Regression With Spatially Misaligned Data

Lisa Madsen and David Ruppert
Department of Statistics
Oregon State University
44 Kidder Hall, Corvallis OR 97331
E-mail: lmadsen@stat.orst.edu

Keywords: Spatial regression, Maximum likelihood, Misaligned data

Abstract

Suppose $\mathbf{X}(s)$ and $\boldsymbol{\epsilon}(s)$ are stationary spatially autocorrelated Gaussian processes and $\mathbf{Y}(s) = \beta_0 + \beta_1 X(s) + \epsilon(s)$ for any location s. Our problem is to estimate the β's, particularly β_1, when $\mathbf{X}$ and $\mathbf{Y}$ are not necessarily observed in the same location. This situation may arise when the data are recorded by different agencies or when there are missing data values.

A natural but naïve approach is to predict ("krige") the missing $\mathbf{X}$'s at the locations $\mathbf{Y}$ is observed, and then use least squares to estimate ("regress") the β's as if these $\mathbf{X}$'s were actually observed. This krige-and-regress estimator is consistent, even when the spatial covariance parameters are estimated. If we use it as a starting value for a Newton-Raphson maximization of the likelihood, the resulting maximum likelihood estimator is asymptotically efficient. We can then use an information-based variance estimator for inference.

1 Introduction

Much of the work in spatial statistics has been in prediction, with kriging (or best linear unbiased prediction) the ubiquitous technique. In this paper, we investigate the problem of regressing one Gaussian spatial process on another when the two processes may be observed at different locations, i.e. the data are misaligned. Our goal is to estimate the regression parameters, especially the slope parameter, and obtain standard errors for the estimates.

We develop two estimation procedures. The first is to krige the predictor process to align the data, then use the kriged values as if they had actually been observed in a standard regression analysis. Call this procedure Krige and Regress (KR). We obtain an expression for the variance of the KR slope parameter estimate, and show under what conditions the KR estimate of the regression parameters is consistent.

The second approach is to use maximum likelihood (ML). Because KR yields consistent estimates, if we use these as initial values in a Newton-Raphson maximization of the log-likelihood, we obtain consistent and fully efficient ML estimates. This justifies an information-based variance estimate.

1.1 Notation and Assumptions

Let $\mathbf{X}$ and $\boldsymbol{\epsilon}$ be second-order stationary spatial Gaussian processes on $S \subseteq \mathbb{R}^d$ with $E[\mathbf{X}(s)] = \mu_X$ and $E[\boldsymbol{\epsilon}(s)] = 0$ for $s \in S$. Assume $\mathbf{X}$ and $\boldsymbol{\epsilon}$ are independent.

We wish to estimate the slope parameter β_1 in the model

$$\mathbf{Y}(s) = \beta_0 + \mathbf{X}(s)\beta_1 + \boldsymbol{\epsilon}(s) \tag{1}$$

for any location s.

Suppose we observe $Y = \left[\mathbf{Y}(s_1) \dots \mathbf{Y}(s_n)\right]'$ and $W = \left[\mathbf{X}(t_1) \dots \mathbf{X}(t_m)\right]'$ where some or all of the s_i and t_i may overlap. Let $X = \left[\mathbf{X}(s_1) \dots \mathbf{X}(s_n)\right]'$ denote the possibly unobserved values of $\mathbf{X}$ at the locations where $\mathbf{Y}$ is observed. Let $\epsilon = \left[\boldsymbol{\epsilon}(s_1) \dots \boldsymbol{\epsilon}(s_n)\right]'$ denote the error vector. For brevity, denote the elements of the vectors Y, W, X, and ϵ as

$$\begin{aligned} Y &= \begin{bmatrix} Y_1 \dots Y_n \end{bmatrix}' \\ W &= \begin{bmatrix} W_1 \dots W_m \end{bmatrix}' \\ X &= \begin{bmatrix} X_1 \dots X_n \end{bmatrix}' \\ \epsilon &= \begin{bmatrix} \epsilon_1 \dots \epsilon_n \end{bmatrix}'. \end{aligned}$$

Let Σ_X, Σ_W, and Σ_ϵ denote the covariance matrices of the vectors X, W, and ϵ respectively, and let Σ_{XW} be the $n \times m$ matrix of covariances $\text{cov}(X_i, W_j) = \text{cov}[\mathbf{X}(s_i), \mathbf{X}(t_j)]$.

Assume the ijth element of the matrices Σ_X, Σ_W, and Σ_{XW} is given by $C_X(\theta_X, h_{ij})$, where h_{ij} is the distance between the ith and jth points and C_X is a covariance function. The definition of $h_{i,j}$ will differ for the three matrices. For Σ_W, $h_{i,j} = \|t_i - t_j\|$, for Σ_X, $h_{i,j} = \|s_i - s_j\|$, and for Σ_{XW}, $h_{i,j} = \|s_i - t_j\|$. In this paper, we use the exponential covariogram function, so the ijth element of Σ_X is

$$\text{cov}(X(s_i), X(s_j)) = \begin{cases} \theta_{X,1} + \theta_{X,2} & h_{ij} = 0 \\ \theta_{X,2} \exp(-h_{ij}\theta_{X,3}) & s_i \neq s_j \end{cases}$$

where $h_{ij} = \|s_i - s_j\|$ and $\theta_X = \begin{bmatrix} \theta_{X,1} & \theta_{X,2} & \theta_{X,3} \end{bmatrix}'$ is a vector of positive covariance parameters. Similarly, assume Σ_ϵ is determined by the exponential covariance and parameters $\theta_\epsilon = (\theta_{\epsilon,1}, \theta_{\epsilon,2}, \theta_{\epsilon,3})$.

For a covariance matrix Σ, determined by θ as described above, we will use the notation $\Sigma(\theta)$ to emphasize that the matrix is a function of θ.

Also, for integers $p, q > 0$, let $\mathbf{1}_{p\times q}$ denote a $p \times q$ matrix of ones, and let I_p be the identity matrix of order p.

Since $\mathbf{X}$ and $\boldsymbol{\epsilon}$ are independent Gaussian processes, Y, X, and W are jointly normally distributed:

$$\begin{pmatrix} Y \\ X \\ W \end{pmatrix} \sim N_{2n+m} \left(\begin{pmatrix} (\beta_0 + \beta_1 \mu_X)\mathbf{1}_{n\times 1} \\ \mu_X \mathbf{1}_{(n+m)\times 1} \end{pmatrix}, \begin{bmatrix} \beta_1^2 \Sigma_X(\theta_X) + \Sigma_\epsilon(\theta_\epsilon) & \beta_1 \Sigma_X(\theta_X) & \beta_1 \Sigma_{XW}(\theta_X) \\ \beta_1 \Sigma_X(\theta_X) & \Sigma_X(\theta_X) & \Sigma_{XW}(\theta_X) \\ \beta_1 \Sigma'_{XW}(\theta_X) & \Sigma'_{XW}(\theta_X) & \Sigma_W(\theta_X) \end{bmatrix} \right). \quad (2)$$

1.2 Estimators

The KR approach is to impute the missing $X's$ by kriging and then estimate β with weighted least squares as if the imputed X's were the actual X's. This krige-and-regress (KR) estimator is given by

$$\hat{\beta}_{KR} = (\hat{\mathbf{X}}' \Sigma_\epsilon^{-1} \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}' \Sigma_\epsilon^{-1} Y$$

where

$$\hat{\mathbf{X}} = \begin{bmatrix} \mathbf{1}_{n\times 1} & \hat{X} \end{bmatrix}$$

and $\hat{X}$ is the vector of kriged predictions of X:

$$\begin{aligned} \hat{X} &= \mathbf{1}_{n\times 1}\hat{\mu}_X + \Sigma_{XW}\Sigma_W^{-1}(W - \mathbf{1}_{m\times 1}\hat{\mu}_X) \\ &= \Lambda W \end{aligned}$$

with

$$\begin{aligned} \hat{\mu}_X &= \frac{\mathbf{1}_{1\times m}\Sigma_W^{-1}}{\mathbf{1}_{1\times m}\Sigma_W^{-1}\mathbf{1}_{m\times 1}} W \\ \Lambda &= \left[\frac{\mathbf{1}_{n\times m} - \Sigma_{XW}\Sigma_W^{-1}\mathbf{1}_{m\times m}}{\mathbf{1}_{1\times m}\Sigma_W^{-1}\mathbf{1}_{m\times 1}} + \Sigma_{XW} \right] \Sigma_W^{-1}. \end{aligned}$$

$\hat{X}$ is the best linear unbiased predictor (BLUP) of X given W.

To obtain the ML estimator, we maximize the joint likelihood of the observed data Y and W. The model in equation (2) implies

$$\begin{bmatrix} Y \\ W \end{bmatrix} \sim N\left(\begin{bmatrix} (\beta_0 + \beta_1 \mu_X \mathbf{1}_{(n)\times 1} \\ \mu_X \mathbf{1}_{m\times 1} \end{bmatrix}, \begin{bmatrix} \beta_1^2 \Sigma_X(\theta_X) + \Sigma_\epsilon(\theta_\epsilon) & \Sigma_{XW}(\theta_X) \\ \Sigma'_{XW}(\theta_X) & \Sigma_W(\theta_X) \end{bmatrix} \right). \tag{3}$$

The joint density of Y and W is therefore

$$f_{YW}(y,w) = \frac{1}{\sqrt{2\pi |\Sigma(\theta_X, \theta_\epsilon)|}} \exp\left(-\frac{1}{2} V' \Sigma^{-1}(\theta_X, \theta_\epsilon) V \right)$$

where

$$\Sigma(\theta_X, \theta_\epsilon) = \begin{bmatrix} \beta_1^2 \Sigma_X(\theta_X) + \Sigma_\epsilon(\theta_\epsilon) & \beta_1 \Sigma_{XW}(\theta_X) \\ \beta_1 \Sigma'_{XW}(\theta_X) & \Sigma_W(\theta_X) \end{bmatrix},$$

and

$$\begin{aligned} V &= \begin{bmatrix} V_Y \\ V_X \end{bmatrix} \\ &= \begin{bmatrix} Y - (\beta_0 + \beta_1 \mu_X) \mathbf{1}_{n\times 1} \\ W - \mu_X \mathbf{1}_{m\times 1} \end{bmatrix}. \end{aligned} \tag{4}$$

The maximum likelihood estimate of β is found by maximizing

$$\log(f_{YW}) = \text{constant} - \frac{1}{2} \log\det(\Sigma) - \frac{1}{2} V' \Sigma^{-1} V \tag{5}$$

with respect to the parameters β, μ_X, θ_X, and θ_ϵ.

2 Krige and Regress

The KR estimator of β_1 under the linear model in equation (1 is

$$\begin{aligned} \hat{\beta}_{KR} &= \frac{\hat{X}'(\mathbf{1}_{1\times n} \Sigma_\epsilon^{-1} \mathbf{1}_{n\times 1} \Sigma_\epsilon^{-1} - \Sigma_\epsilon^{-1} \mathbf{1}_{n\times n} \Sigma_\epsilon^{-1}) Y}{\hat{X}'(\mathbf{1}_{1\times n} \Sigma_\epsilon^{-1} \mathbf{1}_{n\times 1} \Sigma_\epsilon^{-1} - \Sigma_\epsilon^{-1} \mathbf{1}_{n\times n} \Sigma_\epsilon^{-1}) \hat{X}} \\ &= \frac{W' \Lambda' M Y}{W' \Lambda' M \Lambda W} \end{aligned}$$

where $M = \mathbf{1}_{1\times n} \Sigma_\epsilon^{-1} \mathbf{1}_{n\times 1} \Sigma_\epsilon^{-1} - \Sigma_\epsilon^{-1} \mathbf{1}_{n\times n} \Sigma_\epsilon^{-1}$.

2.1 Expectation and Variance

The expectation can be calculated by first conditioning on W and then iterating the expectation. We have

$$\begin{aligned} E(Y|W) &= E(\beta_0 \mathbf{1}_{n\times 1} + \beta_1 X | W) \\ &= \beta_0 \mathbf{1}_{n\times 1} + \beta_1 [\mu_X \mathbf{1}_{n\times 1} + \Sigma_{XW} \Sigma_W^{-1} (\mu_X \mathbf{1}_{m\times 1} - W)]. \end{aligned}$$

Note that $M \mathbf{1}_{n\times 1} = 0$, so

$$\begin{aligned} E(\hat{\beta}_1 | W) &= (W' \Lambda' M \Lambda W)^{-1} W' \Lambda' M E(Y|W) \\ &= (W' \Lambda' M \Lambda W)^{-1} W' \Lambda' M \\ &\quad \cdot \{\beta_0 \mathbf{1}_{n\times 1} + \beta_1 [\mu_X \mathbf{1}_{n\times 1} + \Sigma_{XW} \Sigma_W^{-1} (\mu_X \mathbf{1}_{m\times 1} - W)]\} \\ &= \beta_1 (W' \Lambda' M \Lambda W)^{-1} \\ &\quad \cdot (W' \Lambda' M \Sigma_{XW} \Sigma_W^{-1} W - \mu_X W' \Lambda' M \Sigma_{XW} \Sigma_W^{-1} \mathbf{1}_{m\times 1}) \end{aligned}$$

which is a ratio of quadratic forms. The unconditional expectation of $\hat{\beta}_1$ may be found by the formula for the expectation of a ratio of powers of quadratic forms given in Mathai and Provost (1992).

$$E(\hat{\beta}_1) = \beta_1 E(Q_1 Q_2^{-1})$$

with

$$\begin{aligned} Q_1 &= W'\Lambda' M \Sigma_{XW}\Sigma_W^{-1} W - \mu_X W'\Lambda' M \Sigma_{XW}\Sigma_W^{-1}\mathbf{1}_{m\times 1} \\ Q_2 &= W'\Lambda' M \Lambda W. \end{aligned} \tag{6}$$

The variance of $\hat{\beta}_1$ can be similarly calculated.

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= E[\text{var}(\hat{\beta}_1|W)] + \text{var}[E(\hat{\beta}_1|W)] \\ &= E\{\text{var}[(W'\Lambda' M\Lambda W)^{-1}W'\Lambda' MY|W]\} \\ &\quad + E\{[E(\hat{\beta}_1|W)]^2\} - \{E[E(\hat{\beta}_1|W)]\}^2 \\ &= E[Q_2^{-2}W'\Lambda' M \text{var}(Y|W) M'\Lambda W] + E(Q_1^2 Q_2^{-2}) - [E(Q_1 Q_2^{-1})]^2 \\ &= E\{Q_1^{-2}W'\Lambda' M[\Sigma_\epsilon + \beta_1^2(\Sigma_X - \Sigma_{XW}\Sigma_W^{-1}\Sigma_{XW}')]M'\Lambda W\} \\ &\quad + E(Q_1^2 Q_2^{-2}) - [E(Q_1 Q_2^{-1})]^2 \\ &= E(Q_3 Q_2^{-2}) + E(Q_1^2 Q_2^{-2}) - [E(Q_1 Q_2^{-1})]^2 \end{aligned}$$

where Q_1 and Q_2 are from (6) and

$$Q_3 = W'\Lambda' M[\Sigma_\epsilon + \beta_1^2(\Sigma_X - \Sigma_{XW}\Sigma_W^{-1}\Sigma_{XW}')]M'\Lambda W.$$

The expressions for the variance and expectation of $\hat{\beta}_{KR}$ depend on the unknown parameter β_1 as well as the covariance parameters. It is possible to obtain a variance estimator by substituting estimates of these unknown parameters, but this gives highly negatively biased variance estimates (Madsen, 2004).

2.2 Consistency

To prove consistency of the KR estimator of β_1, we assume that we have N iid copies of the data vector $\begin{bmatrix} Y' & W' \end{bmatrix}'$. This assumption is unrealistic because in practice, we will either have only one observation, or if we do have replication, the replicates will not be iid because the spatial relationships among the data will differ.

If $\hat{\beta}_{ML}$ is obtained from a Newton-Raphson maximization of the log-likelihood with consistent estimators as initial values, then $\hat{\beta}_{ML}$ will be asymptotically normal with covariance given by the inverse information matrix (Lehmann, 1998). This justifies an information-based variance estimate, which turns out to be useful, even with only one multivariate observation.

Suppose we have an iid sequence of multivariate spatially misaligned observations $\begin{bmatrix} Y_i & W_i \end{bmatrix}$, $i = 1, \ldots, N$ where each Y_i is $n \times 1$ and each W_i is $m \times 1$. Let X_i be the $n \times 1$ vector of unobserved covariates associated with Y_i, and let ϵ_i be the $n \times 1$ error vector. Let $\mathbf{X}_i = \begin{bmatrix} \mathbf{1}_{n\times 1} & X_i \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}'$. The model is

$$Y_i = \mathbf{X}_i \beta + \epsilon_i.$$

Assume

$$\begin{bmatrix} X_i \\ W_i \end{bmatrix} \sim \text{iid } N\left(\mu_X \mathbf{1}_{(n+m)\times 1}, \begin{bmatrix} \Sigma_X(\theta_X) & \Sigma_{XW}(\theta_X) \\ \Sigma_{XW}'(\theta_X) & \Sigma_W(\theta_X) \end{bmatrix}\right) \tag{7}$$

and

$$\epsilon_i \sim \text{iid } N(0, \sigma^2 I).$$

Assume that all ϵ's are independent of all X's and W's. The assumptions and notation introduced in Section 1.1 coincides with this framework when $N = 1$.

We now state the result.

Theorem 1 *Assume the asymptotic framework above, and suppose θ_X is estimated by Restricted Maximum Likelihood (Searle, Casella, & McCulloch, 1992) and μ_X is estimated by*

$$\hat{\mu}_X = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{1}_{1\times m}\Sigma_W^{-1}\mathbf{1}_{m\times 1})^{-1}\mathbf{1}_{1\times m}\Sigma_W^{-1}W_i.$$

Suppose $\Sigma_{XW} \neq 0$. Then $\hat{\beta}_{KR}$ is consistent for β. In particular,

$$\sqrt{N}(\hat{\beta}_{KR} - \beta) \xrightarrow{D} N(0, \Sigma_L)$$

where

$$\begin{aligned}
\Sigma_L &= E\left[\hat{\mathbf{X}}_1'\hat{\mathbf{X}}_1\right]^{-1} V_1 E\left[\hat{\mathbf{X}}_1'\hat{\mathbf{X}}_1\right]^{-1} \\
&\quad + E\left[\hat{\mathbf{X}}_1'\hat{\mathbf{X}}_1\right]^{-1}\frac{\partial\hat{\mathbf{X}}_1'}{\partial\mu}E(Y_1)V_2E(Y_1')\frac{\partial\hat{\mathbf{X}}_1}{\partial\mu}E\left[\hat{\mathbf{X}}_1'\hat{\mathbf{X}}_1\right]^{-1} \\
&\quad + E\left[\hat{\mathbf{X}}_1'\hat{\mathbf{X}}_1\right]^{-1}\Bigg\{\left[2\frac{\partial\hat{\mathbf{X}}_1'}{\partial\mu}E(Y_1)C' + \left[\tfrac{\partial\hat{\mathbf{X}}_1'}{\partial\theta_1}\hat{\mathbf{X}}_1\beta \quad \tfrac{\partial\hat{\mathbf{X}}_1'}{\partial\theta_2}\hat{\mathbf{X}}_1\beta \quad \tfrac{\partial\hat{\mathbf{X}}_1'}{\partial\theta_3}\hat{\mathbf{X}}_1\beta\right]\right] \\
&\qquad \cdot [E(\mathcal{I})]^{-1} E\left(\begin{bmatrix}\beta\hat{\mathbf{X}}_1'\frac{\partial\hat{\mathbf{X}}_1}{\partial\theta_1}\\ \beta\hat{\mathbf{X}}_1'\frac{\partial\hat{\mathbf{X}}_1}{\partial\theta_2}\\ \beta\hat{\mathbf{X}}_1'\frac{\partial\hat{\mathbf{X}}_1}{\partial\theta_3}\end{bmatrix}\right)\Bigg\}E\left[\hat{\mathbf{X}}_1'\hat{\mathbf{X}}_1\right]^{-1},
\end{aligned}$$

where

$$\begin{aligned}
V_1 &= E[\hat{\mathbf{X}}_1'(\beta_1^2\Sigma_{X|W} + \sigma^2 I)\hat{\mathbf{X}}_1] \text{ and} \\
V_2 &= (\mathbf{1}_{1\times m}\Sigma_W^{-1}\mathbf{1}_{m\times 1})^{-1},
\end{aligned}$$

and C is a 3×1 vector with

$$\begin{aligned}
C_i &= \operatorname{cov}\Bigg[(\mathbf{1}_{1\times m}\Sigma_W^{-1}\mathbf{1}_{m\times 1})^{-1}\mathbf{1}_{1\times m}\Sigma_W^{-1}W_i - \mu_X\mathbf{1}_{m\times 1} \\
&\qquad -\frac{1}{2}\text{trace}\left(Q^{-1}\frac{\partial Q^{-1}}{\partial\theta_i}\right) - \frac{1}{2}W_1'K'\frac{\partial Q^{-1}}{\partial\theta_i}KW_1\Bigg] \\
&= (\mathbf{1}_{1\times m}\Sigma_W^{-1}\mathbf{1}_{m\times 1})^{-1}\mathbf{1}_{1\times m}\Sigma_W^{-1}\operatorname{cov}\left(W_1, W_1'K'\frac{\partial Q^{-1}}{\partial\theta_i}KW_1\right) \\
&= -\mu_X(\mathbf{1}_{1\times m}\Sigma_W^{-1}\mathbf{1}_{m\times 1})^{-1}\mathbf{1}_{1\times m}K'\frac{\partial Q^{-1}}{\partial\theta_i}K\mathbf{1}_{m\times 1},
\end{aligned}$$

with K any $(m-1)\times m$ known matrix of full row rank such that $K\mathbf{1}_{m\times 1} = 0$ and $Q = K\Sigma_W(\theta)K'$.

The proof is given in Madsen (2004). The asymptotic variance is written as the sum of three terms which can be interpreted as

$$\begin{aligned}
\Sigma_L &= \text{covariance when only } \beta \text{ unknown} \\
&+ \text{loss of efficiency for estimating } \mu_X \\
&+ \text{loss of efficiency for estimating } \theta_X.
\end{aligned}$$

3 Maximum Likelihood

The maximum likelihood estimator is found by maximizing the likelihood in equation (5) with respect to the parameters $\phi = [\beta_0\ \beta_1\ \mu_X\ \theta_X\ \theta_\epsilon]$. This is equivalent to minimizing the objective function

$$-2l = \log|\Sigma| + V'\Sigma^{-1}(\theta_X)V$$

for parameters $\phi = [\beta_0\ \beta_1\ \mu_X\ \theta_X\ \theta_\epsilon]$. From equation (4), V is a function of β and μ_X. Denote the estimator obtained by the minimization as $\hat{\phi}_{ML}$. The second element of this vector is $\hat{\beta}_{ML}$, the maximum likelihood estimator of the slope parameter.

Denote the matrix of second derivatives of l by $\mathcal{I} = \{\frac{\partial^2 l}{\partial\phi_i\partial\phi_j}\}$. The ijth element of the information matrix is

$$\begin{aligned} I_{ij} &= -E\left(\frac{\partial^2 l}{\partial\phi_i\partial\phi_j}\right) \\ &= \frac{1}{2}\text{trace}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\phi_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\phi_j}\right) + \frac{\partial V'}{\partial\phi_i}\Sigma^{-1}\frac{\partial V}{\partial\phi_j}. \end{aligned}$$

Under regularity conditions given in (Lehmann, 1998) and verified for the present model and asymptotic framework in (Madsen, 2004), ϕ_{ML} is asymptotically normal with covariance matrix $\left[\mathcal{I}(\hat{\phi})\right]^{-1}$. This suggests two potential variance estimators. The first is

$$\hat{\text{var}}_1(\hat{\phi}_{ML}) = \mathcal{I}^{-1}(\hat{\phi}_{ML})$$

but in simulations this matrix usually fails to be positive definite. Another information-based variance estimator can be defined as

$$\hat{\text{var}}_2(\hat{\phi}_{ML}) = [E(\mathcal{I})]^{-1}\,|_{\phi=\hat{\phi}_{ML}} \tag{8}$$

which is positive definite under the regularity conditions verified in Madsen (2004).

4 Simulation Study

The theory developed in Sections 2.2 and 3 show that as the number N of independent observation vectors increases, $\hat{\beta}_{ML}$ approaches optimality, i.e. it is asymptotically unbiased and efficient. In this section, we examine the performance of $\hat{\beta}_{ML}$ and its variance estimator, and we compare $\hat{\beta}_{KR}$ and $\hat{\beta}_{ML}$ when $N = 1$ in a simulation study.

A full factorial experiment was done to get an idea of the performance of the estimators for several combinations of true parameter values. For simplicity, we set $\theta_{X,1} = \theta_{\epsilon,1} = 0$. This means that the parameters $\theta_{X,2}$ and $\theta_{\epsilon,2}$ alone control the magnitude of the variance of the X and ϵ processes respectively. The strength of the spatial correlation of the two processes depends on the parameters $\theta_{X,2}$ and $\theta_{\epsilon,2}$. We also fix $\beta = \begin{bmatrix}1 & 1\end{bmatrix}$ and $\mu_X = 1$. This scheme defines sixteen experimental "treatments." For each treatment, 75 replications were simulated and we recorded the Monte Carlo mean squared error

$$\text{MSE} = (\text{number of simulations})^{-1}\sum(\hat{\beta}_1 - \beta)^2.$$

For each replication, vectors W and Y are simulated according to model in equation (3) with $n = m = 50$. The spatial locations of W and Y are shown in Figure 1. The parameter values given in Table 1.

The treatments are listed in Table 2. In this table, "+" denotes the large value of the parameter and "−" denotes the small value.

Taking all 16 experiments together, we plot $\hat{\beta}_{ML}$ vs. $\hat{\beta}_{KR}$ in Figure 2. The maximum likelihood estimator has a smaller MSE than the KR estimator.

We can compare the information-based variance estimate in equation (8) to the observed Monte Carlo MSE. Figure 3 shows this variance estimate for each replicate of each experiment. Superimposed on this

Table 1: Parameter values for the factorial design.

Parameter(s)	Value(s)
μ_X, β_0, β_1	1
θ_{X1}, $\theta_{\epsilon 1}$	0
$\theta_{\epsilon 2}$	$\{0.1, 1\}$
$\theta_{\epsilon 3}$	$\{0.5, 3\}$
θ_{X2}	$\{0.25, 1\}$
θ_{X3}	$\{0.5, 1.5\}$

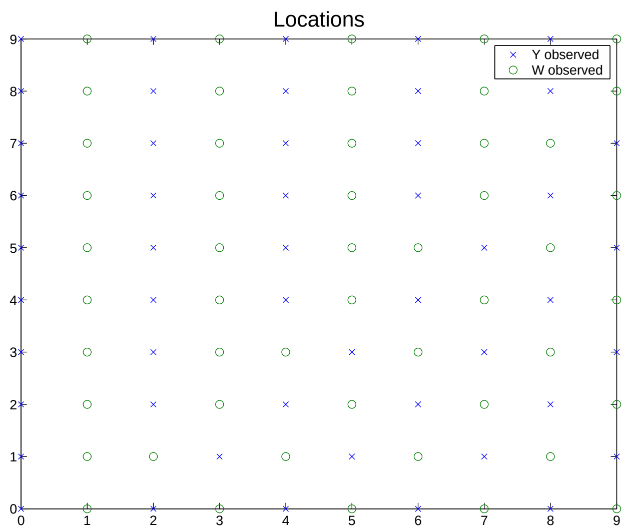

Figure 1: *The spatial locations of the simulated vectors W and Y form a regularly spaced* 10×10 *grid.*

Table 2: Scheme of the factorial simulation experiment. Each of the four parameters θ_{e2}, θ_{e3}, θ_{X2}, and θ_{X3} take on one of two levels leading to sixteen different "treatments." A "+" denotes the larger value and a "−" denotes a small value. The parameter values are given in Table 1.

Treatment	θ_{e2}	θ_{e3}	θ_{X2}	θ_{X3}
1	−	−	−	−
2	+	−	−	−
3	−	+	−	−
4	+	+	−	−
5	−	−	+	−
6	+	−	+	−
7	−	+	+	−
8	+	+	+	−
9	−	−	−	+
10	+	−	−	+
11	−	+	−	+
12	+	+	−	+
13	−	−	+	+
14	+	−	+	+
15	−	+	+	+
16	+	+	+	+

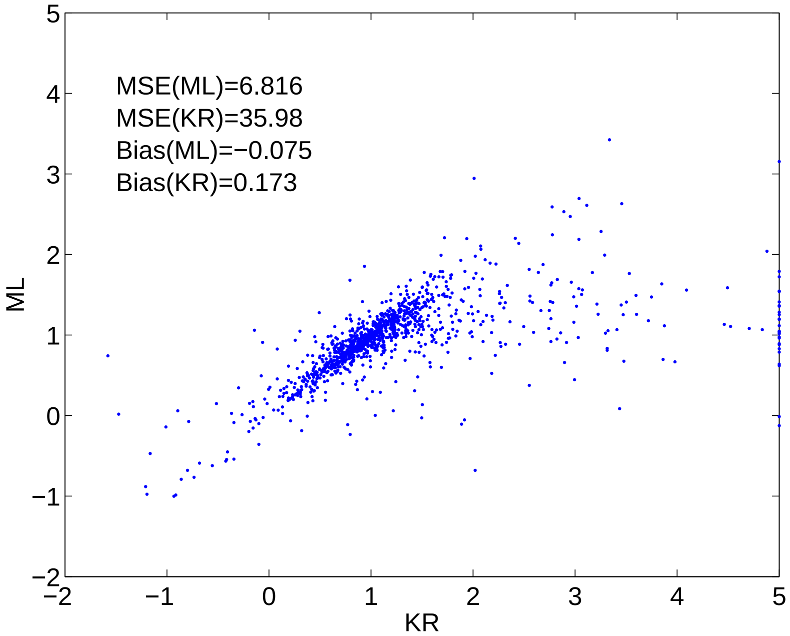

Figure 2: Plot of $\hat{\beta}_{ML}$ vs. $\hat{\beta}_{KR}$ with all 16 treatments combined. $\hat{\beta}_{ML}$ has much smaller MSE and bias than $\hat{\beta}_{KR}$.

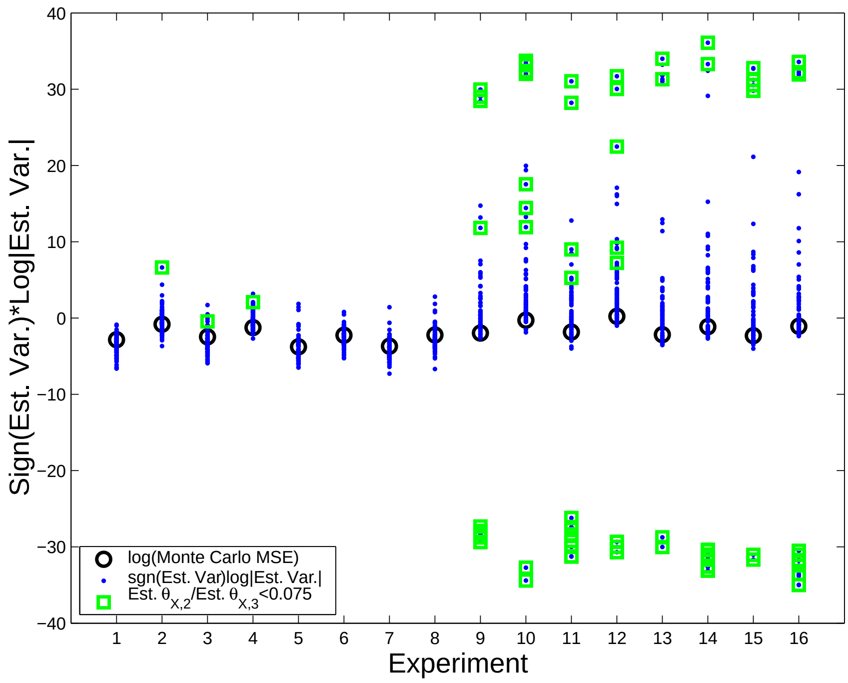

Figure 3: Plot of information-based variance estimates $\widehat{\mathrm{var}}_2(\hat{\beta}_{ML})$ from equation (8) for each of the sixteen experiments. The quantity plotted on the vertical axis is $\mathrm{sgn}[\widehat{\mathrm{var}}(\hat{\beta}_1)] \log|\widehat{\mathrm{var}}(\hat{\beta}_1)|$. Logarithms of the Monte Carlo MSE for each experiment are indicated by open circles. Cases where $\hat{\theta}_{X,2}/\hat{\theta}_{X,3} < 0.075$ are indicated by open squares. Four cases are excluded because $\widehat{\mathrm{var}}(\hat{\beta}_1)$ is infinite. There is one infinite case from each of treatments 9, 10, 13, and 14.

plot are open circles representing the values of the MSE's. In some simulations, problems estimating the covariance parameters lead to numerical instability, and we get variance estimates that are much too large or small, or, in some cases, negative. To show the range of values, as well as some detail for moderate values, the quantity plotted on the vertical axis is $\mathrm{sgn}[\widehat{\mathrm{var}}(\hat{\beta}_1)] \log |\widehat{\mathrm{var}}(\hat{\beta}_1)|$, so that the vertical scale is a log scale, reflected about 0 to accommodate negative values. (None of the $|\widehat{\mathrm{var}}(\hat{\beta}_1)|$'s are closer to 0 than 0.1, so the log function's singularity at 0 doesn't pose a problem.) The numerical problems are usually because the ratio $\hat{\theta}_{X,2}/\hat{\theta}_{X,3}$ is too small. These cases are distinguished by the open squares.

Maximum likelihood estimation has numerical problems when the estimated covariance of the X process is very weak. This occurs when either $\theta_{X,2}$ is small or $\theta_{X,3}$ is large, hence the ratio $\hat{\theta}_{X,2}/\hat{\theta}_{X,3}$ is small. This is usually because the reciprocal range parameter estimate $\hat{\theta}_{X,3}$ exceeds a value of five. Since the locations of the points in the simulation are a minimum of one unit apart (Figure 1), when $\hat{\theta}_{X,3} > 5$, the data do not exhibit enough spatial correlation for W to be an adequate proxy for X.

References

Lehmann, E. L. (1998). *Elements of large-sample theory.* Springer.

Madsen, L. J. (2004). *Regression with spatially misaligned data.* Unpublished Doctoral Dissertation, Cornell University.

Mathai, A. M., & Provost, S. B. (1992). *Quadratic forms in random variables.* Marcel Dekker, Inc.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components.* John Wiley & Sons, Inc.