}, ..., \tilde{T}_i^{(n_i)}),$$

it can be shown to follow an n_i variate joint multivariate normal distribution. As a result, a multivariate Gaussian random field (MGRF) can be constructed using each of the transformed failure times $\tilde{\mathbf{T}}_i$, $i = 1, ..., k$ of all geographical locations, namely

$$\tilde{\mathbf{T}} = \left(\tilde{\mathbf{T}}_1, \tilde{\mathbf{T}}_2, ..., \tilde{\mathbf{T}}_k\right)_{(n_1,...,n_k)}$$

3. SEMI PARAMETRIC ESTIMATING EQUATIONS

This section gives essentials of method of moment estimator, joint modeling and estimation of regression and spatial parameters.

3.1. METHOD OF MOMENT ESTIMATOR

In this section some useful notation similar to those in the previous chapter are introduced to describe the mathematical setup of the problem of interest using stochastic process framework.

For $(i, r) \in \mathcal{L} \times \mathcal{L}_i$, define the counting process $\mathrm{N}_i^{(r)}(t) = \delta_i^{(r)}\mathrm{I}(T_i^{(r)} \leq t)$ where $I(\cdot)$ is an indicator function and $\delta_i = I\left(C_i^{(r)} \leq W_i^{(r)}\right)$ is a non-censoring indicator. At risk process is defined as $\mathrm{Y}_i^{(r)}(t) = \mathrm{I}(T_i^{(r)} \geq t)$. Note that $\mathrm{N}_i^{(r)}(t)$ indicates if an event has occurred by time t, whereas $\mathrm{Y}_i^{(r)}(t)$ indicates if unit (i, r) is at risk at time t. Furthermore, it is assumed that the study ends at a time τ with $\tau \geq \max_{r,i}\mathrm{T}_i^{(r)}$. Therefore, the observation time window is $[0, \tau] = \mathcal{T}$. The entire history at all geostatistical locations at the end of the study is contained in the σ-field $\mathcal{F} = \bigvee_{i=1}^{k}\bigvee_{r=1}^{n_i}\mathcal{F}_{i,\tau}^{(r)}$ with

$$\mathcal{F}_{i,\tau}^{(r)} = \sigma\left(N_i^{(r)}(t), Y_i^{(r)}(t), t \in \mathcal{T}\right).$$

In this dissertation, instantaneous hazard function is assumed to be different from one geographical region to another. From stochastic integration theory, the compensator process of $N_i^{(r)}(t)$ conditional on $Z_i = z_i$ is $A_i^{(r)}(t|z_i)$ given by $A_i^{(r)}(t|z_i) = z_i \int_0^t Y_i^{(r)}(u)\lambda_i(u)du$ so that for each (i, r) the process

$$M_i^{(r)}(t|Z_i) = N_i^{(r)}(t) - \int_0^t Y_i^{(r)}(u) Z_i \lambda_{0i}(u) \exp\left(\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)\right) du \ : t \in \mathcal{T}$$

is a zero-mean square-integrable martingale with respect to the filtration $\mathcal{F}_{i,t}^{(r)}$ conditional on Z_i. Hence by method of moments, an Aalen-Breslow estimator for $\Lambda_{0i}(\cdot) = \int_0^{\cdot} \lambda_{0i}(u)du$, for $i \in \mathcal{L}$ is given by

$$\hat{\Lambda}_{0i}(t|z_i) = \int_0^t \frac{\sum_{r=1}^{n_i} dN_i^{(r)}(u)}{\sum_{r=1}^{n_i} Y_i^{(r)}(u) z_i \exp\left(\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)\right)}. \tag{2}$$

Thus, the estimator of $\Lambda_0(t|\mathbf{Z})$ would be a $k \times 1$ vector

$$\hat{\mathbf{\Lambda}}_\mathbf{0}(t|\mathbf{Z}) = \begin{bmatrix} \hat{\Lambda}_{01}(t|Z_1) \\ \hat{\Lambda}_{02}(t|Z_2) \\ \vdots \\ \hat{\Lambda}_{0k}(t|Z_k) \end{bmatrix}_{k\times 1} . \tag{3}$$

Since $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is an unknown regression parameter vector, $\hat{\mathbf{\Lambda}}_0(t|\mathbf{Z})$ is not yet an estimator. Therefore expression in (3) will be used as substitution for $\lambda_{0i}(t)$ to estimate $\boldsymbol{\beta}$ and also to obtain the in-probability limits of the score matrix.

3.2. JOINT MODELING

With a view towards joint modeling, for a pair of units $(r, s) \in (\mathcal{L}_i, \mathcal{L}_j)$ and $t \in [0, \tau]$, let $[N_i^{(r)}(t), Y_i^{(r)}(t), A_i^{(r)}(t|z_i)]$ and $[N_j^{(s)}(t), Y_j^{(s)}(t), A_j^{(s)}(t|z_j)]$ be their counting, at-risk, and conditional compensator processes respectively. Note that then there are zero mean martingale processes conditional on z_i and z_j namely, $\{M_i^{(r)}(t|z_i) : t \in [0, \tau]\}$ and $\{M_j^{(s)}(t|z_j) : t \in [0, \tau]\}$ with respect to their corresponding filtrations denoted by $\mathcal{F}_{i,t}^{(r)}$ and $\mathcal{F}_{j,t}^{(s)}$ respectively. Next for $(t_1, t_2) \in [0, \tau]^2$, define the joint counting process $N_{ij}^{(r,s)}(\cdot, \cdot)$ by $N_{ij}^{(r,s)}(t_1, t_2) = \mathrm{I}\{T_i^{(r)} \geq t_1, T_j^{(s)} \geq t_2\}$. The covariance function $\langle M_i^{(r)}(t_1|z_i), M_j^{(s)}(t_2|z_j)\rangle$ is defined by

$$\begin{aligned} E(M_i^{(r)}(t_1)M_j^{(s)}(t_2)|T_i^{(r)} > t_1, T_j^{(s)} > t_2, Z_i, Z_j) &= A_{i,j}^{(r,s)}(t_1, t_2|Z_i, Z_j) \\ &= \langle M_i^{(r)}, M_j^{(s)}\rangle(t_1, t_2|Z_i, Z_j). \end{aligned}$$

Further, by stochastic integration theory,

$$E\left(M_i^{(r)}(t_1|Z_i)M_j^{(s)}(t_2|Z_j) - \int_0^{t_1}\int_0^{t_2} Y_i^{(r)}(u_1)Y_j^{(s)}(u_2)Z_iZ_jA_{i,j}^{(r,s)}(du_1, du_2)\right) = 0.$$

Next, for the ith and jth geographical locations $T_i^{(1)}, T_i^{(2)}, ..., T_i^{(n_i)}$ conditional on Z_i are i.i.d.. Similarly, $T_j^{(1)}, T_j^{(2)}, ..., T_j^{(n_j)}$ conditional on Z_j are also i.i.d.. They have the conditional survivor function $\bar{F}_i^{(r)}(t|Z_i)$ and $\bar{F}_j^{(s)}(t|Z_j)$ respectively. Conditional bivariate survivor function $\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}|Z_i, Z_j)$ for $(r, s) \in (\mathcal{L}_i, \mathcal{L}_j)$ is given by

$$\begin{aligned}\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}|Z_i, Z_j) &= \mathrm{P}\left(T_i^{(r)} > t_1, T_j^{(s)} > t_2; \rho_{ij}|Z_i, Z_j\right) \\ &= G\left[\Phi^{-1}(F_i^{(r)})(t_1|Z_i), \Phi^{-1}(F_j^{(s)}(t_2|Z_j))\right],\end{aligned}$$

where ρ_{ij} and $G(\cdot, \cdot; \rho_{ij}|Z_i, Z_j)$ are spatial correlation and conditional bivariate survival function of the transformed failure times $\tilde{T}_i$ and $\tilde{T}_j$ respectively. Following Prentice and Cai (1992) the conditional joint compensator, $A_{ij}^{(r,s)}(t_1, t_2|Z_i, Z_j)$, is given by

$$A_{ij}^{(r,s)}(dt_1, dt_2; \rho_{ij}|Z_i, Z_j) = A_0[\Lambda_i^{(r)}(t_1|Z_i), \Lambda_j^{(s)}(t_2|Z_j); \rho_{ij}]\Lambda_i^{(r)}(dt_1|Z_i)\Lambda_j^{(s)}(dt_2|Z_j),$$

where the baseline joint compensator $A_0[\cdot, \cdot; \rho_{ij}]$ is given by

$$\begin{aligned}A_0(t_1, t_2; \rho_{ij}|Z_i, Z_j) &= \frac{\partial^2}{\partial t_1 \partial t_2}\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}|Z_i, Z_j) + \bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}|Z_i, Z_j) \\ &\quad + \frac{\partial}{\partial t_1}\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}|Z_i, Z_j) + \frac{\partial}{\partial t_2}\bar{F}_{ij}^{(r,s)}(t_1, t_2; \rho_{ij}|Z_i, Z_j).\end{aligned}$$

3.3. ESTIMATION

This section gives the theory on estimating regression and spatial parameters.

3.3.1. Estimating $\boldsymbol{\beta}$. The model in our study was assumed to have independent and identically distributed (i.i.d) frailty random variables that come from a known distribution. Particularly, the gamma distribution with unit mean and variance $1/\alpha$, Gamma(α,α) will be used as the known distribution. It is difficult to apply direct maximum likelihood methods to estimate parameters of interest due to the high dimensionality of the likelihood and the fact

Z_i are not observed. Again we use *pairwise likelihood* approach to estimate the parameters using expectation-maximization algorithm. This approach uses data in two spatial locations that can be the basis of an unbiased estimating function.

More notation in the sequel are introduced with a view toward estimating $\boldsymbol{\beta}$ that accounts for pairwise spatial correlation between two locations $(i, j) \in \mathcal{L}^2$. If $\mathbf{a} = (a_1, a_2)$ is a 1×2 row vector and its transpose denoted by $\mathbf{a}'$ is a 2×1 column vector. Using usual notation used in the first part of this dissertation, for $(r, s) \in (\mathcal{L}_i, \mathcal{L}_j)$, we define $\mathbf{H}_{ij}^{(r,s)}(t|Z_i, Z_j) = (H_i^{(r)}(t|Z_i), H_j^{(s)}(t|Z_j))$ and $\mathbf{M}_{ij}^{(r,s)}(t|Z_1, Z_j) = (M_i^{(r)}(t|Z_i), M_j^{(s)}(t|Z_j))'$.

In our case, the pairwise likelihood can be written as

$$\begin{aligned}
L_{ij}(\boldsymbol{\beta}|\boldsymbol{\delta_0}, Z_i, Z_j) \quad = \quad & \prod_{r=1}^{n_i}\prod_{s=1}^{n_j}\left\{\prod_{u=0}^{t}\left[Y_i^{(r)}(u)z_i e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)\right]^{w_{11}^{ij}dN_i^{(r)}(u)}\right.\\
& \left.\times \exp\left[-\int_0^t\sum_{r=1}^{n_i}Y_i^{(r)}(u)z_i e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du\right]\right\}\\
& \times\left\{\prod_{u=0}^{t}\left[Y_j^{(s)}(u)z_j e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0i}(u)\right]^{w_{12}^{ij}dN_i^{(r)}(u)}\right.\\
& \left.\times \exp\left[-\int_0^t\sum_{s=1}^{n_j}Y_j^{(s)}(u)z_j e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0i}(u)du\right]\right\}\\
& \times\left\{\prod_{u=0}^{t}\left[Y_i^{(r)}(u)z_i e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0j}(u)\right]^{w_{21}^{ij}dN_j^{(r)}(u)}\right.\\
& \left.\times \exp\left[-\int_0^t\sum_{r=1}^{n_i}Y_i^{(r)}(u)z_i e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0j}(u)du\right]\right\}\\
& \times\left\{\prod_{u=0}^{t}\left[Y_j^{(s)}(u)z_j e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0j}(u)\right]^{w_{22}^{ij}dN_j^{(s)}(u)}\right.\\
& \left.\times \exp\left[-\int_0^t\sum_{s=1}^{n_j}Y_j^{(s)}(u)z_j e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0j}(u)du\right]\right\}.
\end{aligned}$$

Then the corresponding log likelihood can be written as

$$
\begin{aligned}
l_{ij}(\boldsymbol{\beta}|\boldsymbol{\delta_0}, Z_i, Z_j) \quad \propto \quad & \sum_{r=1}^{n_i}\sum_{s=1}^{n_j}\left\{\int_0^t w_{11}^{ij}[\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u) - \log S_i^{(0)}(u)]dN_i^{(r)}(u)\right. \\
& + \int_0^t w_{12}^{ij}[\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u) - \log S_i^{(0)}(u)]dN_i^{(r)}(u) \\
& + \int_0^t w_{21}^{ij}[\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u) - \log S_j^{(0)}(u)]dN_j^{(s)}(u) \\
& \left. + \int_0^t w_{22}^{ij}[\boldsymbol{\beta}'\mathbf{x}_j^{sr)}(u) - \log S_j^{(0)}(u)]dN_j^{(s)}(u)\right\}.
\end{aligned}
$$

To estimate $\boldsymbol{\beta}$, if the $\mathbf{Z} = \mathbf{z} = (z_1, z_2, ..., z_k)$ are observed, then the complete likelihood function for the model parameter $\boldsymbol{\beta}$ is given by

$$
\begin{aligned}
L_{ij}(\boldsymbol{\beta}, Z_i, Z_j|\boldsymbol{\delta_0}) = \prod_{r=1}^{n_i}\prod_{s=1}^{n_j} & \left\{ \left[\frac{z_i^{\frac{1}{\alpha}-1}\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} \exp\left(\frac{-z_i}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \right] \left[\frac{z_j^{\frac{1}{\alpha}-1}\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} \exp\left(\frac{-z_j}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \right] \right. \\
& \times \left\{ \prod_{u=0}^{t} \left[Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0i}(u) \right]^{w_{11}^{ij} dN_i^{(r)}(u)} \right. \\
& \left. \times \exp\left[-\int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0i}(u) du \right] \right\} \\
& \times \left\{ \prod_{u=0}^{t} \left[Y_j^{(s)}(u) z_j e^{\boldsymbol{\beta}' \mathbf{x}_j^{(s)}(u)} \hat{\lambda}_{0i}(u) \right]^{w_{12}^{ij} dN_i^{(r)}(u)} \right. \\
& \left. \times \exp\left[-\int_0^t \sum_{s=1}^{n_j} Y_j^{(s)}(u) z_j e^{\boldsymbol{\beta}' \mathbf{x}_j^{(s)}(u)} \hat{\lambda}_{0i}(u) du \right] \right\} \\
& \times \left\{ \prod_{u=0}^{t} \left[Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0j}(u) \right]^{w_{21}^{ij} dN_j^{(r)}(u)} \right. \\
& \left. \times \exp\left[-\int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0j}(u) du \right] \right\} \\
& \times \left\{ \prod_{u=0}^{t} \left[Y_j^{(s)}(u) z_j e^{\boldsymbol{\beta}' \mathbf{x}_j^{(s)}(u)} \hat{\lambda}_{0j}(u) \right]^{w_{22}^{ij} dN_j^{(s)}(u)} \right. \\
& \left. \left. \times \exp\left[-\int_0^t \sum_{s=1}^{n_j} Y_j^{(s)}(u) z_j e^{\boldsymbol{\beta}' \mathbf{x}_j^{(s)}(u)} \hat{\lambda}_{0j}(u) du \right] \right\} \right\}.
\end{aligned}
$$

Note that the above equation, as a function of z_i, is proportional to

$$\begin{aligned} L_{ij}(\boldsymbol{\beta}, Z_i, Z_j|\boldsymbol{\delta_0}) \propto\ & \prod_{r=1}^{n_i}\prod_{s=1}^{n_j} z_i^{\frac{1}{\alpha}+w_{11}^{ij}dN_i^{(r)}(u)+w_{21}^{ij}dN_j^{(s)}(u)-1} \\ & \times \exp\left\{-z_i\left[\frac{1}{\alpha}+\int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du \right.\right. \\ & \left.\left. + \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0j}(u)du\right]\right\} \\ & \times z_i^{\frac{1}{\alpha}+w_{12}^{ij}dN_i^{(r)}(u)+w_{22}^{ij}dN_j^{(s)}(u)-1} \\ & \times \exp\left\{-z_j\left[\frac{1}{\alpha}+\int_0^t \sum_{s=1}^{n_j} Y_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0i}(u)du \right.\right. \\ & \left.\left. + \int_0^t \sum_{s=1}^{n_j} Y_j^{(s)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_j^{(s)}(u)}\hat{\lambda}_{0j}(u)du\right]\right\}. \end{aligned}$$

Since z_i and z_j are independent, given α, $\Lambda_{0i}(\cdot)$, $\Lambda_{0j}(\cdot)$, $N_i^{(r)}$ and $N_j^{(s)}$;

$$L(z_i, data) \sim gamma\,(A, B)$$

where $A = \frac{1}{\alpha} + w_{11}^{ij} N_i^{(r)}(u) + w_{21}^{ij} N_j^{(s)}(u)$ and $B = \frac{1}{\alpha} + \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0i}(u) du + \sum_{r=1}^{n_i} Y_i^{(r)}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0j}(u) du$ and $L_i(\theta, z_i)$ has the form

$$
\begin{aligned}
L_i(\boldsymbol{\beta}, Z_i|\boldsymbol{\delta_0}) \quad = \quad & \prod_{r=1}^{n_i} \prod_{s=1}^{n_j} \left\{ \left[\frac{z_i^{\frac{1}{\alpha}-1} \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} \exp\left(\frac{-z_i}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \right] \right. \\
& \times \left\{ \prod_{u=0}^{t} \left[Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0i}(u) \right]^{w_{11}^{ij} dN_i^{(r)}(u)} \right. \\
& \times \exp \left. \left[- \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0i}(u) du \right] \right\} \\
& \times \left\{ \prod_{u=0}^{t} \left[Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0j}(u) \right]^{w_{21}^{ij} dN_j^{(r)}(u)} \right. \\
& \times \exp \left. \left. \left[- \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) z_i e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0j}(u) du \right] \right\} \right\}. \qquad (4)
\end{aligned}
$$

Letting

$$
\eta = \frac{1}{\alpha} + \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0i}(u) du + \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \hat{\lambda}_{0j}(u) du
$$

and substituting it in (4) we get

$$
\begin{aligned}
L_i(\boldsymbol{\beta}, Z_i|\boldsymbol{\delta_0}) \quad = \quad & \prod_{r=1}^{n_i} \prod_{s=1}^{n_j} \prod_{u=0}^{t} \left[Y_i^{(r)}(u) \hat{\lambda}_{0i}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \right]^{w_{11}^{ij} dN_i^{(r)}(u)} \\
& \times \left[Y_i^{(r)}(u) \hat{\lambda}_{0i}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} \right]^{w_{21}^{ij} dN_j^{(s)}(u)} \\
& \times [\eta z_i]^{\frac{1}{\alpha} + w_{11}^{ij} dN_i^{(r)}(u) + w_{21}^{ij} dN_j^{(s)}(u) - 1} \exp(-\eta z_i) d(\eta z_i) \\
& \times \left[\eta^{\frac{1}{\alpha} + w_{11}^{ij} dN_i^{(r)}(u) + w_{21}^{ij} dN_j^{(s)}(u)} \cdot \Gamma\left(\frac{1}{\alpha}\right) \alpha^{\frac{1}{\alpha}} \right]^{-1}.
\end{aligned}
$$

Since Z_i are not observed, integrating out Z_i we get

$$
\begin{aligned}
L_i(\boldsymbol{\beta}|\boldsymbol{\delta_0}) \;=\; & \prod_{r=1}^{n_i}\prod_{s=1}^{n_j}\prod_{u=0}^{t}\left[Y_i^{(r)}(u)\hat{\lambda}_{0i}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\right]^{w_{11}^{ij}dN_i^{(r)}(u)} \\
& \times\left[Y_i^{(r)}(u)\hat{\lambda}_{0i}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\right]^{w_{21}^{ij}dN_j^{(s)}(u)} \\
& \times\Gamma\left(\frac{1}{\alpha}+w_{11}^{ij}dN_i^{(r)}(u)+w_{21}^{ij}dN_j^{(s)}(u)\right) \\
& \times\left[\frac{1}{\alpha}+\int_0^t\sum_{r=1}^{n_i}Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du\right. \\
& \left.+\int_0^t\sum_{r=1}^{n_i}Y_i^{(r)}(u)e^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0j}(u)du\right]^{-\left(\frac{1}{\alpha}+w_{11}^{ij}dN_i^{(r)}(u)+w_{21}^{ij}dN_j^{(s)}(u)\right)} \\
& \times\left(\Gamma\left(\frac{1}{\alpha}\right)\alpha^{\frac{1}{\alpha}}\right)^{-1}. \qquad (5)
\end{aligned}
$$

And also the full log likelihood is obtained by taking logarithm of (4) which yields,

$$
\begin{aligned}
l_i(\boldsymbol{\beta}|\boldsymbol{\delta_0}) \;=\; & \left[\frac{1}{\alpha}-1+w_{11}^{ij}N_i^{(r)}(t)+w_{21}^{ij}N_j^{(s)}(t)\right]\log z_i \\
& -\frac{n_i}{\alpha}\log\alpha-n_i\log\Gamma\left(\frac{1}{\alpha}\right)-\frac{n_i}{\alpha}z_i \\
& +\sum_{r=1}^{n_i}\int_0^t\left[\log Y_i^{(r)}(u)+\log z_i+\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)+\log\hat{\lambda}_{0i}(u)\right]w_{11}^{ij}dN_i^{(r)}(u) \\
& +\sum_{s=1}^{n_j}\int_0^t\left[\log Y_i^{(r)}(u)+\log z_i+\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)+\log\hat{\lambda}_{0j}(u)\right]w_{21}^{ij}dN_j^{(s)}(u) \\
& -\int_0^t\sum_{r=1}^{n_i}Y_i^{(r)}(u)z_ie^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0i}(u)du \\
& -\int_0^t\sum_{r=1}^{n_i}Y_i^{(r)}(u)z_ie^{\boldsymbol{\beta}'\mathbf{x}_i^{(r)}(u)}\hat{\lambda}_{0j}(u)du.
\end{aligned}
$$

The maximum likelihood estimator of the model parameters is the maximizer of this full likelihood process. Expectation maximization algorithm (EM) is used for the computations of the estimate. We give the main steps of this algorithm below. We will have the algorithm in detail in simulation section later. For the expectation step given $\boldsymbol{\theta}$, $r = 1, \ldots, n_i$ and

$u \in [0, t]$, the conditional expectation of the z_i is given by,

$$E(z_i|\alpha, \Lambda_{0i}(u), \Lambda_{0j}(u), N_i^{(r)}(u), N_j^{(s}(u)) = \frac{\frac{1}{\alpha} + w_{11}^{ij} N_i^{(r)}(u) + w_{21}^{ij} N_j^{(s)}(u)}{C}$$

where, $C = \frac{1}{\alpha} + \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)} [\hat{\lambda}_{0i}(u) + \hat{\lambda}_{0j}(u)] du$. And conditional expectation of the $\log(z_i)$ is given by,

$$\begin{aligned} E[\log z_i|\alpha, \Lambda_{0i}(u), \Lambda_{0j}(u), N_i^{(r)}(u), N_j^{(s}(u)] \quad = \quad & \log\left[\frac{1}{\alpha} + w_{11}^{ij} N_i^{(r)}(u) + w_{21}^{ij} N_j^{(s)}(u)\right] \\ & + \varphi\left[\frac{1}{\alpha} + w_{11}^{ij} N_i^{(r)}(u) + w_{21}^{ij} N_j^{(s)}(u)\right] \\ & - D \end{aligned}$$

where $D = \log E[z_i|\alpha, \Lambda_{0i}(u), \Lambda_{0j}(u), N_i^{(r)}(u), N_j^{(s}(u)]$ and, $\varphi(\cdot)$ is the digamma function. For maximization procedure, the only difference with the case without frailties is that $Y_i^{(r)}(t)$ will be replaced by $z_i Y_i^{(r)}(t)$. Similar to the estimating equation for $\boldsymbol{\beta}$ in the case of model without frailties, given $\mathbf{Z}$ and $\Lambda_0(t|z, \boldsymbol{\beta}, \delta)$, $\boldsymbol{\beta}$ can be estimated by solving the estimating equation below

$$\mathrm{U}^{[ij]}(t, \boldsymbol{\beta}|\boldsymbol{\delta}_0, Z_i, Z_j) = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \int_0^t \mathbf{H}_{ij}^{(r,s)}(u|Z_i, Z_j) \mathbf{W}^{ij}(\boldsymbol{\delta}_0) \mathbf{M}_{ij}^{(r,s)}(u|Z_i, Z_j) du. \tag{6}$$

where

$$\mathbf{H}_{ij}^{(r,s)}(u|Z_i, Z_j) = (H_i^{(r)}(u|Z_i), H_j^{(s)}(u|Z_j)),$$

$$H_i^{(r)}(u|Z_i) = \mathbf{x}_i^{(r)}(u) - \frac{\sum_{r=1}^{n_i} \mathbf{x}_i^{(r)}(u) Y_i^{(r)}(u) Z_i \lambda_{0i}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)}}{\sum_{r=1}^{n_i} Y_i^{(r)}(u) Z_i \lambda_{0i}(u) e^{\boldsymbol{\beta}' \mathbf{x}_i^{(r)}(u)}},$$

and

$$H_j^{(s)}(u|Z_j) = \mathbf{x}_j^{(s)}(u) - \frac{\sum_{s=1}^{n_j} \mathbf{x}_j^{(s)}(u) Y_j^{(s)}(u) Z_j \lambda_{0j}(u) e^{\boldsymbol{\beta}' \mathbf{x}_j^{(s)}(u)}}{\sum_{s=1}^{n_j} Y_j^{(r)}(u) Z_j \lambda_{0j}(u) e^{\boldsymbol{\beta}' \mathbf{x}_j^{(s)}(u)}}.$$

Also, $\mathbf{W}^{ij}(\boldsymbol{\delta}) = (w^{ij}(\boldsymbol{\delta}))$ is a 2×2 matrix whose elements are function of the spatial correlation $\boldsymbol{\delta}$ and the number of units in locations i and j by

$$\mathbf{W}^{ij}(\boldsymbol{\delta}_0) = \begin{pmatrix} w_{11}^{ij}(\boldsymbol{\delta}_0) & w_{12}^{ij}(\boldsymbol{\delta}_0) \\ w_{21}^{ij}(\boldsymbol{\delta}_0) & w_{22}^{ij}(\boldsymbol{\delta}_0) \end{pmatrix}.$$

Then, at time $t \in \mathcal{T}$, we can write the generalized estimating equation for $\boldsymbol{\beta}$ over all pairs as

$$\mathrm{U}(t, \boldsymbol{\beta}|\boldsymbol{\delta}_0, Z_i, Z_j) = \sum_{i \le j} \mathrm{U}^{[ij]}(t, \boldsymbol{\beta}|\boldsymbol{\delta}_0, Z_i, Z_j).$$

Examining $\mathrm{U}^{[ij]}(\cdot, \cdot|\boldsymbol{\delta}_0, Z_i, Z_j)$, it can be written as a sum of four terms each of which is given below

$$U_1^{ij}(t|Z_i) = \sum_{r=1}^{n_i} \int_0^t w_{11}^{ij} H_i^{(r)}(u, \boldsymbol{\beta}|Z_i) M_i^{(r)}(du),$$
$$U_2^{ij}(t|Z_j) = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \int_0^t w_{12}^{ij} H_j^{(s)}(u, \boldsymbol{\beta}|Z_j) M_i^{(r)}(du),$$
$$U_3^{ij}(t|Z_i) = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} \int_0^t w_{21}^{ij} H_i^{(r)}(u, \boldsymbol{\beta}|Z_i) M_j^{(s)}(du),$$
$$U_4^{ij}(t|Z_j) = \sum_{s=1}^{n_j} \int_0^t w_{22}^{ij} H_j^{(s)}(u, \boldsymbol{\beta}|Z_j) M_j^{(s)}(du).$$

3.3.2. Estimating $\boldsymbol{\delta}$. To estimate $\boldsymbol{\delta}$, a function which is an unbiased estimator of 0 is sought. Consequently, the objective is to find a weighted function of $M_i^{(r)}(t_1|Z_i) M_j^{(s)}(t_2|Z_j) - A_{i,j}^{(r,s)}(dt_1, dt_2; \rho_{ij}|Z_i, Z_j)$ which will be an estimating function for $\boldsymbol{\delta}$ with the flavor of score function.

Define the $(k \times k)$ matrix $\mathbf{A}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j) = \big(A_{ij}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j)\big)$ with $(i, j)^{th}$ entry given by

$$A_{ij}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j) = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} A_{ij}^{(r,s)}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j).$$

Let $\nabla_{\delta_l}\mathbf{A}(t_1, t_2; \rho(\boldsymbol{\delta}))$, $l = 1, ..., q$, be the matrix of element-wise derivatives of $\mathbf{A}(t, \rho(\boldsymbol{\delta}))$ with respect to δ_l. Define

$$\Pi_l = \mathbf{A}^{-1}\left[\nabla_{\delta_l}\mathbf{A}\right]\mathbf{A}^{-1},$$

where we use $\mathbf{A}$ for $\mathbf{A}(t_1, t_2; \rho(\boldsymbol{\delta})|z_i, z_j)$ for compactness. Then, for $l = 1, ..., q$, following Cressie (1993), Page 483, it can be shown that $\mathcal{E}(\mathbf{M}(t))\Pi_l\mathcal{E}(\mathbf{M}(t)) + \text{tr}(\Pi_l\mathbf{A}) = 0$, when the frailties are observed, where $\text{tr}(\cdot)$ denotes the trace of a matrix. Consequently, a score function can be defined for estimating the l^{th} component of $\boldsymbol{\delta}$ using two locations by

$$\begin{aligned} \mathrm{U}^{ij}_{\delta_l}(t_1, t_2|Z_i, Z_j) &= \mathbf{M}(t|Z_i, Z_j)\Pi_l\mathbf{M}'(t|Z_i, Z_j) + \text{tr}(\Pi_l\mathbf{A}) \\ &= \mathbf{M}(t|Z_i, Z_j)\Pi_l\mathbf{M}'(t|Z_i, Z_j) + \text{tr}(\mathbf{A}^{-1}\mathbf{A}_{\delta_l}). \end{aligned} \tag{7}$$

The expression in (7) can be viewed as a score process and its sum over all pairwise spatial locations (i, j) can serve as an estimating function for δ_l. So, the estimating function over all pairs of spatial locations for $\boldsymbol{\delta}$ is the $q \times 1$ vector $\mathrm{U}_{\boldsymbol{\delta}}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j) = (\mathrm{U}_{\delta_l}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j), l = 1, ..., q)'$ where

$$\mathrm{U}_{\delta_l}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j) = \sum_{(i,j), i \leq j} \mathrm{U}^{ij}_{\delta_l}(t_1, t_2; \rho(\boldsymbol{\delta})|Z_i, Z_j).$$

4. NUMERICAL IMPLEMENTATION

This section gives a description on the simulation design and discusses the simulation results obtained.

4.1. SIMULATION DESIGN

We are in the same simulation setting as in the first part of this dissertation. The only difference is we generate the unobservables frailty.

The parameter α which governs the gamma frailty variable was set to 1 and 80. The choice of 1 was to mimic the presence of frailty where as 80 was chosen to mimic the absence of frailty.

300 replications were performed with each sample size and parameter combination. We have listed the results obtained in Table 1, Table 2, Table 3, Table 4, Table 5, Table 6, Table 7 and Table 8 where CP stands for censoring percentage.

Basically, in this simulation part of this chapter, the procedures developed for the case without frailties in the first part of the dissertation were adapted for the case with frailties.

4.2. EXPECTATION-MAXIMIZATION ALGORITHM

Expectation-maximization (EM) algorithm is typically used in the presence of frailty, in order to estimate regression coefficients and frailty parameters. We use this algorithm in our work.

Mainly, this algorithm has two steps which are called the expectation step (E-Step) and maximization step (M- Step). In Expectation step, we calculate conditional expectation of unobserved frailties conditional on the observed information and obtain the current parameter estimates. In the maximization step, we take these expected values found in E-step as the true information. Then by maximizing the likelihood we obtain new estimates of the parameters of interest, given the expected values.

4.2.1. Initialization Step. First, by setting Z_i to 1, an ordinary Cox model is fitted and $\boldsymbol{\beta}_{initial}$ is estimated. So, we call $\boldsymbol{\beta}_{initial} = \hat{\boldsymbol{\beta}}^{(0)}$. Then the initial value for the cumulative hazard function is estimated by

$$\hat{\Lambda}_{0i}^{(0)}(t|\hat{Z}_i^{(0)}, \hat{\boldsymbol{\beta}}^{(0)}) = \int_0^t \frac{\sum_{r=1}^{n_i} dN_i^{(r)}(u)}{\sum_{r=1}^{n_i} Y_i^{(r)}(u)\hat{Z}_i^{(0)} \exp\left(\hat{\boldsymbol{\beta}}^{(0)\prime}\mathbf{x}_i^{(r)}(u)\right)}. \tag{8}$$

Next, an initial estimate $\hat{\alpha}^{(0)}$ of α is also specified.

4.2.2. E-Step. Using initial values specified in initialization step, namely $\hat{\alpha}^{(0)}$, $\hat{\boldsymbol{\beta}}^{(0)}$ and $\hat{\Lambda}_{0i}^{(0)}(t|\hat{Z_i}^{(0)}, \hat{\boldsymbol{\beta}}^{(0)})$ obtain $\hat{Z_i}^{(1)}$ and $\widehat{\log Z_i}^{(1)}$ by

$$\hat{Z_i}^{(1)} = \frac{\frac{1}{\hat{\alpha}^{(0)}} + w_{11}^{ij} N_i^{(r)}(u) + w_{21}^{ij} N_j^{(s)}(u)}{\frac{1}{\hat{\alpha}^{(0)}} + \int_0^t \sum_{r=1}^{n_i} Y_i^{(r)}(u) e^{\hat{\boldsymbol{\beta}}^{(0)\prime} \mathbf{x}_i^{(r)}(u)} [\hat{\lambda}_{0i}^{(0)}(u) + \hat{\lambda}_{0j}^{(0)}(u)] du}$$

and

$$\begin{aligned}\widehat{\log Z_i^{(1)}} \quad = \quad & \log\left[\frac{1}{\hat{\alpha}^{(0)}} + w_{11}^{ij} N_i^{(r)}(u) + w_{21}^{ij} N_j^{(s)}(u)\right] \\ & + \varphi\left[\frac{1}{\hat{\alpha}^{(0)}} + w_{11}^{ij} N_i^{(r)}(u) + w_{21}^{ij} N_j^{(s)}(u)\right] \\ & - \log E[z_i|\hat{\alpha}^{(0)}, \Lambda_{0i}^{(0)}(u), \Lambda_{0j}^{(0)}(u), N_i^{(r)}(u), N_j^{(s}(u)]\end{aligned}$$

where φ is the di gamma function. i.e., the first derivative of the logarithm of the gamma function.

4.2.3. M-Step.

1. Using (8), obtain $\hat{\Lambda}_{0i}^{(1)}(t|\hat{Z_i}^{(1)}, \hat{\boldsymbol{\beta}}^{(0)})$.

2. Obtain $\hat{\boldsymbol{\beta}}^1$ by substituting $\hat{Z_i}^{(1)}$ for Z in equation (6).

3. Calculate $\hat{\alpha}$ by maximizing the full likelihood in (5) with respect to α given the current values $(\hat{\Lambda}_{0i}^{(1)}, \hat{\boldsymbol{\beta}}^{(1)})$.

4. Compare the values $(\mathbf{Z}^{(1)}, \boldsymbol{\alpha}^{(1)})$ with the values $(\mathbf{Z}^{(0)}, \boldsymbol{\alpha}^{(0)})$ until the values of $\hat{\mathbf{Z}}$ and $\hat{\boldsymbol{\alpha}}$ have stabilized. After that, terminate the algorithm. At that time the estimates will be the final values. If not, replace $(\hat{\alpha}^{(0)}, \hat{\theta}_i^{(0)})$ by $(\hat{\alpha}^{(1)}, \hat{\theta}_i^{(1)})$ and proceed to step 1 of the algorithm.

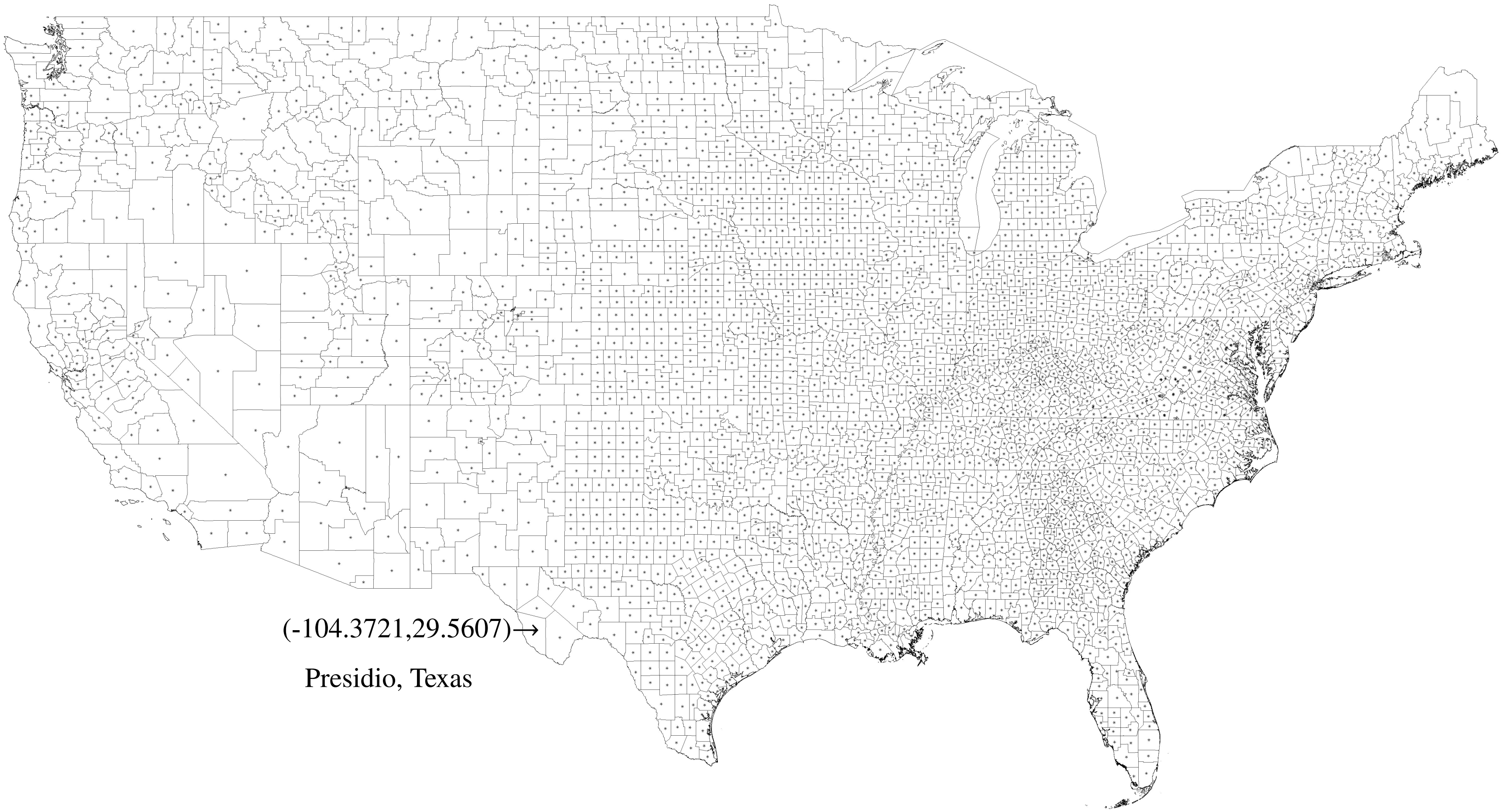

Figure 1. United States counties map

Table 1. Numerical results for the Matérn exponential correlation function when $\eta_2 = 1.5$ and $\alpha = 1$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$	$\hat{\alpha}$
5%	93510	1.029815	0.06729599	2.057801	0.13552010	0.4955121	0.12053941	1.488432	0.11112740	0.9731942
	155850	1.014823	0.06495556	2.042032	0.11051278	0.4992394	0.11334371	1.610395	0.10472624	0.9545107
	311700	1.006431	0.05473459	2.018490	0.09951939	0.5051744	0.12691218	1.549215	0.15831554	0.9090093
10%	93510	1.047730	0.06449028	2.104573	0.10631068	0.4874499	0.10649691	1.511984	0.07699964	0.9771219
	155850	1.030775	0.06013435	2.068024	0.10654855	0.5061238	0.11293126	1.438576	0.13941386	0.9563781
	311700	1.027802	0.05337257	2.051298	0.09327863	0.4930497	0.11546449	1.474758	0.11076289	0.9090040
20%	93510	1.075200	0.06433280	2.164343	0.11327967	0.5288665	0.11365795	1.603472	0.05575933	0.9735515
	155850	1.061672	0.06576021	2.131705	0.11733459	0.4982928	0.11438367	1.387781	0.04881755	0.9571072
	311700	1.050594	0.05264146	2.092443	0.09957561	0.4907045	0.11473691	1.505192	0.09484126	0.9085904
25%	93510	1.092659	0.06124843	2.184846	0.10802976	0.4950278	0.11362695	1.491698	0.10371179	0.9738883
	155850	1.074765	0.04774915	2.146604	0.09454380	0.4682383	0.06982386	1.503596	0.04716646	0.9539029
	311700	1.076244	0.04235159	2.140288	0.10131059	0.4526317	0.10066346	1.477103	0.09831267	0.9099881

Table 2. Numerical results for the Matérn exponential correlation function when $\eta_2 = 0.8$ and $\alpha = 1$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$	$\hat{\alpha}$
5%	93510	1.0298147	0.06729608	2.057801	0.13552022	0.5134468	0.10098063	1.562462	0.10580093	0.9731942
	155850	1.0341206	0.06870927	2.064386	0.12306539	0.4860302	0.10792099	1.446218	0.05659222	0.9552199
	311700	0.9984789	0.06281266	1.998170	0.11672808	0.4168675	0.10805437	1.576111	0.12014724	0.9101237
10%	93510	1.0508675	0.05790747	2.116513	0.10568375	0.5203977	0.12890535	1.516446	0.11313952	0.9721177
	155850	1.0421226	0.06186197	2.082221	0.10465472	0.4818415	0.10779321	1.431234	0.13094104	0.9542291
	311700	1.0096367	0.05760414	2.025881	0.10460210	0.4446421	0.09892984	1.538571	0.12164347	0.9087586
20%	93510	1.0686054	0.06495788	2.142292	0.11109328	0.4921543	0.12045447	1.551658	0.09672958	0.9723732
	155850	1.0787732	0.05432879	2.155540	0.10254942	0.5174955	0.09462939	1.529183	0.10424762	0.9558403
	311700	1.0490556	0.05306500	2.104589	0.10102172	0.5026691	0.10533320	1.400833	0.06760286	0.9096832
25%	93510	1.0883004	0.06627846	2.186066	0.10689041	0.4394719	0.10531318	1.557365	0.08627364	0.9713455
	155850	1.0710977	0.05418177	2.138554	0.09712947	0.5290640	0.12375550	1.545007	0.04283120	0.9518796
	311700	1.0723802	0.04334608	2.139368	0.08662750	0.4739613	0.13053625	1.438792	0.08938201	0.9068976

Table 3. Numerical results for the powered correlation function when $\eta_2 = 1.5$ and $\alpha = 1$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$	$\hat{\alpha}$
5%	93510	1.0297976	0.06677012	2.057658	0.13441018	0.4904419	0.12436872	1.411924	0.08566766	0.9726771
	155850	1.0335716	0.06575936	2.065785	0.12156014	0.4296011	0.12230111	1.520703	0.06406158	0.9551784
	311700	0.9975607	0.05692530	1.991894	0.11209812	0.5008602	0.11447406	1.494406	0.13094789	0.9102763
10%	93510	1.0504947	0.05954997	2.095521	0.11696553	0.4458940	0.12688910	1.481853	0.10223561	0.9705856
	155850	1.0307966	0.06081119	2.070436	0.11371851	0.5303000	0.08873202	1.520669	0.11810360	0.9528246
	311700	1.0132854	0.05070085	2.028760	0.09346839	0.4717819	0.11292959	1.487179	0.10299681	0.9086204
20%	93510	1.0762099	0.05967611	2.162980	0.12003897	0.4650780	0.08370460	1.398178	0.13148639	0.9762355
	155850	1.0617857	0.05963679	2.127436	0.11811524	0.5033934	0.10682083	1.489601	0.09856559	0.9555101
	311700	1.0477046	0.04812786	2.097496	0.08945003	0.4953500	0.11579610	1.481760	0.10042311	0.9024484
25%	93510	1.0902096	0.06692188	2.185399	0.10838223	0.5327789	0.10797154	1.446979	0.09649122	0.9734237
	155850	1.0652719	0.05276319	2.128149	0.09299337	0.5427218	0.09420071	1.495957	0.12162412	0.9510271
	311700	1.0640897	0.05085197	2.122893	0.09741570	0.4730951	0.10361080	1.477261	0.09989270	0.9084603

Table 4. Numerical results for the powered correlation function when $\eta_2 = 0.8$ and $\alpha = 1$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$	$\hat{\alpha}$
5%	93510	1.0297977	0.06677008	2.057658	0.13441010	0.5222716	0.11342932	1.520958	0.11338022	0.9726771
	155850	1.0151447	0.06551430	2.042825	0.11076297	0.4323340	0.09598532	1.429305	0.07981490	0.9546027
	311700	0.9854927	0.06039379	1.980880	0.11169694	0.5126132	0.09959637	1.451956	0.12727739	0.9086424
10%	93510	1.0595446	0.06569072	2.124685	0.11785367	0.4743243	0.11603366	1.423637	0.07538284	0.9734053
	155850	1.0335135	0.06296638	2.060783	0.12117217	0.4972412	0.10441265	1.471154	0.10266954	0.9541632
	311700	1.0266617	0.05709201	2.051110	0.10730734	0.4673013	0.11080530	1.434374	0.10030536	0.9090585
20%	93510	1.0880303	0.07545429	2.179516	0.12622636	0.5090908	0.11967975	1.407115	0.07692050	0.9672630
	155850	1.0673455	0.06249530	2.147557	0.10674219	0.4954051	0.10779545	1.517965	0.11503578	0.9543658
	311700	1.0388419	0.05024930	2.090638	0.08477245	0.4564342	0.10418029	1.453947	0.09000802	0.9090062
25%	93510	1.1040061	0.06091270	2.209749	0.11073527	0.4656196	0.10760151	1.576729	0.09819009	0.9686993
	155850	1.0725780	0.05559012	2.151594	0.09826625	0.4725774	0.09112407	1.446028	0.04689739	0.9562210
	311700	1.0741758	0.04558763	2.146315	0.08796930	0.4354300	0.09859633	1.531692	0.10315632	0.9068350

Table 5. Numerical results for the Matérn exponential correlation function when $\eta_2 = 1.5$ and $\alpha = 80$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.7822744	0.10042276	1.560656	0.14991148	0.4781593	0.10571357	1.490568	0.13263675
	155850	0.8641238	0.05030686	1.719723	0.09014568	0.5249068	0.12352137	1.430255	0.09321591
	311700	0.9656768	0.04445717	1.922655	0.08738155	0.4993314	0.11028005	1.519357	0.10024713
10%	93510	0.7734162	0.10425711	1.547982	0.14145007	0.4847504	0.10627189	1.493118	0.12062017
	155850	0.8633048	0.05595902	1.720429	0.09784264	0.5144859	0.11971700	1.480974	0.11968495
	311700	0.9608139	0.04628279	1.913547	0.09002798	0.5002486	0.12825714	1.388380	0.06354968
20%	93510	0.7702852	0.10150908	1.540944	0.14711942	0.5100953	0.09477460	1.442438	0.08088030
	155850	0.8519345	0.05426860	1.699649	0.09511546	0.4621895	0.09403900	1.531357	0.11815007
	311700	0.9557099	0.04525894	1.899531	0.09004018	0.4843984	0.09691497	1.513268	0.09512435
25%	93510	0.7711973	0.11114501	1.536129	0.14618372	0.5441557	0.10871051	1.524998	0.10970698
	155850	0.8462685	0.05329960	1.684695	0.09413872	0.4634573	0.10042738	1.497794	0.14067757
	311700	0.9458192	0.04436356	1.880800	0.08789099	0.5446168	0.11982288	1.503245	0.11052694

Table 6. Numerical results for the Matérn exponential correlation function when $\eta_2 = 0.8$ and $\alpha = 80$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.7748998	0.10183450	1.546166	0.14201940	0.4882348	0.11443885	1.474876	0.09779100
	155850	0.8659328	0.05193996	1.722957	0.09106426	0.4807230	0.11170969	1.481125	0.11471019
	311700	0.9594776	0.04402847	1.910161	0.08585052	0.4655450	0.09277994	1.466099	0.10876257
10%	93510	0.7768684	0.10800230	1.545498	0.13933938	0.4462199	0.11451939	1.438839	0.12150385
	155850	0.8576143	0.05277886	1.705625	0.09764180	0.4269675	0.11660109	1.454666	0.10952768
	311700	0.9596312	0.04141641	1.909722	0.08038290	0.4667544	0.12044353	1.400070	0.10626081
20%	93510	0.7697540	0.10568070	1.527925	0.15124119	0.4463510	0.10309318	1.501104	0.11256934
	155850	0.8541032	0.05115987	1.703836	0.09514274	0.4263876	0.10085830	1.547133	0.10169458
	311700	0.9531669	0.04368928	1.895089	0.08648275	0.4526783	0.10419536	1.529238	0.08352670
25%	93510	0.7591065	0.10913657	1.528585	0.15268638	0.4755711	0.11954677	1.497023	0.09135412
	155850	0.8452898	0.05435434	1.686046	0.09768092	0.4849959	0.11653679	1.494860	0.09271720
	311700	0.9475552	0.04440450	1.884488	0.08801135	0.5250469	0.10591207	1.439092	0.07686285

Table 7. Numerical results for the powered correlation function when $\eta_2 = 1.5$ and $\alpha = 80$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.7715399	0.10051988	1.556166	0.14011544	0.4499527	0.0906412	1.501031	0.10911225
	155850	0.8657452	0.05036445	1.728770	0.08969768	0.4717423	0.1037612	1.476357	0.09088143
	311700	0.9613030	0.04205719	1.915105	0.08209854	0.5027827	0.1201323	1.494042	0.08246186
10%	93510	0.7726270	0.09616186	1.542451	0.14118200	0.5051200	0.1165951	1.450969	0.09633239
	155850	0.8592544	0.05180736	1.712537	0.09289018	0.4594674	0.1298239	1.538667	0.12832220
	311700	0.9568549	0.04320916	1.904039	0.08672308	0.4346925	0.1044191	1.457736	0.11169946
20%	93510	0.7703975	0.10521271	1.535270	0.15692027	0.4873341	0.1211949	1.514300	0.15230805
	155850	0.8532241	0.04680881	1.698623	0.08649556	0.4573149	0.1040705	1.573742	0.09072474
	311700	0.9529985	0.04335377	1.895621	0.08454086	0.4691986	0.1214077	1.521292	0.11410709
25%	93510	0.7700280	0.10767472	1.526667	0.13793426	0.4898497	0.1127624	1.457671	0.12063205
	155850	0.8482416	0.05371444	1.691411	0.09647265	0.4882707	0.1081157	1.524964	0.10627590
	311700	0.9481910	0.04650724	1.884490	0.09025366	0.4928822	0.1215053	1.430373	0.09325126

Table 8. Numerical results for the powered correlation function when $\eta_2 = 0.8$ and $\alpha = 80$

CP	n	$\bar{\beta}_1$	$\hat{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\hat{\sigma}_{\beta_2}$	$\bar{\delta}_1$	$\hat{\sigma}_{\delta_1}$	$\bar{\delta}_2$	$\hat{\sigma}_{\delta_2}$
5%	93510	0.7787358	0.10437284	1.558627	0.13280294	0.4779905	0.1184450	1.547133	0.10169458
	155850	0.8666975	0.05167690	1.725039	0.09239407	0.5304549	0.1027453	1.513268	0.09512435
	311700	0.9601151	0.04147368	1.912017	0.08142999	0.4832991	0.1053006	1.520087	0.11196502
10%	93510	0.7798831	0.10110653	1.547783	0.13724420	0.4582216	0.1318242	1.441526	0.09739045
	155850	0.8570229	0.04953480	1.710144	0.08424719	0.5070480	0.1127849	1.424603	0.08090544
	311700	0.9581474	0.04503409	1.908502	0.08799876	0.4313484	0.1125438	1.493118	0.12062017
20%	93510	0.7762263	0.11036121	1.543216	0.14155035	0.5333254	0.1002577	1.511277	0.11847563
	155850	0.8570660	0.04933424	1.702625	0.09308367	0.4789544	0.1132957	1.513660	0.12258979
	311700	0.9549337	0.04469491	1.899613	0.08754672	0.4652688	0.1211030	1.443540	0.11838361
25%	93510	0.7608348	0.10863988	1.515490	0.14329973	0.4301427	0.1213539	1.495397	0.05561716
	155850	0.8427310	0.05288128	1.681707	0.09422412	0.4588298	0.1209582	1.536580	0.13753493
	311700	0.9476600	0.04510312	1.884357	0.08812584	0.4939713	0.1110672	1.523255	0.10580306

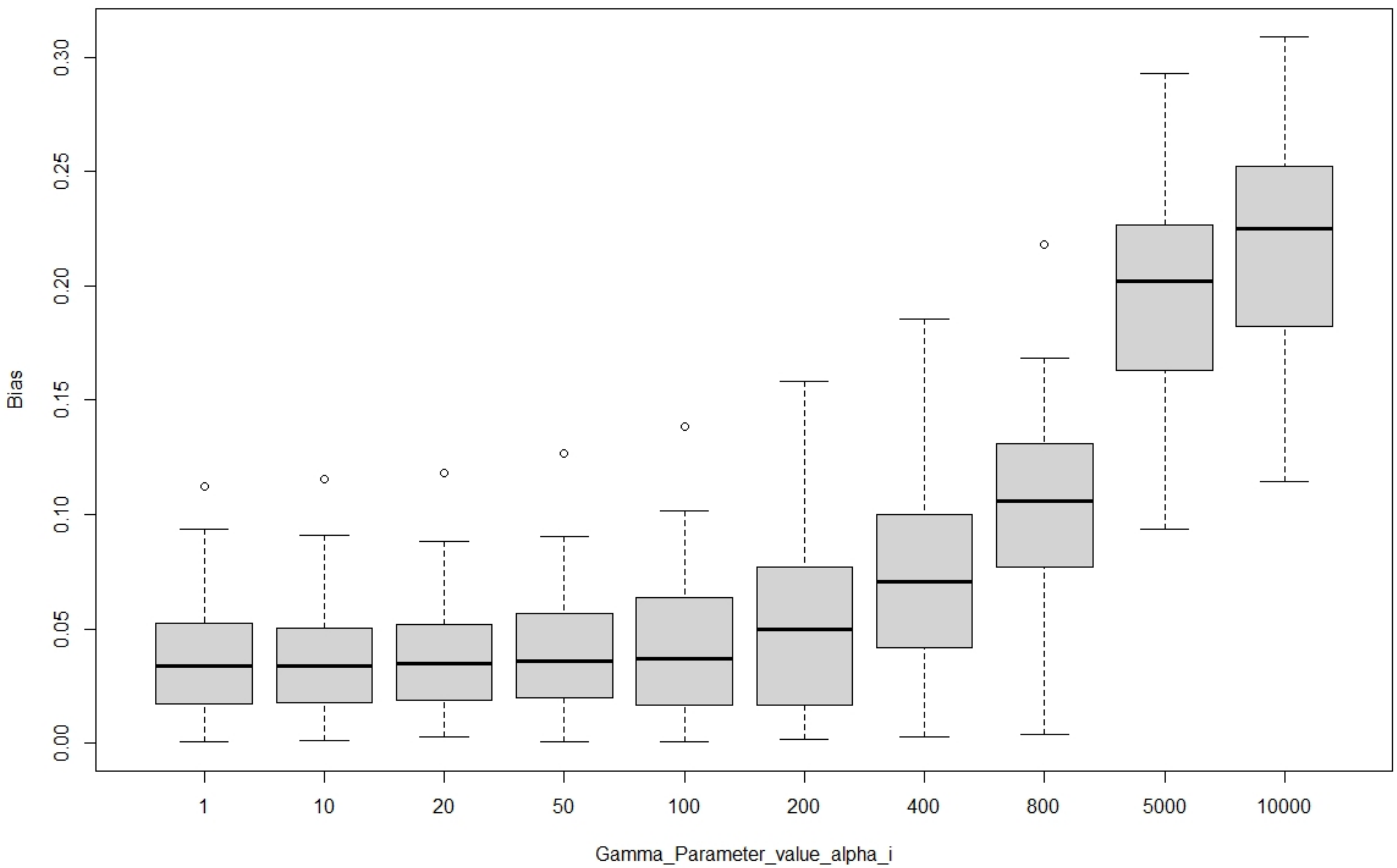

Figure 2. Box plots of bias vs. α. (This is for exponential correlation function. Weibull shape parameter of η_2 = 0.8 and n=100)

4.3. DISCUSSIONS OF SIMULATION RESULTS

Table 1, Table 2, Table 3, Table 4, Table 5, Table 6, Table 7 and Table 8 list mean values and the standard deviations of the estimators of β_1, β_2, δ_1 and δ_2 according to the values of η_2 (0.8, 1.5), sample size (30×3117, 50×3117, 100×3117) and spatial correlation model (exponential, powered exponential). Additionally, Table 1, Table 2, Table 3 and Table 4 contain mean values of α. Effects of changing the values of sample size, CP, η_2, α and spatial correlation model on the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are investigated.

As the sample size increases the estimators of the regression coefficient $\boldsymbol{\beta}$ improves, with deceasing biases and standard errors. This is true for both exponential and powered exponential spatial correlation models.

Moreover, the estimates of frailty parameter which were obtained using EM algorithm were close to the true values.

In Figure 2 we also observe that bias of the estimator of $\boldsymbol{\beta}$ decreases as the α decreases. Regardless of the sample size, it is noted that as censoring percentage (CP) increases, the bias of $\boldsymbol{\beta}$ increases. It is justifiable since, higher censoring means less failure times. On the other hand, regardless of the sample size, the biases of the $\boldsymbol{\delta}$ remain very steady and are not affected by the change in CP. This is clear since the spatial correlation parameter is the correlation between two areas and hence it is not affected by sample size.

No significant difference in the results between exponential and powered exponential was observed. It was not a surprise since both model have exponential components.

However, the standard deviations of the estimates of $\boldsymbol{\delta}$ do not have any noticeable pattern with the increasing sample size. This is also true because of above mentioned reason.

It is worthwhile noting the fact that accounting for frailty in the model improved the performance of the regression parameter in the presence of frailty. This was evident as bias of estimator of regression parameter decreases when frailty variance increase. This can be justified because subjects in the same location are not independent due to the effects of frailty. So, it becomes clear that, ignoring the frailties when they are present can make unreliable estimates of the regression parameters.

5. CONCLUDING REMARKS

This research was conducted based on the assumption that the subjects in each geographical region are concentrated in the center. In reality, this may not be the case. Therefore, it would be worthwhile to the situation where the geostatistical location of each unit is considered. Another aspect needing further investigation is the possibility of allowing the event to recur which has applications in many area. Techniques in Adekpedjou and Niang Dabo (2021) can be used. It will be also interesting to investigate asymptotic properties of the estimators with frailty. Techniques in Murphy (1995) and Parner (1998) can be used. In this work, we have assumed the Cox model as a model for failure time. Others such as Accelerated failure time, additive model, or additive/multiplicative models can be considered in future.

REFERENCES

Adekpedjou, A. and Niang Dabo, S. (2021). Semiparametric estimation with spatially correlated recurrent event data. *To appear*.

Banerjee, S., Wall, M., and Carlin, B. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in minnesota. *Biostatistics*, 4(1):123–142.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144. Update of, and a supplement to, the 1986 original.

Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

Kosorok, M. R., Lee, B. L., and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *The Annals of Statistics*, 32(4):1448–1491.

Li, Y. and Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *J. Amer. Statist. Assoc.*, 101(474):591–603.

Murphy, S. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics*, 23:182–198.

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

Petersen, J. H. (1998). An additive frailty model for correlated life times. *Biometrics*, pages 646–661.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, 79(3):495–512.

SECTION

2. CONCLUSIONS

In this research a method was explored for modeling failure time data in the presence of possible spatial correlation by probit transforming failure times and constructing multivariate Gaussian random fields. Particularly, we account for spatial correlation by including a spatially dependent variance-covariance matrix for the Gaussian random fields, whose elements are a function of Euclidean distance between geographical locations, and those elements represent the spatial dependency among all pairs of geographical locations of the study. In literature, there have been some work to model spatially correlated failure time data which only consider one subject per location. We generalize this setting to have many units per any given geographical location, since this is the more practical situation according to real world applications. For Paper I we considered the case with no frailty. We obtained weighted pairwise semi parametric estimating equations in order to estimate regression and spatial parameters. Our estimators were shown to be consistent and asymptotically normally distributed under infill asymptotic. The simulation study we conducted gave results that were in agreement with developed methods. In Paper II we included frailty variables with a view towards investigating their effects in estimating procedure. Specifically, we assume that each subject in a given geographical region has the same frailty, where as the frailty is different from one geographical region to the other. We then used the same methods that we used in Paper I adapting to the case with frailty. Finally, we conducted a separate simulation study to examine the effects of frailty when modeling failure times.

REFERENCES

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726.

Adekpedjou, A. and Niang Dabo, S. (2021). Semiparametric estimation with spatially correlated recurrent event data. *To appear*.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120.

Banerjee, S. and Carlin, B. (2003). Semi-parametric spatiotemporal analysis. *Environmetrics*, 14:523–535.

Banerjee, S. and Dey, D. K. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Anal.*, 11(2):175–191.

Banerjee, S., Wall, M., and Carlin, B. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in minnesota. *Biostatistics*, 4(1):123–142.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144. Update of, and a supplement to, the 1986 original.

Bronnenberg, B. J. (2005). Spatial models in marketing research and practice. *Appl. Stoch. Models Bus. Ind.*, 21(4-5):335–343.

Chung, K. L., Williams, R. J., and Williams, R. (1990). *Introduction to stochastic integration*, volume 2. Springer.

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(1):141–151.

Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

Diva, U., Banerjee, S., and Dey, D. K. (2007). Modelling spatially correlated survival data for individuals with multiple cancers. *Stat. Model.*, 7(2):191–213.

Diva, U., Dey, D. K., and Banerjee, S. (2008). Parametric models for spatially correlated survival data for individuals with multiple cancers. *Stat. Med.*, 27(12):2127–2144.

Engen, S. (2007). Stochastic growth and extinction in a spatial geometric Brownian population model with migration and correlated noise. *Math. Biosci.*, 209(1):240–255.

Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*. John Wiley & Sons.

Foutz, R. V. (1976). On the consistency of locally asymptotically most stringent tests. *Canad. J. Statist.*, 4(2):211–219.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555.

Gneiting, T., Kleiber, W., and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.*, 105(491):1167–1177.

Goldstein, H. (1995). *Multilevel Statistcal Models*. New York. Edward Arnold.

Gorst (1995). North west leukaemia register. *Technical report*.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics. XLIX. On the Matérn correlation family. *Biometrika*, 93(4):989–995.

Guyon, X. (1995). *Random fields on a network*. Probability and its Applications (New York). Springer-Verlag, New York. Modeling, statistics, and applications, Translated from the 1992 French original by Carenne Ludeña.

Handcock, M. S. and Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, 35(4):403–410.

Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *J. Amer. Statist. Assoc.*, 93(443):1099–1111.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *J. Amer. Statist. Assoc.*, 97(460):965–972.

Hjort, N. L. and Omre, H. (1994). Topics in spatial statistics. *Scand. J. Statist.*, 21(4):289–357. With discussion and a reply by the authors.

Hunt, F. Y. (1978). *GENETIC AND SPATIAL VARIATION IN SOME SELECTION-MIGRATION MODELS*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–New York University.

Jacod, J. (1975). Multivariate point processes: predictable projection, radon-nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31(3):235–253.

Jacod, J. (1976). Un théorème de représentation pour les martingales discontinues. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(3):225–244.

Kosorok, M. R., Lee, B. L., and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *The Annals of Statistics*, 32(4):1448–1491.

Lawless, J. F. and Yilmaz, Y. E. (2011). Semiparametric estimation in copula models for bivariate sequential survival times. *Biom. J.*, 53(5):779–796.

Lele, S. and Taper, M. L. (2002). A composite likelihood approach to (co)variance components estimation. *J. Statist. Plann. Inference*, 103(1-2):117–135. C. R. Rao 80th birthday felicitation volume, Part I.

Li, Y. and Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *J. Amer. Statist. Assoc.*, 101(474):591–603.

Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics*, 58(2):287–297.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Stat. Med.*, 13:2233–2247.

Lindsay, B. G. (1988). Composite likelihood methods. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)*, volume 80 of *Contemp. Math.*, pages 221–239. Amer. Math. Soc., Providence, RI.

Lindsay, B. G., Yi, G. Y., and Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statist. Sinica*, 21(1):71–105.

Matérn, B. (1986). *Spatial variation*, volume 36 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, second edition. With a Swedish summary.

Matheron, G. (1962). *Traité de gostatistique appliquée*. Statistics for Biology and Health. Editions Technip, France.

Murphy, S. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics*, 23:182–198.

Paik, J. and Ying, Z. (2012). A composite likelihood approach for spatially correlated survival data. *Comput. Statist. Data Anal.*, 56(1):209–216.

Pan, C., Cai, B., Wang, L., and Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data. *Comput. Statist. Data Anal.*, 74:198–208.

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

Petersen, J. H. (1998). An additive frailty model for correlated life times. *Biometrics*, pages 646–661.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, 79(3):495–512.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.*, 11(2):453–466.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.*, 93(443):1164–1175.

Stein, M. L. (1999). *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York. Some theory for Kriging.

Varin, C., Hø st, G., and Skare, O. i. (2005). Pairwise likelihood inference in spatial generalized linear mixed models. *Comput. Statist. Data Anal.*, 49(4):1173–1191.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica*, 21(1):5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.*, 84(408):1065–1073.

Yilmaz, Y. E. and Lawless, J. F. (2011). Likelihood ratio procedures and tests of fit in parametric and semiparametric copula models with censored data. *Lifetime Data Anal.*, 17(3):386–408.

VITA

Fathima Zahra Sainul Abdeen attended Badi-Ud-Din Mahmud Girl's College, Kandy from 1994-2004 and Girls' High School, Kandy from 2004-2007. She entered the University of Peradeniya, Sri Lanka in 2008 and completed her bachelor's degree specializing in Statistics in 2014. Soon after graduating she started teaching in the same university and worked there as an instructor for two years. Zahra entered Missouri University of Science and Technology for graduate studies in 2016. She received a Master of Science in Applied Mathematics with Statistics emphasis in May 2020. In May, 2022 she successfully defended her dissertation and received her Doctor of Philosophy in Applied Mathematics with Statistics emphasis from Missouri University of Science and Technology in July 2022.