Statistica Sinica **19** (2009), 1077-1093

SPATIAL LINEAR MIXED MODELS WITH COVARIATE MEASUREMENT ERRORS

Yi Li[1,3], Haicheng Tang[2] and Xihong Lin[3]

[1]*Dana Farber Cancer Institute,* [2]*American Express and* [3]*Harvard School of Public Health*

Abstract: Spatial data with covariate measurement errors have been commonly observed in public health studies. Existing work mainly concentrates on parameter estimation using Gibbs sampling, and no work has been conducted to understand and quantify the theoretical impact of ignoring measurement error on spatial data analysis in the form of the asymptotic biases in regression coefficients and variance components when measurement error is ignored. Plausible implementations, from frequentist perspectives, of maximum likelihood estimation in spatial covariate measurement error models are also elusive. In this paper, we propose a new class of linear mixed models for spatial data in the presence of covariate measurement errors. We show that the naive estimators of the regression coefficients are attenuated while the naive estimators of the variance components are inflated, if measurement error is ignored. We further develop a structural modeling approach to obtaining the maximum likelihood estimator by accounting for the measurement error. We study the large sample properties of the proposed maximum likelihood estimator, and propose an EM algorithm to draw inference. All the asymptotic properties are shown under the increasing-domain asymptotic framework. We illustrate the method by analyzing the Scottish lip cancer data, and evaluate its performance through a simulation study, all of which elucidate the importance of adjusting for covariate measurement errors.

Key words and phrases: Asymptotic bias, consistency and asymptotic normality, EM algorithmMeasurement error, increasing domain asymptotics, spatial data, structural modeling, variance components.

1. Introduction

Spatial data are common in ecology, environmental health and epidemiology, where sampling units are geographical areas or spatially located individuals (Cressie (1993)). Analysis of spatial data is challenged by the spatial correlation among the observations. Mixed effects models provide a convenient framework to model the spatial correlation using random effects that are assumed to follow some spatial correlation structure, such as the conditional autoregressive (CAR) structure (Yasui and Lele (1997) and Waller, Carlin, Xia and Gelfand (1997)) or

the Matèrn correlation structure (Stein (1999)). Asymptotic theory for spatial linear mixed models was established by Mardia and Marsh (1984).

Spatial data are susceptible to measurement errors in covariates. For example in ecological studies, covariates are often collected from a small survey sample in each area and sample averages are used as surrogates for the true population aggregated values, such as the percentage of smokers in a county (Xia and Carlin (1998)). Measurement errors can be substantial when the areas are small, especially in nutritional ecological studies (Prentice and Sheppard (1995)), where additional measurement errors arise due to inaccuracy in measuring nutrition intakes, such as fat intake, using conventional instruments, and 24 hour food recall. In environmental health studies, the air pollution level, e.g., PM10 or ozone, in an area is difficult to measure and is often approximated by using the distance from a polluted site or by using the measures at a few monitoring sites (Carroll, Chen, George, Li, Newton, Schmiediche and Wang (1997)).

There is a vast literature on modeling measurement error for independent data. For an overview, see Carroll, Ruppert and Stefanski (1995). Several authors have considered measurement error in covariates in generalized linear mixed models for clustered data, such as longitudinal data (Wang and Davidian (1996) and Wang, Lin, Gutierrez and Carroll (1998)). However, only limited work has been done in modeling measurement error in covariates for spatial data. Bernardinelli, Pascutto, Best and Gilks (1997) and Xia and Carlin (1998) accounted for measurement error in covariates using hierarchical models in disease mapping. These authors mainly concentrated on parameter estimation using Gibbs sampling. Little is understood about the theoretical effect of measurement error on the asymptotic biases in regression coefficients and variance components when measurement error is ignored. To our knowledge, our work is the first attempt to understand the theoretical properties of maximum likelihood estimation in spatial measurement error mixed effects models.

We first study the asymptotic bias in the naive estimator when measurement error is ignored. Our results show that ignoring measurement error results in attenuated regression coefficients and inflated variance components. We then proceed by applying the structural modeling approach to make valid maximum likelihood inference by accounting for measurement error. An EM algorithm is proposed to compute the maximum likelihood estimator. The proposed methods are illustrated through an application to the Scottish lip cancer data and their performance is evaluated through a simulation study.

2. The Spatial Linear Mixed Measurement Error Model

Suppose that the data are obtained from n geographical areas with continuous outcome variable Y_i, unobserved true covariate X_i (assumed to be a scalar),

observed X_i-related covariate W_i, and other accurately observed covariates $\mathbf{Z}_i$ at the ith area ($i = 1, \cdots, n$). Conditional on the site-specific random effects b_i that model the spatial correlation, the spatial linear mixed model of Y given X and Z can be written as

$$Y_i = \beta_0 + X_i\beta_x + \mathbf{Z}_i^T\boldsymbol{\beta}_z + b_i + \epsilon_i, \tag{2.1}$$

where the random effect vector $(b_1, \ldots, b_n)$ is $N\{0, \mathbf{V}(\boldsymbol{\theta})\}$ and $\boldsymbol{\theta}$ is a vector of variance components, the residuals ϵ_i are $N(0, \sigma_\epsilon^2)$, and b_i and ϵ_i are independent of each other and of the covariates X and $\mathbf{Z}$.

The covariance matrix $\mathbf{V}(\boldsymbol{\theta})$ models the spatial correlation and admits many choices. For instance, we might parameterize the (i, j)th component of $\mathbf{V}(\boldsymbol{\theta})$ as $V_{ij}(\boldsymbol{\theta}) = \theta\rho(\|s_i - s_j\|)$, where correlation function $\rho(.)$ is an isotropic correlation function that decays as the Euclidean distance $d_{ij} = \|s_i - s_j\|$ between two individuals increases. A widely adopted choice for this correlation function is the Matèrn function $[(2\eta\sqrt{\nu}d)^\nu K_\nu(2\eta\sqrt{\nu}d)]/[2^{\nu-1}\Gamma(\nu)]$, where η measures the correlation decay with the distance and ν is a smoothness parameter, $\Gamma(\cdot)$ is the conventional Gamma function and $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν (see, e.g., Abramowitz and Stegun (1965)). This spatial correlation model is rather general, special cases including the exponential model

$$\rho(d) = \exp(-d) \tag{2.2}$$

when the smoothness parameter $\nu = 0.5$ and the 'decay parameter' $\eta = 1$, and the Gaussian correlation model

$$\rho(d) = \exp(-d^2) \tag{2.3}$$

corresponding to $\nu \to \infty$ and $\eta = 1$ (see, e.g., Waller and Gotway (2004, p.279)). Our theoretical development in the ensuing sections focuses on these two widely used cases of the Matèrn family.

The conditional auto-regressive (CAR) structure is also a popular choice. It has appealing theoretical properties, computational advantages, and attractive interpretation (Cressie (1993)). A common CAR structure takes the form (Yasui and Lele (1997))

$$\mathbf{V} = \theta(\mathbf{I} - \gamma\mathbf{M}\mathbf{Q})^{-1}\mathbf{M} = \theta(\mathbf{M}^{-1} - \gamma\mathbf{Q})^{-1}, \tag{2.4}$$

where $\mathbf{Q} = \{q_{ij}\}$ is an $n \times n$ symmetric matrix; $\mathbf{M}$ is an $n \times n$ diagonal matrix with diagonal elements $1/q_{i+}$, with $q_{i+} = \sum_j q_{ij}$, $-1 < \gamma < 1$ is the spatial dependence parameter that controls the amount of information in an area provided by its neighbors, and θ is a scale parameter. The quantity q_{ij} controls the strength of connection between areas i and j, and often takes value 0 when areas i, j are not

neighbors. When area i and area j are neighbors, a common choice is $q_{ij} = 1$ to reflect equal weights from neighboring areas. Note the flexibility of the CAR structure that allows a more general neighborhood concept than geographical proximity.

In the presence of measurement error we cannot observe X directly, but see instead its error-contaminated version W. The spatial linear mixed measurement error model is completed by assuming an additive measurement error model to relate W and X as

$$W_i = X_i + U_i, \tag{2.5}$$

where U_i is the measurement error and is $N(0, \sigma_u^2)$ independent of the unobserved covariate X_i. Note that the measurement error variance σ_u^2 often needs to be estimated using replicates or a validation data set.

Since the covariate X is unobserved, we use the structural modeling approach in the measurement error literature (Carroll, Ruppert and Stefanski (1995)) by assuming a parametric model for X, and proceed with maximum likelihood estimation. The classical measurement error model often assumes X to be an independent and identically distributed Gaussian random variable. However, since we are dealing with spatial data, it is likely that spatial correlation exists not only in the outcome variable Y, but also in the covariate X. Hence we assume a spatial linear mixed model for the unobserved covariate X,

$$X_i = \alpha_0 + \mathbf{Z}_i^T \boldsymbol{\alpha}_z + a_i + e_i, \tag{2.6}$$

where the random effect vector $(a_1, \ldots, a_n) \sim N\{0, \boldsymbol{\Sigma}(\boldsymbol{\zeta})\}$, $\boldsymbol{\Sigma}(\boldsymbol{\zeta})$ models the spatial correlation among the X_i, and the residuals e_i are independent $N(0, \sigma_e^2)$. We assume the a_i and the e_i are independent of the $\mathbf{Z}_i$. Let $\mathbf{W} = (W_1, \ldots, W_n)^T$, with $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{a}$, $\mathbf{b}$ defined similarly. Note that we allow the spatial correlation structure $\boldsymbol{\Sigma}(\boldsymbol{\zeta})$ among the X_i to be different from the spatial correlation structure $\mathbf{V}(\boldsymbol{\theta})$ among the Y_i. In practice, since X and Y both come from the same area, it is often reasonable to assume that they share the same spatial correlation structure with possibly different parameter $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$.

It follows that the likelihood of the observed data $\mathbf{Y}$, $\mathbf{W}$ conditional on $\mathbf{Z}$ is

$$L(\mathbf{Y}, \mathbf{W}|\mathbf{Z}) = \int L(\mathbf{Y}|\mathbf{X}, \mathbf{Z}) L(\mathbf{W}|\mathbf{X}, \mathbf{Z}) L(\mathbf{X}|\mathbf{Z}) d\mathbf{X}.$$

Since all the conditional distributions inside the integral are Gaussian, the joint distribution of $(\mathbf{Y}, \mathbf{W}|\mathbf{Z})$ has the closed form

$$\ell(\mathbf{Y}, \mathbf{W}|\mathbf{Z}) = -\frac{(2n)}{2} \ln(2\pi) - \frac{1}{2} \ln|\boldsymbol{\Lambda}| - \frac{1}{2} \begin{pmatrix} \mathbf{Y} - \boldsymbol{\mu}_y \\ \mathbf{W} - \boldsymbol{\mu}_w \end{pmatrix}^T \boldsymbol{\Lambda}^{-1} \begin{pmatrix} \mathbf{Y} - \boldsymbol{\mu}_y \\ \mathbf{W} - \boldsymbol{\mu}_w \end{pmatrix}, \tag{2.7}$$

where $\boldsymbol{\mu}_y = (\beta_0 + \beta_x\alpha_0)\mathbf{1} + \mathbf{Z}(\beta_x\boldsymbol{\alpha}_z + \boldsymbol{\beta}_z)$, $\mu_w = \alpha_0\mathbf{1} + \mathbf{Z}\boldsymbol{\alpha}_z$, and

$$\boldsymbol{\Lambda} = \mathrm{cov}(\mathbf{Y}, \mathbf{W}|\mathbf{Z}) = \begin{pmatrix} \beta_x^2\boldsymbol{\Sigma}(\boldsymbol{\zeta}) + \mathbf{V}(\boldsymbol{\theta}) + (\beta_x^2\sigma_e^2 + \sigma_\epsilon^2)\mathbf{I} & \beta_x\{\boldsymbol{\Sigma}(\boldsymbol{\zeta}) + \sigma_e^2\mathbf{I}\} \\ \beta_x\{\boldsymbol{\Sigma}(\boldsymbol{\zeta}) + \sigma_e^2\mathbf{I}\} & \boldsymbol{\Sigma}(\boldsymbol{\zeta}) + (\sigma_e^2 + \sigma_U^2)\mathbf{I} \end{pmatrix},$$

with $\mathbf{I}$ an n-dimensional identity matrix.

3. The Asymptotic Bias Analysis

It is of substantial interest to investigate the effect of measurement error by investigating the bias caused by ignoring measurement error, i.e., simply replacing X in model (2.1) by its error-prone version W. This problem, albeit common in spatial data and cautioned by many authors, is never formally addressed. Specifically, the direction and magnitude of biases in naive estimators obtained by ignoring measurement error are not well understood. The goal of this section is to study their asymptotic biases. Our asymptotic bias analysis shows that ignoring measurement error results in an attenuated regression coefficient estimator and an inflated variance component estimator.

We assume the spatial linear mixed measurement error model (2.1) only contains a single covariate X (no $\mathbf{Z}$) with

$$\begin{aligned} Y_i &= \beta_0 + X_i\beta_x + b_i + \epsilon_i, \\ X_i &= \alpha_0 + a_i + e_i, \end{aligned} \tag{3.1}$$

where the distributions of b_i, ϵ_i, a_i, e_i are the same as those in (2.1) and (2.6). The naive estimators of $(\beta_0, \beta_x, \theta, \sigma_\epsilon^2)$ are obtained by simply replacing X_i with the error-prone observation W_i and fitting

$$Y_i = \beta_{0,\mathrm{naive}} + W_i\beta_{x,\mathrm{naive}} + b_{i,\mathrm{naive}} + \epsilon_{i,\mathrm{naive}}, \tag{3.2}$$

where $b_i \sim N\{0, \mathbf{V}(\theta_{\mathrm{naive}})\}$ and $\epsilon_i \sim N(0, \sigma^2_{\epsilon,\mathrm{naive}})$. Let $\boldsymbol{\mathcal{W}} = (\mathbf{1}, \mathbf{W})$, $\boldsymbol{\beta}_{\mathrm{naive}} = (\beta_{0,\mathrm{naive}}, \beta_{x,\mathrm{naive}})^T$, $\boldsymbol{\Lambda}_{\mathrm{naive}}(\boldsymbol{\vartheta}) = \mathbf{V}(\theta_{\mathrm{naive}}) + \sigma_\epsilon^2\mathbf{I}_n$, and $\boldsymbol{\vartheta}_{\mathrm{naive}} = (\theta_{\mathrm{naive}}, \sigma^2_{\epsilon,\mathrm{naive}})^T \stackrel{def}{=} (\vartheta_1, \vartheta_2)^T$. The naive estimates would be obtained by maximizing the likelihood which ignores measurement error,

$$-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Lambda}_{\mathrm{naive}}| - \frac{1}{2}(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\mathrm{naive}})^T\boldsymbol{\Lambda}_{\mathrm{naive}}^{-1}(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\mathrm{naive}}). \tag{3.3}$$

Specifically, they solve

$$\begin{aligned} \frac{1}{n}\boldsymbol{\mathcal{W}}^T\boldsymbol{\Lambda}_{\mathrm{naive}}^{-1}(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\mathrm{naive}}) = 0 \\ \frac{1}{2n}\Big[(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\mathrm{naive}})^T\frac{\partial\boldsymbol{\Lambda}_{\mathrm{naive}}}{\vartheta_j}\boldsymbol{\Lambda}_{\mathrm{naive}}^{-1}\frac{\partial\boldsymbol{\Lambda}_{\mathrm{naive}}}{\vartheta_j}(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\mathrm{naive}}) \\ -\mathrm{tr}\Big(\boldsymbol{\Lambda}_{\mathrm{naive}}^{-1}\frac{\partial\boldsymbol{\Lambda}_{\mathrm{naive}}}{\vartheta_j}\Big)\Big] = 0. \end{aligned} \tag{3.4}$$

We seek the probability limits of the naive estimates as functions of the true values as $n \to \infty$; with a slight abuse of notation, these are $\boldsymbol{\beta}_{\text{naive}}$ and $\boldsymbol{\vartheta}_{\text{naive}}$.

We resort to the increasing domain asymptotics framework when studying bias, this as opposed to infill asymptotics. Zhang and Zimmerman (2005) compared these two frameworks and found that, for certain consistently estimable parameters of exponential covariograms, approximations corresponding to the two frameworks perform about equally well, but for those parameters that cannot be estimated consistently or are highly correlated, infill asymptotic approximation may be preferable. It is usually difficult to derive infill asymptotic properties, so the increasing domain asymptotic framework is used in this work.

Consider $\boldsymbol{\Lambda}$ in (3.3), which depends on $\boldsymbol{\vartheta} = (\theta, \sigma_\epsilon) \stackrel{def}{=} (\vartheta_1, \vartheta_2)$. Let $\boldsymbol{\Lambda}_i = \partial/\partial\vartheta_i \boldsymbol{\Lambda}(\boldsymbol{\vartheta})$ and $\boldsymbol{\Lambda}_{ij} = \partial^2/\partial\vartheta_i\partial\vartheta_j \boldsymbol{\Lambda}(\boldsymbol{\vartheta})$, where the differentiation is element-wise for $i, j = 1, 2$. Now let $\lambda_1 \le \cdots \le \lambda_n$ be the eigen-values of $\boldsymbol{\Lambda}$, and let those of $\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Lambda}_{ij}$ be λ_k^i and λ_k^{ij} for $k = 1, \ldots, n$, respectively, with $|\lambda_1^i| \le \cdots \le |\lambda_n^i|$ and $|\lambda_1^{ij}| \le \cdots \le |\lambda_n^{ij}|$ for $i, j = 1, 2$. We consider the following modified regularity conditions of Mardia and Marsh (1984).

(c.1) $\limsup \lambda_n < \infty$, $\limsup |\lambda_n^i| < \infty$, $\limsup |\lambda_n^{ij}| < \infty$, for all $i, j = 1, 2$.

(c.2) $||\boldsymbol{\Lambda}_i||^{-2} = O(n^{-(1/2)-\delta})$ for some $\delta > 0$, $i = 1, 2$, $||\mathbf{A}|| = \sqrt{tr(\mathbf{A}^T\mathbf{A})}$.

(c.3) $\mathbf{A} = (a_{ij})_{2\times 2}$ is invertible, where for all $i, j = 1, 2$, $a_{ij} = \{t_{ij}/(t_{ii}t_{jj})^{1/2}\}$ exists and $t_{ij} = tr(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_j)$.

(c.4) $\lim(\boldsymbol{\mathcal{W}}^T\boldsymbol{\mathcal{W}})^{-1} = 0$.

These conditions ensure the growth and convergence of the information matrix from (3.3), which allows the usage of the general results of Sweeting (1980) to guarantee the convergence of the naive estimates. In practice, (c.1) and (c.2) are difficult to verify. However, using some basic matrix norm properties, we show in Appendix A.0 that the common geostatistical models, for example the exponential, Gaussian, and CAR models, satisfy (c.1) and (c.2). Condition (c.3) is an identifiability condition, ensuring that the variance components $(\vartheta_1, \vartheta_2)$ are not linear dependent, which is satisfied in our settings. Condition (c.4) ensures that the observed covariates are not trivial and is satisfied for the measurement error models (2.5) and (2.6). Then if (c.1)−(c.4) hold, required limits exist (Sweeting (1980)) and satisfy the asymptotic equations (Harville (1977)),

$$\lim_{n\to\infty} \frac{1}{n} E\Big\{\boldsymbol{\mathcal{W}}^T \boldsymbol{\Lambda}_{\text{naive}}^{-1}(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\text{naive}})\Big\} = 0,$$

$$\lim_{n\to\infty} \frac{1}{2n}\Big[E\Big\{(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\text{naive}})^T \frac{\partial \boldsymbol{\Lambda}_{\text{naive}}}{\vartheta_j} \boldsymbol{\Lambda}_{\text{naive}}^{-1} \frac{\partial \boldsymbol{\Lambda}_{\text{naive}}}{\vartheta_j}(\mathbf{Y} - \boldsymbol{\mathcal{W}}\boldsymbol{\beta}_{\text{naive}})\Big\} - \text{tr}\Big(\boldsymbol{\Lambda}_{naive}^{-1} \frac{\partial \boldsymbol{\Lambda}_{\text{naive}}}{\vartheta_j}\Big)\Big] = 0, \quad (3.5)$$

where the expectations are taken under the true law of $(\mathbf{Y}, \mathbf{W})$ in (2.1) (omitting $\mathbf{Z}$). In particular, we can calculate the asymptotic biases in the naive regression coefficients $\boldsymbol{\beta}_{naive}$. The result is summarized in Theorem 1 and the proof is given in Appendix A.1 (on-line supplement), which can be found on-line at `http://www.stat.sinica.edu.tw/statistica`.

Theorem 1. (*Asymptotic Biases in the Regression Coefficients*) *Under* (c.1)–(c.4), *the following hold.*

(i) *The probability limit of the naive estimator* $\boldsymbol{\beta}_{\text{naive}}$ *is*

$$\beta_{x,\text{naive}} = \lambda_* \beta_x, \qquad \beta_{0,\text{naive}} = \beta_0 + \alpha_0(1-\lambda_*)\beta_x, \tag{3.6}$$

where

$$\lambda_* = \lim_{n\to\infty} \frac{\text{tr}\Big[\{\mathbf{V}(\theta_{\text{naive}}) + \sigma^2_{\epsilon,\text{naive}}\mathbf{I}\}^{-1}\{\boldsymbol{\Sigma}(\boldsymbol{\zeta}) + \sigma_e^2\mathbf{I}\}\Big]}{\text{tr}\Big[\{\mathbf{V}(\theta_{\text{naive}}) + \sigma^2_{\epsilon,\text{naive}}\mathbf{I}\}^{-1}\{\boldsymbol{\Sigma}(\boldsymbol{\zeta}) + \sigma_e^2\mathbf{I}\}\Big] + \sigma_u^2 \text{tr}\Big[\{\mathbf{V}(\theta_{\text{naive}}) + \sigma^2_{\epsilon,\text{naive}}\mathbf{I}\}^{-1}\Big]}, \tag{3.7}$$

and hence $0 \le \lambda_* \le 1$.

(ii) *If* Y *and* X *have the same spatial covariance structure with different scale parameters,*

$$\mathbf{V}(\boldsymbol{\theta}) = \theta\mathbf{R}, \quad \boldsymbol{\Sigma}(\boldsymbol{\zeta}) = \sigma_\Sigma^2\mathbf{R}, \tag{3.8}$$

where $\mathbf{R}$ *is a known matrix, then* λ_* *in* (3.7) *is*

$$\lambda_* = \lim_{n\to\infty} \frac{\sum_{l=1}^n (\delta_l\sigma_e^2 + \sigma_\Sigma^2)/(\delta_l\sigma^2_{\epsilon,\text{naive}} + \theta_{\text{naive}})}{\sum_{l=1}^n \{\delta_l(\sigma_e^2 + \sigma_u^2) + \sigma_\Sigma^2\}/(\delta_l\sigma^2_{\epsilon,\text{naive}} + \theta_{\text{naive}})}, \tag{3.9}$$

where $\{\delta_l\}$ *are the eigenvalues of* $\mathbf{R}^{-1}$.

(iii) *For regular* (*square*) *grid data and the conditional auto-regressive spatial correlation structure* (2.4) *defined using the adjacent neighborhood spatial correlation structure of* Breslow and Clayton (1993),

$$\lambda_* \ge \frac{\sigma_\Sigma^2 + \sigma_e^2(4+4\gamma)}{\sigma_\Sigma^2 + (\sigma_e^2 + \sigma_u^2)(4+4\gamma)}; \tag{3.10}$$

for regular grid data and an exponent or Gaussian spatial correlation structure,

$$0 \le \lambda_* \le 1 - \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 + 4\sigma_\Sigma^2/(1-e^{-1/\sqrt{2}})^2}. \tag{3.11}$$

The results in Theorem 1 show that ignoring the measurement error causes the regression coefficient estimates to be attenuated. Calculations of the attenuation factor λ_* can be quite complicated in general. Therefore the results in (ii) are particularly useful for numerically computing λ_*, since it avoids the inversion of large matrices. Note that the eigenvalues therein do not depend on data if the spatial dependence parameter γ is known. For grid data, exponential and Gaussian correlation structures are often used. In these cases, (iii) provides a bound of the attenuation factor that can be easily computed.

We state in Theorem 2 the asymptotic bias in the naive variance component $\boldsymbol{\vartheta}_{\text{naive}}$; the proof is given in Appendix A.2 (on-line supplement).

Theorem 2. (*Asymptotic Biases in Variance Components*) *Suppose Y and X have the same spatial covariance structure with different scale parameters as at* (3.8). *Under* (c.1)−(c.4), *the asymptotic limits of the naive estimators of the spatial variance component and the residual variance satisfy,*

$$\theta_{\text{naive}} = \theta + (1-\lambda_*)^2\sigma_\Sigma^2\beta_x^2, \quad \sigma^2_{\epsilon,\text{naive}} = \sigma_\epsilon^2 + \{(1-\lambda_*)^2\sigma_e^2 + \lambda_*^2\sigma_U^2\}\beta_x^2, \qquad (3.12)$$

where λ_* is defined in (3.9).

Theorem 2 shows that when Y and X possess the same spatial covariance structure, a reasonable assumption in practice since they come from the same spatial area, the naive estimators of the spatial variance component and the residual variance both overestimate the corresponding true values. For more general cases when the spatial covariance structure of Y and X differ, the asymptotic limits of the naive estimators are difficult to calculate, and no analytic expressions are available.

The asymptotic relative biases in the naive estimators of the regression coefficients and the variance components, assuming the adjacent neighborhood spatial correlation structure, is illustrated in Figure 1. Since the computation of λ_* involves $n \to \infty$, we approximate λ_* with $n = 1,024$ on a 32×32 lattice. The spatial dependence parameter γ in (2.4) are taken as $\gamma = 0.2$ and $\gamma = 0.95$. The regression coefficient is $\beta_x = 0.5$, variance components are $\theta = 1$ and $\sigma_\epsilon^2 = 0.3$. The parameters in the X models are $\alpha_0 = 1.4$, $\sigma_\Sigma^2 = 1.2$, and $\sigma_e^2 = 0.5$. We iteratively calculate λ_* using (3.9) and (3.12). In our experience, convergence is often achieved within five iterations. Then we obtain the expected naive estimates from (3.6) and (3.12). The bias curves for the naive estimates of β_x and θ are plotted as a function of the measurement error variance σ_u^2. It should be noted that the bias curves in fact correspond to the finite sample *exact* bias.

Figure 1 shows that the naive estimate of the regression coefficient β_x is attenuated, while the naive estimate of the variance component θ is inflated. The biases increase with the measurement error variance σ_u^2, but decrease with

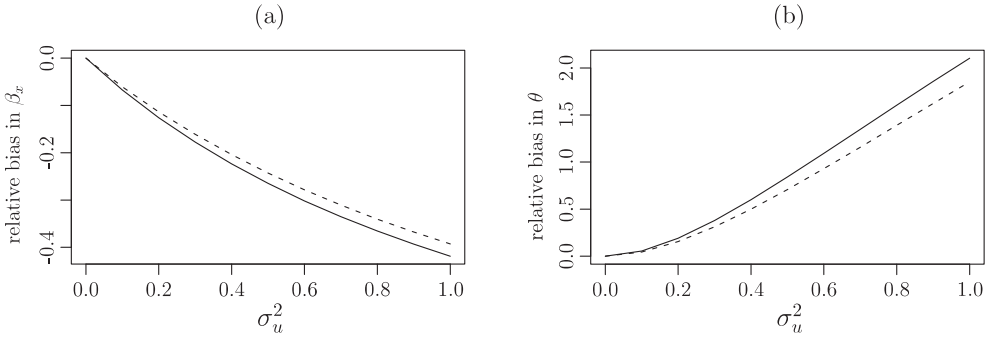

Figure 1. Asymptotic relative biases in the naive estimates of β_x and θ. The CAR spatial covariance structure with spatial dependence parameter $\gamma = 0.2$ and 0.95 was used. The true parameter values were $\beta_x = .5$, $\theta = 1$, $\sigma_\epsilon^2 = 0.5$, $\sigma_\Sigma^2 = 1.2$, and $\sigma_e^2 = 0.5$. Variance of measurement error σ_u^2 varied between 0 and 1.0. The two curves in each plot correspond to the spatial dependence parameters $\gamma = 0.2$ and $\gamma = 0.95$.

the spatial dependence parameter γ. The reason for the latter phenomenon is explained by the fact that stronger dependence implies that neighbor areas can provide more information, and hence the estimates are more resistant to the effect of measurement error.

4. Maximum Likelihood Estimation

We consider the large sample results for the maximum likelihood estimator for the spatial linear mixed measurement error models (2.1), (2.5) and (2.6). In particular, we show for some commonly used spatial models, the MLEs are consistent and asymptotically normal. To proceed, we assume mild regularity conditions on the parameter space and the observed covariate $\mathbf{Z}$.

(d.1) The unknown parameters $\mathbf{\Omega}$ in (2.1), (2.5) and (2.6) lie in a compact set of an Euclidean space.

(d.2) Let $\tilde{\mathbf{Z}} = (\mathbf{1}, \mathbf{Z})$, where $\mathbf{1}$ is an $n \times 1$ vector of 1's, $\lim n^{-1}\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}} = \mathcal{Z}_0$ in probability, where $\mathcal{Z}_0$ is a positive definite matrix.

It follows that, for the common geostatistical models, such as the exponential, Gaussian and CAR models, the maximum likelihood estimator is consistent and asymptotically normal, as summarized in the following theorem. The proof is deferred to Appendix A.3 (on-line supplement).

Theorem 3. (*Consistency and Asymptotic Normality of MLEs*) *Let* $\mathbf{\Omega}_0$ *be the true unknown parameters in* (2.1), (2.5) *and* (2.6) *and* $\widehat{\mathbf{\Omega}}$ *be its maximum likelihood estimator. Suppose that* Y *and* X *have the exponential, Gaussian or CAR*

(2.4) *spatial covariance structure on regular grid. Then, under* (d.1) *and* (d.2), $\widehat{\mathbf{\Omega}}$ *is consistent and* $\mathbf{\Gamma}^{1/2}(\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0) \to N(0, \mathbf{I}_p)$ *in distribution, where* $\mathbf{\Gamma}^{1/2}$ *is the Cholesky decomposition of* $\mathbf{\Gamma} \stackrel{def}{=} E_{\mathbf{\Omega}_0}\{-\partial^2\ell/\partial\mathbf{\Omega}\partial\mathbf{\Omega}^T\}$, $\mathbf{\Omega}_0$ *is the truth,* ℓ *is as defined in* (2.7), *and* $\mathbf{I}_p$ *is the identity matrix of dimension of* p, *the dimension of* $\mathbf{\Omega}_0$.

Theorem 3 does not require X and Y to have the same correlation structure, but, since X and Y both come from the same area, it may be reasonable to assume that they do. In such a situation we propose an EM algorithm to compute the MLEs; in particular, we assume the spatial covariance structures of the random effects $\mathbf{b}$ and $\mathbf{a}$ take the same form (3.8) with different scale parameters. The EM algorithm for a general spatial covariance structure is similar. The complete data are $(\mathbf{Y}, \mathbf{W}, \mathbf{X}, \mathbf{Z}, \mathbf{b}, \mathbf{a})$, where $(\mathbf{Y}, \mathbf{W}, \mathbf{Z})$ are observed data and $\mathbf{X}$, $\mathbf{b}$, and $\mathbf{a}$ are missing data. The complete data loglikelihood is

$$\begin{aligned}\ell(\mathbf{Y}, \mathbf{W}, \mathbf{X}, \mathbf{b}, \mathbf{a}|\mathbf{Z}) = &-\frac{n}{2}\log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\|\mathbf{Y} - \beta_0\mathbf{1} - \beta_x\mathbf{X} - \mathbf{Z}\boldsymbol{\beta}_z - \mathbf{b}\|^2 - \frac{n}{2}\log(\theta)\\ &-\frac{1}{2\theta}\mathbf{b}^T\mathbf{R}^{-1}\mathbf{b} - \frac{n}{2}\log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\|\mathbf{W} - \mathbf{X}\|^2 - \frac{n}{2}\log(\sigma_e^2)\\ &-\frac{1}{2\sigma_e^2}\|\mathbf{X} - \alpha_0\mathbf{1} - \mathbf{Z}\boldsymbol{\alpha}_z - \mathbf{a}\|^2 - \frac{n}{2}\log(\sigma_\Sigma^2) - \frac{1}{2\sigma_\Sigma^2}\mathbf{a}^T\mathbf{R}^{-1}\mathbf{a},\end{aligned}$$

where $\|\cdot\|$ denotes the square norm.

Let $\tilde{\mathbf{X}} = (\mathbf{1}, \mathbf{X}, \mathbf{Z})$, $\tilde{\mathbf{Z}} = (\mathbf{1}, \mathbf{Z})$, $\boldsymbol{\beta} = (\beta_0, \beta_x, \boldsymbol{\beta}_z^T)^T$, and $\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}_z^T)^T$. At the $(t+1)$th step, let the estimator of $\boldsymbol{\beta}$ be $\hat{\boldsymbol{\beta}}^{(t+1)}$ and the estimator of $\boldsymbol{\alpha}$ be $\boldsymbol{\alpha}^{(t+1)}$, and define the variance component estimates similarly. In the M step, we update the regression coefficients

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{(t+1)} &= E(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}|\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)})^{-1}E(\tilde{\mathbf{X}}^T(\mathbf{Y} - \mathbf{b})|\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)})\\ \hat{\boldsymbol{\alpha}}^{(t+1)} &= (\tilde{\mathbf{Z}}^T\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^T E(\mathbf{X} - \mathbf{a}|\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)}),\end{aligned}$$

where $E(\cdot|\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)})$ is the expectation conditional on the observed data $(\mathbf{Y}, \mathbf{W})$ with all parameters taking the values of the current estimates $\hat{\xi}^{(t)}$. We update the variance components by

$$\begin{aligned}\hat{\theta}^{(t+1)} &= \frac{1}{n}E(\mathbf{b}^T\mathbf{R}^{-1}\mathbf{b}|\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)}),\\ \hat{\sigma}_\epsilon^{2(t+1)} &= \frac{1}{n}E(\|\mathbf{Y} - \hat{\beta}_x^{(t+1)}\mathbf{X} - \mathbf{Z}\hat{\boldsymbol{\beta}}_z^{(t+1)} - \mathbf{b}\|^2\ |\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)}),\\ \hat{\sigma}_\Sigma^{2(t+1)} &= \frac{1}{n}E(\mathbf{a}^T\mathbf{R}^{-1}\mathbf{a}|\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)}),\\ \hat{\sigma}_e^{2(t+1)} &= \frac{1}{n}E(\|\mathbf{X} - \hat{\alpha_0}^{(t+1)}\mathbf{1} - \mathbf{Z}\boldsymbol{\alpha}_z^{(t+1)} - \mathbf{a}\|^2\ |\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \hat{\xi}^{(t)}).\end{aligned}$$

In the E step, we compute the conditional expectations that appear in the above equations. The closed-form expressions of these conditional expressions are derived and can be found in Appendix A.4 (on-line supplement). These steps can be easily implemented since all the quantities involved have closed form and no numerical integration is needed. Finally, the standard errors of the maximum likelihood estimates can be obtained by inverting the Fisher information matrix.

5. Simulation Study

Our simulation study aims at evaluating the finite sample performance of the naive estimates obtained by ignoring the measurement error and the maximum likelihood estimates obtained by accounting for the measurement error. We took the data to be on a regular grid. We considered the Y model (2.1) with a single covariate X. We assumed the adjacent neighborhood CAR spatial correlation structure (2.4) for both the random effects $\{b_i\}$ and $\{a_i\}$ in the Y and X models, neighbors being defined as the four adjacent areas for each location except for those on the edge. The weight q_{ij} was set to be 1 if areas i and j are neighbors and 0 otherwise. The spatial dependence parameter was $\gamma = 0.95$, mimicking what was obtained in the data example in the next section. The unobserved covariate X was generated under model (2.6) with mean 1.4 and variance components 1.2 and 0.3, respectively, for the spatial covariance and residual error term. The observed error-contaminated version W was generated by adding Gaussian noise with variance $\sigma_u^2 = 0.50$ to X. To generate the outcome variable Y, the regression coefficients were taken as $(\beta_0, \beta_x)^T = (0.0, 2.0)^T$, and the variance components for the spatial covariance and residual error term were taken as 1.0 and 0.5. For each generated data set, we computed the naive estimates that ignored the measurement error and the maximum likelihood estimates that accounted for the measurement error using the EM algorithm. We varied the grid size to be $7(n = 7 \times 7)$, $10(n = 10 \times 10)$ and $20(n = 20 \times 20)$. The averages and variances of the estimates from 500 replications are given in Table 1.

We next examined the performance of the MLE when the spatial correlation structure was specified to be the exponential model, as well as the Gaussian model. The locations of subjects were sampled uniformly over region $[0, \sqrt{n}]^2$, where n is the number of subjects. We set $n = 49, 100, 400$ in our simulations. The results are documented in Tables 2 and 3.

All the results (Tables 1−3) show that the naive estimate of β_x is attenuated while the naive estimates for θ and σ_ϵ^2 are inflated, agreeing with our asymptotic bias analysis. The maximum likelihood estimates computed using the EM algorithm, on the other hand, performed very well. The mean of the estimates of the regression coefficients and the variance components were very close to the corresponding true values. As expected, there was a bias-variance tradeoff. The

Table 1. Results of a simulation study from 500 replications under the CAR model. A regular 20×20 grid design and an adjacent neighborhood covariance structure with $\gamma = 0.95$ were used. The true parameters were $\beta_0 = 0$, $\beta_x = 0.5$, $\theta = 1\sigma_\epsilon^2 = 0.3$, $\sigma_\Sigma^2 = 1.2$, $\sigma_e^2 = 0.5$ and $\alpha_0 = 1.4$. The measurement error variance was $\sigma_u^2 = 0.5$. Inside the brackets are estimated standard errors.

Sample Size	Para-meter	Mean of naive est	model based Var. of naive est		MSE	Mean of MLE	model based Var. of MLE		MSE
49	β_0	0.221	0.095	(0.161)	0.200	-0.045	0.150	(0.210)	0.210
	β_x	0.326	0.011	(0.012)	0.040	0.513	0.030	(0.043)	0.043
	θ	1.141	0.355	(0.383)	0.401	1.041	0.283	(0.367)	0.368
	σ_ϵ^2	0.461	0.054	(0.063)	0.090	0.374	0.045	(0.056)	0.061
100	β_0	0.261	0.062	(0.072)	0.14	-0.042	0.079	(0.090)	0.09
	β_x	0.323	0.005	(0.005)	0.036	0.507	0.0157	(0.0157)	0.016
	θ	1.066	0.222	(0.265)	0.269	0.957	0.209	(0.239)	0.240
	σ_ϵ^2	0.399	0.0268	(0.033)	0.043	0.326	0.0312	(0.029)	0.030
400	β_0	0.247	0.0160	(0.0154)	0.076	0.0032	0.020	(0.022)	0.022
	β_x	0.318	0.0012	(0.0012)	0.034	0.503	0.0035	(0.0035)	0.0035
	θ	1.015	0.067	(0.073)	0.073	0.989	0.062	(0.072)	0.072
	σ_ϵ^2	0.376	0.0068	(0.0069)	0.012	0.304	0.0068	(0.0072)	0.0072

MLEs effectively eliminated the biases in the naive estimators but had larger variances. As an overall measure of performance using the MSE, the MLEs had smaller MSEs than the naive estimators. The MSE gain was more apparent as n increased.

Finally, to compare the empirical results with our theoretical asymptotic bias analysis results, we computed the theoretical values of the naive estimate using the results in Theorems 1 and 2. For example, under the CAR model with $\gamma = 0.95$, these values were 0.254, 0.318, 1.039, 0.367 for β_0, β_x, θ and σ_ϵ^2, compared with 0.247, 0.318, 1.027, 0.376 of the average naive estimates based on 500 simulations for grid size 20 ($n = 400$) (see Table 1). Hence, our theoretical values do match with our simulation results.

6. Analysis of Scottish Lip Cancer Incidence Data

The Scottish lip cancer incidence data were collected in each of the 56 counties of Scotland (Breslow and Clayton (1993)). For each county, the number of lip cancer cases among males from 1975-1980 and the percentage of AFF employment in all employed population were reported. Earlier analysis found that the rates were higher in counties with higher proportion of the population employed in agriculture, forestry, and fishing (AFF) – the professions that require working outdoors. This observation reflects the biological plausible causal relationship between ultraviolet rays and lip cancer. Breslow and Clayton (1993)

Table 2. Results of a simulation study from 500 replications under the Gaussian model. The locations of subjects were sampled uniformly over the region $[0, \sqrt{n}]^2$, where n is the number of subjects. The true parameters were $\beta_0 = 0$, $\beta_x = 0.5$, $\theta = 1\sigma_\epsilon^2 = 0.3$, $\sigma_\Sigma^2 = 1.2$, $\sigma_e^2 = 0.5$ and $\alpha_0 = 1.4$. The measurement error variance was $\sigma_u^2 = 0.5$. Inside the brackets are estimated standard errors.

Sample Size	Parameter	Mean of naive est	model based Var. of naive est		MSE	Mean of MLE	model based Var. of MLE		MSE
49	β_0	0.276	0.107	(0.110)	0.186	-0.038	0.136	(0.191)	0.192
	β_x	0.320	0.010	(0.012)	0.044	0.532	0.031	(0.038)	0.039
	θ	1.076	0.168	(0.180)	0.185	0.970	0.150	(0.156)	0.156
	σ_ϵ^2	0.365	0.019	(0.018)	0.022	0.300	0.020	(0.021)	0.021
100	β_0	0.218	0.043	(0.041)	0.119	-0.022	0.055	(0.057)	0.057
	β_x	0.338	0.0057	(0.0056)	0.032	0.512	0.0128	(0.0134)	0.013
	θ	1.049	0.077	(0.078)	0.081	0.982	0.071	(0.066)	0.066
	σ_ϵ^2	0.373	0.0153	(0.0158)	0.0211	0.298	0.0150	(0.0150)	0.0150
400	β_0	0.176	0.0076	(0.0084)	0.039	0.034	0.009	(0.006)	0.006
	β_x	0.369	0.0015	(0.0018)	0.0189	0.498	0.002	(0.002)	0.002
	θ	1.027	0.026	(0.022)	0.022	1.019	0.023	(0.024)	0.024
	σ_ϵ^2	0.386	0.013	(0.014)	0.021	0.303	0.0108	(0.0108)	0.011

applied spatial mixed models to study the association between the percentage of the AFF employment and the lip cancer incidence. However the exposure of main interest is the exposure to sunlight, a known risk factor for lip cancer. The AFF employment variable serves as a surrogate for the degree of exposure to sunlight. Since we mainly focused on the association between lip cancer and the exposure to sunlight, we need to account for the measurement error in using the AFF employment variable to measure the degree of exposure to sunlight.

Breslow and Clayton (1993) modeled the standardized morbidity ratios calculated by dividing the observed number of cancer cases by the age-adjusted expected cancer cases using a Poisson regression model. To apply our methodology, we first took a square root transformation of the observed SMR; the transformed SMR approximated a normal distribution well, which was verified using the Shapiro-Wilks test. We applied the spatial linear mixed measurement error model to account for the measurement error.

Following Breslow and Clayton (1993), we assumed the adjacent neighborhood spatial correlation structure for the square-root transformed SMR. These authors also noted that the covariate, the percentage of the AFF employment, exhibited the same spatial aggregation as the SMR. We hence assumed the same spatial correlation structure with a different scale parameter for the AFF variable.

Table 3. Results of a simulation study from 500 replications under the exponential model. The locations of subjects were sampled uniformly over the region $[0, \sqrt{n}]^2$, where n is the number of subjects. The true parameters were $\beta_0 = 0$, $\beta_x = 0.5$, $\theta = 1\sigma_\epsilon^2 = 0.3$, $\sigma_\Sigma^2 = 1.2$, $\sigma_e^2 = 0.5$ and $\alpha_0 = 1.4$. The measurement error variance $\sigma_u^2 = 0.5$. Inside the brackets are estimated standard errors.

Sample Size	Parameter	Mean of naive est	model based Var. of naive est	MSE	MLE	Mean of model based Var. of MLE	MSE
49	β_0	0.223	0.145 (0.179)	0.228	-0.030	0.187 (0.239)	0.239
	β_x	0.339	0.010 (0.012)	0.038	0.528	0.0270 (0.0323)	0.033
	θ	1.079	0.220 (0.268)	0.274	0.966	0.191 (0.235)	0.236
	σ_ϵ^2	0.408	0.0535 (0.0670)	0.079	0.364	0.052 (0.063)	0.067
100	β_0	0.216	0.062 (0.073)	0.108	-0.021	0.071 (0.079)	0.079
	β_x	0.350	0.0057 (0.0058)	0.028	0.509	0.012 (0.013)	0.013
	θ	1.037	0.114 (0.130)	0.131	0.982	0.113 (0.134)	0.134
	σ_ϵ^2	0.399	0.037 (0.042)	0.038	0.314	0.036 (0.037)	0.037
400	β_0	0.168	0.009 (0.010)	0.038	-0.002	0.011 (0.011)	0.011
	β_x	0.374	0.0015 (0.0015)	0.017	0.495	0.0026 (0.0026)	0.0026
	θ	1.037	0.040 (0.030)	0.031	1.010	0.037 (0.045)	0.045
	σ_ϵ^2	0.387	0.022 (0.014)	0.021	0.298	0.020 (0.024)	0.024

The analysis results are given in Table 4. The naive analysis showed a strong effect of the AFF employment on the SMR ($\beta_x = 0.139$ and SE=0.091), and the spatial correlation seemed to dominate in the total variation ($\hat{\theta} = 0.310$, $\hat{\sigma}_\epsilon^2 = 0.0389$). We next considered the spatial linear mixed measurement error model to account for the measurement error in the AFF employment. Since no validation data set was available, the measurement error variance σ_u^2 could not be estimated directly from the data. We fit a linear random intercept CAR model on W. This allowed us to estimate the sum of σ_e^2 and σ_u^2 as 0.041. We then did a sensitivity analysis by varying σ_u^2 from 0.0, naive analysis, to moderate measurement error, $\sigma_u^2 = 0.020$, to severe measurement error, $\sigma_u^2 = 0.035$. The estimates of the dependence parameter γ were 0.922 when $\sigma_u^2 = 0$, 0.928 when $\sigma_u^2 = 0.02$ and 0.932 when $\sigma_u^2 = 0.035$, all of which were close to the estimate of 0.93 obtained by Yasui and Lele (1997), and indicated a strong spatial dependence. Second, all the analyses indicated that working outdoors was associated with the risk of lip cancer. Third, ignoring measurement error did attenuate the regression coefficient estimates. As σ_u^2 increased, the estimates of the regression coefficients became larger. For example, the estimate of the coefficient of 'AFT', with estimated standard error in brackets, increased from 0.132 (0.093) when $\sigma_u^2 = 0$, to 0.153 (0.099) when $\sigma_u^2 = 0.02$, and to 0.172 (0.104) when $\sigma_u^2 = 0.035$, while the variance component for the spatial correlation part was estimated as 0.434 (0.245) when $\sigma_u^2 = 0$, 0.414(0.234) when $\sigma_u^2 = 0.02$, and 0.394 (0.228) when

Table 4. Sensitivity analysis of Scottish Lip Cancer Incidence Data: Outcome variable is the square root of SMR; the covariate is AFF/10. The measurement error variance varied between 0 (naive), 0.02 (moderate) and 0.035 (severe).

Estimate ± standard error			
Parameter	naive $\sigma_u^2 = 0.0$	moderate $\sigma_u^2 = 0.02$	severe $\sigma_u^2 = 0.035$
γ	0.922 ± 0.072	0.928 ± 0.044	0.932 ± 0.043
β_0	0.939 ± 0.164	0.923 ± 0.168	0.908 ± 0.171
β_x	0.132 ± 0.093	0.153 ± 0.099	0.172 ± 0.104
θ	0.434 ± 0.245	0.414 ± 0.234	0.394 ± 0.228
σ_ϵ^2	0.017 ± 0.045	0.021 ± 0.044	0.024 ± 0.044
σ_Σ^2		1.258 ± 0.258	1.183 ± 0.256
σ_e^2		0.0001 ± 0.0005	0.0001 ± 0.0003
α_0		0.8033 ± 0.258	0.8030 ± 0.258

$\sigma_u^2 = 0.035$. These results indicated that accounting for measurement error increased the magnitude of the estimated effects of 'AFT' while it decreased the overestimation of the spatial variance component.

7. Discussion

In this paper we have proposed spatial linear mixed measurement error models to account for covariate measurement error and spatial correlation in spatial data. Our asymptotic bias analysis shows that, by ignoring the measurement error, the naive estimators of the regression coefficients are attenuated and the naive estimators of the variance components are inflated. We give formulae for calculating these biases for a general case, and provide simplified forms or bounds for some commonly-used spatial correlation structures. Our numerical calculation also shows that the biases are related to the spatial dependence parameter γ for an adjacent neighborhood structure.

We have developed a structural modeling approach to accounting for the covariate measurement error in spatial data, where spatial linear mixed models are assumed for both the outcome and the unobserved covariate, and an additive model is assumed for the observed error-prone covariate. An EM algorithm is developed to compute the maximum likelihood estimate. Our simulation study shows that the maximum likelihood estimator works well in finite samples and appropriately corrects for the bias in the naive estimator. We also find that the maximum likelihood estimators correct the biases in naive estimators, but are associated with larger variances.

On the computational side, our algorithm requires operations on matrices of large size, including inversion of large matrices. We alleviate the computational

burden by diagonalizing the matrices simultaneously. Since the sizes of the matrices involved increase rapidly with the grid size of the spatial areas, many operations on these matrices are needed in each EM iteration. These cause problems in handling large data sets with the EM algorithm. Here it might be more convenient to adopt an MCMC algorithm, especially if one uses the conditional autoregressive spatial covariance structure.

Our structural modeling approach, where a parametric model is assumed for the unobserved covariate X, might be sensitive to misspecification of the distribution of X. An alternative estimation in the measurement error literature is functional modeling, such as SIMEX (Carroll et al. (1995)), which makes no distributional assumption on X. However it can be less efficient than the MLE when the distribution of X is correctly specified. It is of interest in future research to compare these two approaches in terms of their robustness and efficiency.

We have concentrated on Gaussian spatial outcomes. Work is underway to extend the results to non-Gaussian spatial outcomes, with measurement error in the covariate, within the framework of spatial generalized linear mixed models (e.g., Diggle, Moyeed and Tawn (1998)).

Acknowledgement

The authors wish to thank the Editor, an Associate Editor and two anonymous referees for their insightful suggestions, which significantly improved this work.

References

Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Function.* Dover, New York.

Bernardinelli, L., Pascutto, C., Best, N. G. and Gilks, W. R. (1997). Disease mapping with errors in covariates. *Statist. Medicine* **16**, 741-752.

Breslow, N. E. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88**, 9-25

Carroll, R. J., Chen, R., George, E. I., Li, T. H., Newton, H. J., Schmiediche, H. and Wang, N. (1997). Ozone exposure and population density in Harris county, Texas. *J. Amer. Statist. Assoc.* **92**, 392-415.

Carroll, R., Ruppert, D. and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models.* Champman and Hall, New York.

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

Diggle, P., Moyeed, R. and Tawn, J. (1998). Model-based geostatistics. *Appl. Statist.* **47**, 299-350.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72**, 320-340.

Lehman, E. L. and Casella, G. (1998). *Theory of Point Estimation.* 2nd edition. Springer, New York.

Mardia, K. V. and Marsh, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.

Prentice, R. and Sheppard, L. (1995). Aggregate data studies of disease risk factors. *Biometrika* **82**, 113-125.

Schervish, M. (1995). *Theory of Statistics.* Springer, New York.

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

Sweeting, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8**, 1375-1381.

Waller, L., Carlin, B. P., Xia, H. and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *J. Amer. Statist. Assoc.* **92**, 607-617.

Waller, L. and Gotway, C. (2004). *Applied Spatial Statistics for Public Health Data.* Wiley, New York.

Wang, N. and Davidian, M. (1996). Effects of covariate measurement error on nonlinear mixed effects models. *Biometrika* **83**, 801-812.

Wang, N., Lin, X., Gutierrez, R. G. and Carroll, R. J. (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *J. Amer. Statist. Assoc.* **93**, 249-261.

Wilkinson, J. H. (1965). *The Algebraic Eigenvalue Problem.* Oxford University, Oxford.

Xia, H. and Carlin, B. P. (1998). Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Statist. Medicine* **17**, 2025-2043.

Yasui, Y. and Lele, S. (1997). A regression methods for spatial disease rates: an estimating function approach approach. *J. Amer. Statist. Assoc.* **92**, 21-32.

Zhang, H. and Zimmerman, D. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92**, 921-936.

Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, 44 Binney St, Boston, MA 02115, U.S.A.
Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave, Boston, MA 02115, U.S.A.

E-mail: yili@hsph.harvard.edu

American Express, 200 Vesey Street, New York, NY 10285, U.S.A.

E-mail: haicheng.k.tang@aexp.com

Department of Biostatistics, Harvard School of Public Health, 655 Huntington Ave, Boston, MA 02115, U.S.A.

E-mail: xlin@hsph.harvard.edu

(Received July 2007; accepted March 2008)