INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Université d'Ottawa · University of Ottawa

LINEAR REGRESSION WITH SPATIALLY CORRELATED DATA

By
Cynthia Jacqueline Bocci, B. Comm., M. Sc.
August 1999

A Thesis
submitted to the School of Graduate Studies and Research
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Mathematics[1]

[1]The Ph.D. Program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

In this dissertation, the analysis of spatial data through regression is investigated. Multiple observations taken from sites are assumed to be spatially dependent. Our linear model includes a pure error and a spatial error term whose covariance structure is given by an unknown linear combination of 2 known covariograms. The pure and spatial error terms also have separate scale parameters.

Our first concern is with the estimation of the parameters of this model. An algorithm to estimate these parameters is proposed as well as a consistent estimator for one of the spatial parameters. Numerical simulations support the use of our algorithm.

The second main issue is that of asymptotics. To that end, a formula for the inverse of the variance-covariance matrix of observations is developed. Limits of the asymptotic variance of the parameter estimates as the number of observations per site increases are found with this formula. On the other hand, specific sampling schemes are studied when considering the asymptotics for the number of sites going to infinity.

From the simulation and asymptotic results, some rules for experimental design are given. Extensions to more general models are made and areas of future research including possible applications are suggested.

Acknowledgements

My deepest gratitude goes to Dr. André Robert Dabrowski, my supervisor, for his infinite patience, guidance, and expertise. I also appreciate his efforts to broaden my mathematical perspective by encouraging me and making it possible for me to attend many conferences and workshops. In addition, his generousity in the form of research assistantships is much appreciated.

I wish to thank the University of Ottawa for their financial support and a valuable learning experience.

Dedication

to my parents, my sister, and Joe

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter provides an overview of spatial statistics and a discussion of recent developments in this field. The contribution of the thesis is outlined in Section 1.4.

1.1 What is Spatial Statistics?

Spatial statistics has its origins in geostatistics and was pioneered by Matheron (1962). It was first used in mining applications where problems of determining spatial correlation to predict the quantity of ore in mine deposits typically arise. For example, mineralization, a process converting organic material into a mineral, within a fixed finite region may vary spatially due to heat and pressure. Samples from a geological survey can be analyzed to estimate quantities of minerals at various locations.

Unlike classical experiments, it is not reasonable to assume independence across different sampling locations. Any statistical procedure based on such data must take this dependence into account. One underlying assumption is that observations that are taken far apart are less alike than those taken close together. Generally, the farther apart the observations, the weaker the dependence between them. This dependence structure can be a function not only of distance but also of direction. As an illustration, mineralization may occur differently in different directions according to the direction of pressures. Spatial models capturing the dependence between observations at different locations are used extensively in earth and environmental

sciences. Today, applications include mining, meteorology and air and water quality studies. Many examples are illustrated in Cressie (1993) and Davis (1986). More recently Sampson and Guttorp (1992) and Smith (1996) have developed statistical spatial techniques to assess environmental monitoring networks. Guttorp (1997) has analyzed spatial data on carbon dioxide concentrations. Other studies have analyzed spatial data to assess changes in the environment. The impact of spatial statistics is expected to grow with the development of Geographical Information Systems (GIS) such as *ArcInfo*.

1.2 Tools of the Trade

This section introduces the terminology and objects frequently encountered in spatial statistics and necessary for this thesis. It is assumed that the reader is familiar with the theory of classical linear models, see Hocking (1985) or Searle (1971). In addition, familiarity with basic probability theory is also assumed. A selected summary of matrix operations and multivariate statistics can be found in the appendices to this thesis.

There are many aspects and types of data considered in spatial statistics (geostatistical data, lattice data, and point processes). Here we will concentrate on geostatistical and lattice data. Let $s \in \mathbb{R}^d$ be the spatial location and consider $Z(s)$ a random variable of observations at location $s \in \mathcal{D}$ with $\mathcal{D}$ a fixed subset of $\mathbb{R}^d$. Geostatistical data is usually associated with sampling over a continuous region $\mathcal{D}$ with d-dimensional volume. The set $\mathcal{D}$ can also be a countable collection of spatial sites as in the case of lattice data where observations are obtained from a region which is regularly spaced or defined on a grid. Typically in geostatistics, the random process $Z(s)$ is assumed to have a spatially dependent joint normal distribution rather than independent values for distinct s. This assumption makes theoretical calculations and prediction possible. The assumption of normality is also widely adopted in applications outside of geostatistics.

We first discuss methods of characterizing the degree of spatial dependence. The definitions follow those of Cressie (1993). In spatial statistics, the *variogram* is a

very important tool used to measure the degree of spatial dependence between observations. It differs from the ordinary variance of a process in that it incorporates the correlation between observations at different sites. For locations s_i and $s_{i'}$, the variogram $2\gamma(s_i, s_{i'})$ is defined by

$$2\gamma(s_i, s_{i'}) = \mathrm{Var}(Z(s_i) - Z(s_{i'})).$$

Here $\mathrm{Var}(X)$ denotes the variance of X. The quantity $\gamma(s_i, s_{i'})$ is generally called the *semi-variogram*. Let $h_{ii'} = s_i - s_{i'}$. A random process $Z(\cdot)$ is said to be *intrinsically stationary* if

$$E(Z(s_i) - Z(s_{i'})) = 0,$$
$$\gamma(s_i, s_{i'}) = \gamma(s_i - s_{i'}) = \gamma(h_{ii'}).$$

Such a process has constant mean and a variogram which is a function only of the distance between locations, $h_{ii'} = s_i - s_{i'}$. A *covariogram* $c(\cdot)$, is a covariance function such that

$$\mathrm{Cov}(Z(s_i), Z(s_{i'})) = c(s_i - s_{i'}) = c(h_{ii'}). \tag{1}$$

That is, the covariance depends only on the spatial displacement between two sites. A random function $Z(\cdot)$ with constant mean satisfying (1) is said to be *second-order stationary*. Note that second-order stationarity implies intrinsic stationarity because of the relation

$$\mathrm{Var}(Z(s_i) - Z(s_{i'})) = \mathrm{Var}(Z(s_i)) + \mathrm{Var}(Z(s_{i'})) - 2\mathrm{Cov}(Z(s_i), Z(s_{i'})).$$

In fact, for a second-order stationary process

$$\gamma(h) = c(0) - c(h).$$

A process is said to be *strictly stationary* if it is distributionally invariant under shifts in the index set. Specifically, the distribution of $\{Z(s_1), Z(s_2), \ldots, Z(s_n)\}$ is identical to that of $\{Z(s_1 + h), Z(s_2 + h), \ldots, Z(s_n + h)\}$. The variogram or covariogram is called *isotropic* if it is a function only of the norm of the displacement, the distance $||h||$. The process Z is *anisotropic* if the dependence between $Z(s_i)$ and $Z(s_{i'})$ is a function of direction as well as distance.

Variograms and covariograms must possess certain analytic properties. Ideally, the variogram is an increasing function of distance that eventually levels off to the variance; $\gamma(\infty) = \text{Var}(Z(\cdot))$. Alternatively, from the underlying assumption of asymptotic independence, the covariance is a decreasing function of distance so that $c(\infty) = 0$. As a result, $\gamma(\infty) = c(0)$. In addition, notice that $\gamma(h) = \gamma(-h)$ and $\gamma(0) = 0$. Similarly, $c(h) = c(-h)$ with $c(0) = \text{Var}(Z(\cdot))$. If the semi-variogram is discontinuous at the origin, that is $\gamma(h) \to a_o$ as $h \to 0$, then $a_o > 0$ is called the *nugget*. The discontinuity can occur because of the scale of observation. Observations are not available for spatial separations less than $\min ||s_i - s_{i'}||$ and consequently, a linear projection of the semi-variogram to the origin may not reflect the true nature of that curve. The nugget effect will not by considered in this thesis because we will include a local error in our model. Another property is that the variogram $2\gamma(\cdot)$ must be conditionally negative-definite while the covariogram $c(\cdot)$, of a second-order stationary process, must be positive-definite (see Appendix A.5 for definitions). One further property that variogram models must satisfy is $2\gamma(h)/||h||^2 \to 0$ as $||h|| \to \infty$.

Some popular isotropic semi-variograms are the exponential, rational quadratic, spherical and wave models shown in Figure 1. The spherical and wave models are only valid in $\mathbb{R}^d$ for $d \leq 3$.

$$
\begin{array}{lrcl}
\text{exponential} & \gamma(h) & = & b_e\{1 - exp(-||h||/a_e\} \\
\text{rational quadratic} & \gamma(h) & = & b_r||h||^2/(1 + ||h||^2/a_r) \\
\text{spherical} & \gamma(h) & = & \begin{cases} b_s\{1.5(||h||/a_s) - .5(||h||/a_s)^3\} & 0 \leq ||h|| \leq a_s \\ b_s & ||h|| \geq a_s \end{cases} \\
\text{wave} & \gamma(h) & = & b_w\{1 - a_w sin(||h||/a_w)/||h||\}
\end{array}
$$

with $b_e, b_r, b_s, b_w \geq 0$ and $a_e, a_r, a_s, a_w \geq 0$.

Define the *sill* of a semi-variogram to be the limit of the semi-variogram as $||h|| \to \infty$. The b coefficients in the above models represent the sill. The *range* of the variogram is the distance at which the semi-variogram becomes equal to its sill. This is infinite for all of the above models except the spherical one.

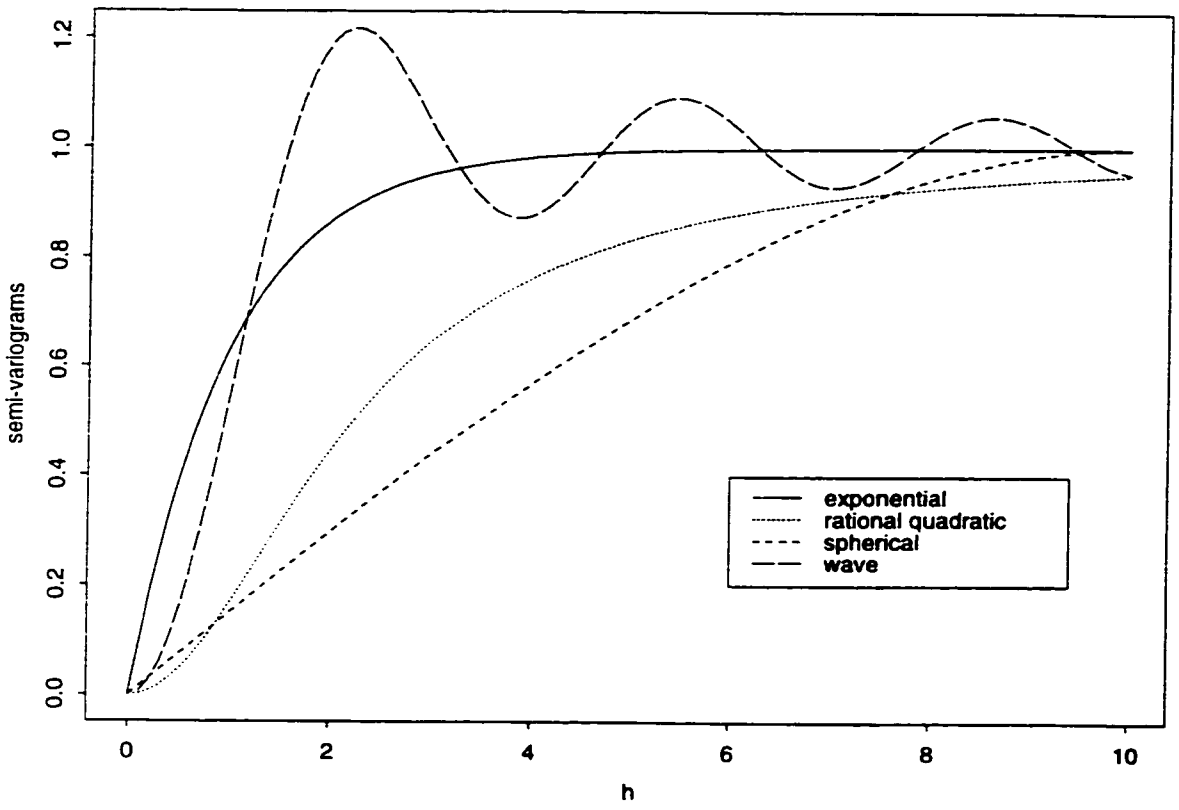


Figure 1: Graphs of four semi-variogram curves described on page 4.

1.3 Recent Developments

Spatial statistics is a relatively young discipline that generates research in several aspects of spatial data analysis. Many applications arise in earth, environmental and health sciences. For example, spatial data can be used to predict the amount of pollutant at an unmonitored site from observations at nearby sites or to detect spatial climatic trends. Spatial data can also be analyzed to create an optimal sampling design under cost constraints. This involves specifying the number of stations (monitoring sites) and their locations. Prediction, trend estimation and spatial design all require a good understanding of the spatial dependence structure for the underlying random process. Hence, the modelling or estimation of the (co)variogram is an active area of research in spatial statistics. The motivation for developing better estimates of the (co)variogram is to ultimately ensure better prediction.

A simplified typical application in geostatistics might proceed as follows: First, a theoretical linear model with parameters is chosen on some basis for the random process $Z(s) = a + bx_s + e$ where x_s is non-random. Data is then collected, detrended, and an empirical variogram is calculated. Next, the empirical variogram for $Z(s)$ is fitted by a completely specified valid variogram and the parameter estimates of a and b for the model are found using the dependence structure of the fitted variogram. Finally, predictions relying on the fitted variogram are generated by a method known as *Kriging*. Basically, Kriging finds the best linear unbiased estimator of the mean value for the variable in question over a limited domain. Of course, verification occurs at many stages. In each phase, there is a variety of techniques at one's disposal.

Naturally, there now exist several estimates of the variogram. In classical geostatistics, the spatial random process is assumed to be stationary and isotropic. Many estimators of the (co)variogram exist under these assumptions. The traditional estimate, $2\hat{\gamma}(h)$ of Matheron (1962), the robust estimate, $2\bar{\gamma}(h)$ of Cressie and Hawkins (1980) and the highly robust estimate, $2\tilde{\gamma}(h)$ of Genton (1998) are just a few of them. For $h \in \mathbb{R}^d$, define $N(h) = \{(s_i, s_{i'}) : s_i - s_{i'} = h\}$, $|N(h)|$ = cardinality of $N(h)$, and $V_i(h) = Z(s_i + h) - Z(s_i)$. Then

$$
\begin{aligned}
2\hat{\gamma}(h) &\equiv \tfrac{1}{|N(h)|}\textstyle\sum_{N(h)}(Z(s_i)-Z(s_{i'}))^2, \\
2\bar{\gamma}(h) &\equiv \left\{\tfrac{1}{|N(h)|}\textstyle\sum_{N(h)}|Z(s_i)-Z(s_{i'})|^{1/2}\right\}^4/(.457+\tfrac{.494}{|N(h)|}), \\
2\tilde{\gamma}(h) &\equiv \left(2.2191\{|V_i(h)-V_{i'}(h)|; i<i'\}_{(k)}\right)^2,
\end{aligned} \tag{2}
$$

where

$$k = \begin{pmatrix} [\mathrm{N(h)}/2]+1 \\ 2 \end{pmatrix},$$

and $[N(h)/2]$ denotes the integer part. The k in the estimator $2\tilde{\gamma}(h)$ refers to the k^{th} quantile for the ordered set of absolute differences $|V_i(h)-V_{i'}(h)|$ for $i<i'$. All of these empirical variograms try to capture the spatial dependence between locations. The latter ones attempt to diminish the influence of outliers in the data.

The assumptions of stationarity and isotropy may not always hold, however, due to factors such as topography, currents, winds, etc. Direction can also play an important role in capturing variability. Consequently, there is considerable interest in estimating spatial correlation for nonstationary, anisotropic processes. Sampson and Guttorp (1992) develop a complex non-parametric estimator in this case. Briefly, it involves a spatial transformation of the geographic map of sites in a so-called G-space into an alternative D-space within which the structure becomes stationary and isotropic. An isotropic Gaussian variogram γ_o is fitted in this D-space. Next, a smooth function f that maps the G-space into the D-image is found using thin plate splines. The original G-space dispersion structure can be captured through the relation

$$\gamma(s_1,s_2) = \gamma_o(f(s_1),f(s_2)). \tag{3}$$

Brown, Le and Zidek (1994) use a Bayesian approach to spatial data whereby the covariance structure has a prior distribution whose covariance structure is in turn estimated using the techniques of Sampson and Guttorp. Smith (1996) also subscribes to the idea of spatial deformation but proposes an alternative method of fitting (3) based on maximum likelihood. One difference here is that f is represented by a linear combination of radial basis functions and the variogram in D-space is fitted using a different structure. These techniques were developed mainly to evaluate spatial networks or to assist in spatial design of experiments.

Some researchers choose to focus instead on the estimation of parameters of a specific linear model. The work of Mardia and Marshall (1984) and Cook and Pocock (1983) lies in this domain. The latter set up a regression model with spatially correlated errors and use the maximum likelihood approach to estimate parameters. An exponential form is chosen for the covariance structure and the maximization is done numerically. Mardia and Marshall (1984) develop theorems for the consistency and asymptotic normality of estimators attained through maximum likelihood in the case of normal dependent observations.

We will also consider regression, but for a different class of variance structures than those discussed in Mardia and Marshall (1984). Our model includes a pure error term and a spatial error term whose covariance structure is given by an unknown linear combination of two known covariograms. The pure and spatial error terms also have a separate scale parameter. The need to determine the spatial covariance structure in our setting is similar to the questions leading to the previously discussed estimates of the covariogram. Although maximum likelihood is a natural approach, we find that it is numerically ineffective, and we develop alternate methods.

1.4 Guide to the Thesis

This thesis will explore estimation problems and experimental design questions when the experimenter wishes to estimate regression coefficients in the presence of spatially-correlated data. In particular, the spatial dependence is partially known; the covariogram is an unknown combination of known covariograms. A description of the contents and highlights of new results for each chapter follows.

Chapter 2 introduces a simple regression model with both regression coefficients and spatial parameters. An alternative algorithm to maximum likelihood estimation is proposed to estimate these parameters in a more reliable fashion. As part of this algorithm, a consistent estimator is found for one of the spatial parameters. Small sample simulation investigates these estimates in comparison to simultaneous numeric maximization of the likelihood function.

In Chapter 3, an inverse for the variance-covariance matrix of observations of

Chapter 2 is obtained. This formula for the inverse is required for the sequel, and is a novel contribution of this thesis.

Chapter 4 studies the asymptotics of the parameter estimates as the number of observations per site and, separately, as the number of sites increases. Theoretical expressions are obtained when the number of observations goes to infinity. To investigate the case when the number of sites goes to infinity, however, small sample simulations are performed on three different sampling schemes. Chapter 5 summarizes these results into general principles to guide the design of a spatial experiment.

Chapter 6 considers a more general model than that of Chapter 2 for which the bulk of the results of Chapter 4 still hold. The last section of Chapter 6 looks at future directions for this research and a possible application incorporating topography.

Chapter 2

Regression with Spatially Correlated Errors

In this chapter, we pursue a simple model for data whose dependence is captured through spatially correlated errors. Various techniques are investigated to obtain parameter estimates. Finally an algorithm is proposed and small sample simulations are executed to compare its performance with standard global optimization. A technical results section follows.

2.1 The Simple Model

The spatial dependence of the data will be specified by semi-variograms chosen from

$$\Gamma = \{\gamma_i : \gamma_i \text{ semi-variogram}, \gamma(0) = 0,\ \gamma_i \neq a\gamma_r \text{ if } i \neq r \text{ for any real } a > 0\}.$$

The last condition above guarantees that one $\gamma \in \Gamma$ is not a scalar multiple of another.

Assume that for each site i, $i = 1, \ldots, n_s$, there correspond the observations (s_i, y_{ij}, x_i) where s_i is the location of site i; y_{ij} denotes the response of the jth sample at site i, $j = 1, \ldots, n_p$; and x_i is the value of a predictor variable, x, at site i. This notation suggests multiple samples at each site with the same number of observations per site. Allow s_i to be d-dimensional. A visual realization of this model is given in

Section 2.1.1. Suppose further that

$$y_{ij} = \beta_o + \beta_1 x_i + \sigma_s \varepsilon_{s_i} + \sigma_p e_{ij} \tag{4}$$

where

$$e_{ij} \sim \mathcal{N}(0,1),$$

with the e_{ij}'s independent of each other and the ε_{s_i}'s. We interpret ε_{s_i} as spatially correlated error and e_{ij} as pure (eg. local or measurement) error. Fix $\gamma_1, \gamma_2 \in \Gamma$. In addition, we assume ε_{s_i} is a second-order stationary process with distribution $\mathcal{N}(0, \mathrm{Var}(\varepsilon_{s_i})) = \mathcal{N}(0, \gamma_\varepsilon(\infty))$ with

$$\gamma_\varepsilon \in \mathcal{G} = \{\gamma \in \Gamma : \gamma = \alpha\gamma_1 + (1-\alpha)\gamma_2 \text{ and } \alpha \in [0,1]\}.$$

Hence, in classical linear models terminology, the model in (4) is a mixed model with the first two terms forming the fixed effects and the last two terms contributing to the random effects, while β_o and β_1 are the usual regression coefficients and σ_p^2 and σ_s^2 are the variance components. As such, there is an analogy to the analysis of variance components in linear models. The parameter to be estimated in this scenario is $\Phi = (\beta_o, \beta_1, \sigma_p^2, \sigma_s^2, \alpha)$.

It is convenient to adopt a multivariate notation at this point. The model (4) becomes

$$\mathrm{y} = \mathrm{Xb} + \sigma_s \varepsilon_s + \sigma_p e \tag{5}$$

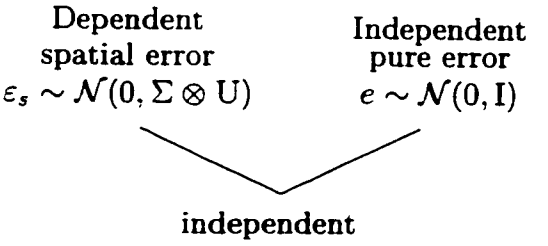
with fixed matrix $\mathrm{X}_{n_s n_p \times 2}$ and $\mathrm{b} = (\beta_0 \ \beta_1)'$ and random $n_s n_p \times 1$ vectors ε_s and e. It is assumed that X is of full rank. In an effort to find the variance of y, note that

$$\mathrm{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_s^2 \mathrm{Var}(\varepsilon_{s_i}) + \sigma_p^2 & i = i', j = j' \\ \sigma_s^2 \mathrm{Var}(\varepsilon_{s_i}) & i = i', j \neq j' \\ \sigma_s^2 \mathrm{Cov}(\varepsilon_{s_i}, \varepsilon_{s_{i'}}) & i \neq i' \end{cases} \tag{6}$$

A more concise expression can be obtained using Kronecker products as defined in Appendix A.2. Let $\Sigma = \left(\mathrm{Cov}(\varepsilon_{s_i}, \varepsilon_{s_{i'}})\right)_{n_s \times n_s}$ and U be a $n_p \times n_p$ matrix of ones. Denote $\mathrm{A} = \Sigma \otimes \mathrm{U}$, so that the variance of y can be written as

$$\begin{aligned} \mathrm{V} = \mathrm{Var}(\mathrm{y}) &= \sigma_p^2 \mathrm{I}_{n_s n_p} + \sigma_s^2 (\Sigma \otimes \mathrm{U}) \\ &= \sigma_p^2 (\mathrm{I} + \delta \mathrm{A}). \end{aligned} \tag{7}$$

with $\delta = \sigma_s^2/\sigma_p^2$. Then y is normally distributed with expectation Xb and variance V, an $n_s n_p \times n_s n_p$ positive definite matrix. In summary,



$$\text{model: } \mathrm{y} = \mathrm{Xb} + \sigma_s \varepsilon_s + \sigma_p e$$
$$\Rightarrow \ \mathrm{y} \sim \mathcal{N}(\mathrm{Xb}, \mathrm{V}) \ \text{ where } \ \mathrm{V} = \sigma_p^2(\mathrm{I} + \delta \mathrm{A}),$$
$$\delta = \sigma_s^2/\sigma_p^2 \ \text{ and } \ \mathrm{A} = \Sigma \otimes \mathrm{U}.$$

Remark: Recall that the variance of a stationary process is the variogram of that process evaluated at infinity. In our model, the process ε_{s_i} has a variogram which in turn is a linear combination of two known variograms. Therefore, for the set $\mathcal{G}$ above, it is desirable that a linear combination of variograms (covariograms) be a variogram (covariogram). The following provides this result.

Lemma 2.1 *The linear combination of variograms (covariograms) with positive coefficients is a variogram (covariogram).*

Proof: See Journel and Huijbregts (1978) p.162. ■

It will be convenient to express the $\mathrm{Var}(\varepsilon_{s_i})$ in terms of covariograms rather than variograms. For a second order stationary process, the covariogram and variogram are related through the equations

$$\begin{aligned} c(h) &= \gamma(\infty) - \gamma(h) \\ \gamma(h) &= c(0) - c(h). \end{aligned}$$

Hence from $\gamma_\epsilon \in \mathcal{G}$, an equivalent covariogram expression arises. In fact,

$$\Sigma = \alpha \mathrm{C}_1 + (1 - \alpha)\mathrm{C}_2,$$

where C_1 and C_2 are matrices whose entries are composed from known valid covariogram models.

Note that if Γ were the set of all variograms then σ_p^2, σ_s^2 and α would not be unique and the model not identifiable. It is necessary to impose the condition that our set of variograms, or equivalently covariograms, cannot be scalar multiples of each other. However, for γ_1 and γ_2 fixed, this is not a problem. There are other possible scenarios that could arise to make the model not identifiable. For example, suppose that two covariogram models c_1 and c_2 are such that $c_1(0) = c_2(0)$ and $c_1(h_o) = c_2(h_o)$. Consider only two sites whose separation is exactly h_o. Then the matrices C_1 and C_2 are equal,

$$C_1 = \begin{pmatrix} c_1(0) & c_1(h_o) \\ c_1(h_o) & c_1(0) \end{pmatrix} = \begin{pmatrix} c_2(0) & c_2(h_o) \\ c_2(h_o) & c_2(0) \end{pmatrix} = C_2,$$

and Σ is no longer a function of α.

It is also possible to standardize the covariogram matrices by writing

$$ST_r = C_2 - \sum_{t=1}^{r} p_t D_t$$

where $D_1, D_2, \ldots, D_r$ are such that $\langle D_t, D_m \rangle = 0 \ \forall t \neq m$, $p_1 = \langle C_2, D_1 \rangle / \langle D_1, D_1 \rangle$ and $p_t = \langle ST_{t-1}, D_t \rangle / \langle D_t, D_t \rangle$ for $t = 2, \ldots, r$. The D_t's represent the orthogonal versions of powers of C_1, C_1^t obtained by letting

$$D_t = C_1^t - \langle C_1^t, C_1 \rangle / \langle C_1, C_1 \rangle C_1.$$

with $D_1 = C_1$. However the D_t's are not necessarily covariograms. This represention can perhaps be used to help numeric convergence of optimization routines.

2.1.1 Example

Small sample simulations throughout this thesis are run using the following example unless otherwise specified. Suppose there are n_s=10 sites with n_p observations per site. The locations of these sites, given in terms of Cartesian coordinates, and the predictor variables (covariates) are shown in Table 1. Figure 3 shows the 2-dimensional site locations. The first four sites are depicted by a plus symbol and the remaining six sites are denoted by triangles. Most simulations throughout this thesis use 10 sites. When using only 4 sites, the first four are chosen.

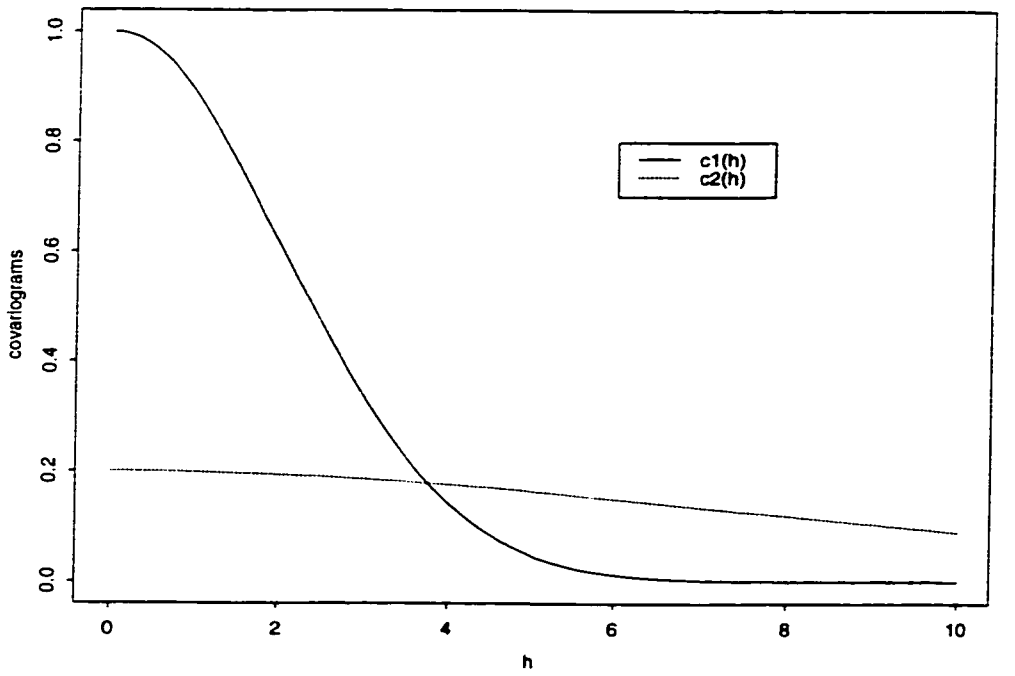


Figure 2: Covariograms $c_1(h)$ and $c_2(h)$

We have chosen two isotropic covariogram models whose graphs are depicted in Figure 2.

$$\begin{aligned} c_1(h) &= 1e^{-(.35)^2\|h\|^2}, \quad h \in \mathbb{R}^d \\ c_2(h) &= .2e^{-(.09)^2\|h\|^2}. \end{aligned} \tag{8}$$

Note that $c_1(h)$ has a strong local dependence in comparison to $c_2(h)$. Figure 4a graphs the given set of predictors (see Table 1) against given site locations; whereas Figure 4b plots simulated observations against site coordinates using the model (4) with values from Table 1.

Table 1: Site Locations and Predictor Values. The first four sites are chosen when only four sites are used.

Site s_i	Predictor x_i
(-4,0)	12
(0,3)	100
(2,-2)	435
(5,1)	215
(4,2)	309
(3,6)	400
(-1,1)	72
(1,1)	75
(-2,-3)	59
(-1,6)	17

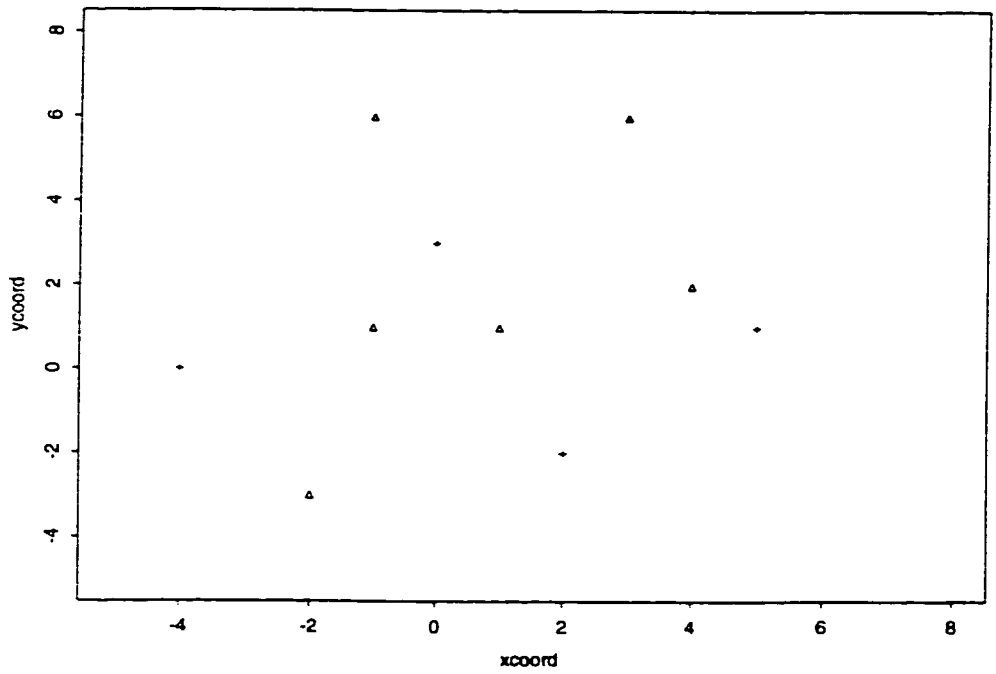
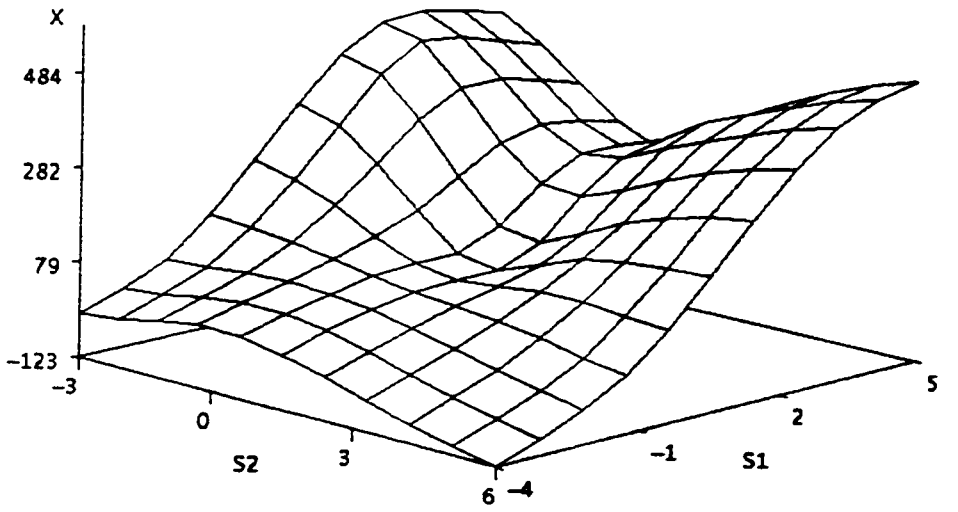
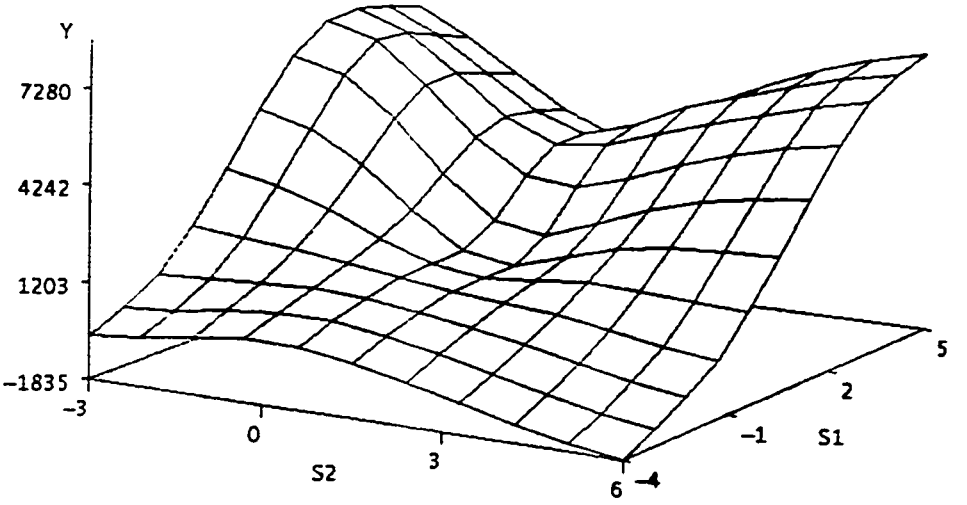


Figure 3: Site locations: 2-dimensional coordinate system. The four-site set of the Example is indicated by a +. The ten-site set comprises all the indicated points.



a) predictors plotted against location coordinates



b) observations plotted against location coordinates

Figure 4: Example of predictor surface and observations surface. These are smoothed using a spline fit to the sites given in Table 1.

2.2 Estimates of Φ

At this point, the techniques of Mardia and Marshall (1984) and Cook and Pocock (1983) for estimating spatial parameters in linear models, are briefly reviewed. Mardia and Marshall (1984) set up a regression model with spatially correlated errors, and use the method of maximum likelihood to estimate not only the regression coefficients but also the set of covariance parameters. They discuss computational aspects and suggest methods of scoring using updating equations to numerically maximize the likelihood function. Sufficient conditions for asymptotic normality and weak consistency (see Appendix B.3 for definition) of the covariance parameters are found. Cook and Pocock (1983) also use a regression model with spatially correlated errors. An exponentially decreasing function is chosen for the spatial correlation of these errors. The maximum likelihood estimation is done numerically. They compute the maximum likelihood estimates by carrying out a grid search. This is feasible since the variance-covariance matrix of the spatial errors has a particular parametrization which reduces calculations by simplifying the likelihood function significantly. Our model differs in that it incorporates both spatial and pure error and a more difficult covariance structure. The variance-covariance matrix V is a function not only of σ_p^2 but also of the two spatial parameters σ_s^2 and α. The parameter α is embedded in the variance matrix of the spatial errors $\Sigma(\alpha) = \alpha C_1 + (1-\alpha)C_2$. Hence, the matrix V has a structure which cannot be exploited in the same manner as the variance matrix of Mardia and Marshall (1984) or Cook and Pocock (1983). It appears that their maximum likelihood approach cannot successfully estimate the parameters of our model.

2.2.1 Maximum Likelihood Equations

Consider the model in (5) setting $b = (\beta_0, \beta_1)$ and $\theta = (\theta_1, \theta_2, \theta_3) = (\sigma_p^2, \sigma_s^2, \alpha)$. Given that the distribution of y is $\mathcal{N}_{n_s n_p}(Xb, V(\theta))$, the log-likelihood function is

$$L = \frac{-n_s n_p}{2} \ln(2\pi) - \frac{1}{2} \ln \mid V(\theta) \mid -\frac{1}{2}(y - Xb)' \, V^{-1}(\theta) \, (y - Xb). \qquad (9)$$

Ordinarily, to compute the maximum likelihood estimates of b and θ, the derivatives of the log-likelihood function with respect these parameters are set equal to zero and solved. The first partial derivatives are given below.

$$\begin{aligned} \mathrm{L_b} = \frac{\partial \mathrm{L}}{\partial \mathrm{b}} &= -\mathrm{X}'\,\mathrm{V}^{-1}\,\mathrm{X}\,\mathrm{b} + \mathrm{X}'\,\mathrm{V}^{-1}\,\mathrm{y} \\ (\mathrm{L}_\theta)_l = \frac{\partial \mathrm{L}}{\partial \theta_l} &= -\frac{1}{2}\,tr\left(\mathrm{V}^{-1}\,\frac{\partial \mathrm{V}}{\partial \theta_l}\right) + \frac{1}{2}(\mathrm{y}-\mathrm{Xb})'\,\mathrm{V}^{-1}\,\frac{\partial \mathrm{V}}{\partial \theta_l}\,\mathrm{V}^{-1}\,(\mathrm{y}-\mathrm{Xb}) \\ \text{for } l = 1,2,3. \end{aligned}$$

They give rise to the likelihood equations

$$(\mathrm{X}'\,\mathrm{V}^{-1}\,\mathrm{X})\mathrm{b} = \mathrm{X}'\,\mathrm{V}^{-1}\,\mathrm{y} \tag{10}$$

$$tr\left(\mathrm{V}^{-1}\,\frac{\partial \mathrm{V}}{\partial \theta_l}\right) = (\mathrm{y}-\mathrm{Xb})'\,\mathrm{V}^{-1}\,\frac{\partial \mathrm{V}}{\partial \theta_l}\,\mathrm{V}^{-1}\,(\mathrm{y}-\mathrm{Xb}) \tag{11}$$

$$\text{for } l = 1,2,3.$$

Exact solutions to equations (10) and (11) cannot in general be obtained, so numeric techniques must be employed. These methods will depend on the curvature of the likelihood.

There are some consequences to estimates obtained by the method of maximum likelihood. Under general conditions (see Appendix B.2), the asymptotic variance matrix of the maximum likelihood estimates is given by the inverse of the Fisher information matrix defined by

$$\mathcal{I}(\Phi) = [-E_\Phi\,\frac{\partial^2 L}{\partial \Phi^2}]\Big|_\Phi .$$

The standard errors for the parameter estimates are the square root of the diagonal elements of the estimated asymptotic variance.

2.2.2 Traditional Approaches

Various methods are needed to estimate the parameters because the likelihood equations are intractable. Although many approaches are conceivable not all of them work well. In fact, due to the dependence structure, there are some practical problems.

Table 2: 100 Simulations of *nummax*

	β_0	β_1	σ_p^2	σ_s^2	α
Real value	2	15	16	49	.25
Sample mean	1.613	14.999	15.989	16.191	.582
Standard error	3.63	.009	6.96	38.35	.47
Number of Boundary Estimates				41	78

These are discussed before we outline our preferred procedure. Highlights of simulation results for the various techniques follow while a comparison is reserved until Section 2.2.4. Technical results regarding estimation techniques are found in Section 2.3. For the rest of this chapter and unless otherwise specified, the empirical results are based on 100 simulations using the example of Section 2.1.1, with n_s=10, n_p=2, and observations coming from a $\mathcal{N}(\mathrm{Xb}, \mathrm{V})$ distribution with $\Phi = (2, 15, 16, 49, .25)$. All the *S-Plus* programs used in the simulations are listed in Appendix C.

Numerical Global Maximization

A straight-forward numerical approach maximizing the likelihood of y over all five parameters simultaneously gives disappointing results. Specifically, σ_s^2 and α are poorly estimated as shown in Table 2. The routine restricts $\alpha \in [0, 1]$ and imposes the boundaries $.001 < \sigma_s^2 < \infty$. Simulations yield a number of boundary values for both α and σ_s^2. In fact, from 100 repetitions, there are 41 infeasible estimates for σ_s^2 and 78 boundary estimates for α. The number of estimates in which both α and σ_s^2 are boundary values is 20. This implies that 99 out of 100 estimates involve infeasible or boundary estimates for either α or σ_s^2. The simulations are generated by the program *nummax* and require an initial guess. The traditional regression estimates are used for the initial guesses of β_0, β_1, and σ_p^2 while the initial values of σ_s^2 and α are 81 and .1 respectively.

Numerical Gauss-Seidal Algorithm

Next, notice that by (10), b can be expressed as a function of θ, that is

$$\mathrm{b} = (\mathrm{X}'\mathrm{V}(\theta)^{-1}\mathrm{X})^{-1}\mathrm{X}'\mathrm{V}(\theta)^{-1}\mathrm{y}.$$

An approach taken by Chiu, Leonard, and Tsui (1996) can be adapted to our problem. Their procedure is as follows: First select an initial estimate for θ to obtain $\hat{\mathrm{b}}$. Next, substitute $\hat{\mathrm{b}}$ into (11) and perform a Newton-Raphson-type algorithm to obtain an updated estimate of θ. Use the latter to update the estimate of b and so on. Continue updating until convergence of $\hat{\theta}$ is achieved. Mardia and Marshall (1984) use this technique with a different updating equation. This technique is sometimes referred to as a nonlinear Gauss-Seidal algorithm. In our simulations, however, estimates for θ are rarely obtained in this manner. In the program *GSeidal*, the very first Newton-Raphson step (i.e. the first update) for θ is never realized because either σ_p^2 or σ_s^2 or both parameters fail to converge. Generally, only β_0 and β_1 are relatively well estimated. It should be noted that due of the complexity of equation (11), the Jacobian in the Newton-Raphson step is approximated by the difference quotient with discretization parameter equal to .01.

Analysis of Variance Technique

Efforts now shift to more accurately estimating the numerically sensitive parameters, σ_s^2 and α. An analysis of variance technique, such as used in random effects models, is unsuccessful in capturing σ_s^2. This method produces estimates for the variance components σ_p^2 and σ_s^2 as follows. Define

$$\begin{aligned} SSE &= \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2, \\ SSA &= n_p \sum_{i=1}^{n_s} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2, \end{aligned}$$

and denote

$$v = \mathrm{Var}(\varepsilon_{s_i}),$$

$$\begin{aligned} \mu_i &= \beta_0 + \beta_1 x_i, \\ \text{and } c(h_{ii'}) &= \text{Cov}(\varepsilon_{s_i}, \varepsilon_{s_{i'}}). \end{aligned}$$

Let $r_X = rank(\mathrm{X})$. Equate the expectations

$$\begin{aligned} E(SSE) &= (n_s n_p - r_X)\sigma_p^2, \\ E(SSA) &= n_p \sum_i (\mu_i - \bar{\mu}.)^2 + (n_s n_p v - \frac{n_p}{n_s} \sum_i \sum_{i'} c(h_{ii'}))\sigma_s^2 + (n_s - 1)\sigma_p^2, \end{aligned}$$

to their respective sums of squares and solve for σ_p^2 and σ_s^2. The estimates $\bar{\sigma}_p^2 = SSE/(n_s n_p - r_X)$ and

$$\bar{\sigma}_s^2 = \frac{SSA - n_p \sum_i (\mu_i - \bar{\mu}.)^2 - (n_s - 1)\bar{\sigma}_p^2}{n_s n_p v - \frac{n_p}{n_s} \sum_i \sum_{i'} c(h_{ii'})}$$

are theoretically unbiased (as proved in Lemma 2.2). Nevertheless, simulation results obtained from the program *anovsig* suggest a large variance for $\bar{\sigma}_s^2$. It is known that the analysis of variance technique can produce negative estimates for the variance components. In fact, a large proportion of estimates for σ_s^2 here are negative. See Table 3.

Method of Moments

Equally disappointing is the estimate for σ_s^2 obtained by a modified method of moments technique. From (4), the dependent y_{ij} 's are $N(\mu_i, \sigma_s^2 v + \sigma_p^2)$. For each site i, set

$$\begin{aligned} \frac{\sum_j y_{ij}}{n_p} &= \mu_i, \\ \frac{\sum_j {y_{ij}}^2}{n_p} &= \sigma_s^2 v + \sigma_p^2 + \mu_i^2. \end{aligned}$$

Solving for σ_s^2 suggests an estimate of the variance component associated with the spatial error at each site of

$$\tilde{\sigma}_{s_i}^2 = \left(\sum_j {y_{ij}}^2 / n_p - \mu_i^2 - \sigma_p^2 \right) / v.$$

Table 3: 100 Simulations of *anovsig* and *methmom*

	anovsig		*methmom*
	$\tilde{\sigma}_p^2$	$\tilde{\sigma}_s^2$	$\tilde{\sigma}_s^2$
Sample mean	27.5	-3.4×10^3	-8.9×10^3
Sample variance	127.9	7.6×10^8	2.5×10^9
Number of Negative Estimates.	0	56	59

Put $\tilde{\sigma}_s^2 = \sum_i \tilde{\sigma}_{s_i}^2 / n_s$. Then

$$\text{Var}(\tilde{\sigma}_s^2) = \frac{1}{n_s^2 v^2} \sum_{i=1}^{n_s} \sum_{r=1}^{n_s} 2(\sigma_s^2 c(h_{ir}))^2 + 4\mu_i \mu_r \sigma_s^2 c(h_{ir}).$$

Details of this derivation are shown in Lemma 2.3. Table 3 shows a large variance for $\tilde{\sigma}_s^2$ using the program *methmom* to obtain 100 estimates of $\tilde{\sigma}_s^2$. Again, a large number of negative estimates are generated. The simulations of *nummax, GSeidal, anovsig* and *methmom* are run using the same data set.

2.2.3 Algorithm

While explicit expressions, however ineffectual, can be obtained for σ_s^2, α cannot be isolated. At this point it seems clear not only that a more specialized approach is needed to estimate σ_s^2 and α but also that simultaneous estimation may not work reliably. We propose the procedure outlined below followed by comments on convergence whenever possible. The procedure estimates β_0, β_1 and σ_p^2 simultaneously then α and σ_s^2 separately using different approaches.

Setup

STEP 1: Select an initial estimate for σ_s^2 and α. Use $(X'X)^{-1}X'y$ and $SSE/(n_s n_p - n_s)$ for the initial guesses of b and σ_p^2 respectively. These familiar estimates are borrowed from standard regression theory.

Gauss-Seidal Procedure

STEP 2: Fix σ_s^2 and α. Consider the likelihood of y to be

a function of b and σ_p^2. Numerically maximize this likelihood to obtain the estimates $\hat{b}$ and $\widehat{\sigma_p^2}$.

STEP 3: Fix σ_s^2 and use $\hat{b}$ and $\widehat{\sigma_p^2}$ from step two. Consider the likelihood of y to be a function of α only. Numerically maximize this likelihood to obtain the estimate $\hat{\alpha}$.

STEP 4: Substitute the estimates obtained in Steps 2 and 3 to yield an estimate of σ_s^2 through regression by defining a new variable $D_{ir} = ((\bar{y}_{i\cdot} - \mu_i) - (\bar{y}_{r\cdot} - \mu_r))^2$, and regressing D_{ir} on a difference of covariances of spatial errors.

Iterative Step

STEP 5: Continue updating the estimates by repeating Steps 2 through 4, until convergence of $\hat{\alpha}$ and $\widehat{\sigma_s^2}$ is achieved.

Details for STEP 2 : Here σ_s^2 and α are fixed and the likelihood of y is considered to be a function of β_0, β_1 and σ_p^2 only. The built-in command **nlminb** in *S-plus* minimizes a non-linear function and requires a starting value. The program *opt3* uses this command to minimize the negative log-likelihood numerically. This is equivalent to maximimizing the likelihood. This seems to be reasonable in view of the good performance of **nlminb** in estimating β_0, β_1 and σ_p^2.

Details for STEP 3: Fix all the parameters except α and consider the likelihood as a function of α only. The program *optalpha* also uses the function **nlminb** to minimize the negative log-likelihood while imposing the bounds on α, i.e. $0 \leq \alpha \leq 1$. To analyze the performance of this program alone, suppose all the parameters except alpha are set to their real values in our 10-site Example 2.1.1. The program *optalpha*, which requires an initial guess for α, appears to be insensitive to the starting value of α and works well even for a small number of sites. Evidence of this is shown in Table 4.

For each initial guess, *optalpha* is run 200 times for both n_s=10 and n_s=4 sites using Example in Section 2.1.1 and $\Phi = (2, 15, 16, 49, .25)$. There were 29 boundary

Table 4: 200 Simulations of *optalpha*: This table shows that *optalpha* is insensitive to starting values.

		n_s= 10		n_s=4	
True α	Starting Value	Average Estimate	Standard Error	Average Estimate	Standard Error
.25	.1	.297652	.2565	.2842178	.3299
.25	.3	.297652	.2565	.2842178	.3299
.25	.75	.297652	.2565	.2842178	.3299
.25	1	.297652	.2565	.2842178	.3299

estimates in the case of $n_s = 10$ for all starting values, and 77 boundary estimates in the case $n_p = 4$ for all starting values.

Details for STEP 4: Finally, a different approach is used to estimate σ_s^2 which exploits the representation $\Sigma = \alpha C_1 + (1-\alpha)C_2$. Assume all parameters but σ_s^2 are known. Recalling the definitions $\mu_i = \beta_0 + \beta_1 x_i$, $v = \mathrm{Var}(\varepsilon_{s_i})$ and $c(h_{ir}) = \mathrm{Cov}(\varepsilon_{s_i}, \varepsilon_{s_r})$, then $y_{ij} = \mu_i + \sigma_s \varepsilon_{s_i} + \sigma_p e_{ij}$ with $y_{ij} \sim \mathcal{N}(\mu_i, \sigma_s^2 v + \sigma_p^2)$. Now define

$$D_{ir} = [(\bar{y}_{i\cdot} - \mu_i) - (\bar{y}_{r\cdot} - \mu_r)]^2, \ i, r = 1, \ldots, n_s.$$

Let $R_{ir} = 2\mathrm{Var}(\varepsilon_{s_i}) - 2\mathrm{Cov}(\varepsilon_{s_i}, \varepsilon_{s_r}) = 2v - 2c(h_{ir})$ and fit $D_{ir} - 2\sigma_p^2/n_p$ on R_{ir} by least squares with no intercept to obtain an estimate of the slope, σ_s^2. This estimate is both unbiased and consistent by Theorem 2.1 below.

Theorem 2.1 *For $i, i', r, r' = 1, \ldots, n_s$ define*

$$D = (D_{ir})_{n_s^2 \times 1}, \quad R = (R_{ir})_{n_s^2 \times 1}, \quad CST = (\tfrac{2\sigma_p^2}{n_p})_{n_s^2 \times 1},$$

and

$$g = g(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'}) = c(h_{ii'}) - c(h_{ir'}) - c(h_{ri'}) + c(h_{rr'}).$$

The estimate $\widehat{\sigma_s^2} = (R'R)^{-1}R'(D - CST)$ is unbiased and consistent provided

$$\frac{\sum_{i=1}^{n_s}\sum_{r=1}^{n_s}\sum_{i'=1}^{n_s}\sum_{r'=1}^{n_s} g(0, h_{ir}, h_{ir}, 0) g(0, h_{i'r'}, h_{i'r'}, 0) g^2(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'})}{\left[\sum_{i=1}^{n_s}\sum_{r=1}^{n_s} g^2(0, h_{ir}, h_{ir}, 0)\right]^2} \to 0$$

as both n_p and n_s go to infinity.

Proof: Note that $\bar{y}_{i\cdot} = \mu_i + \sigma_s \varepsilon_{s_i} + \sigma_p \bar{e}_{i\cdot}$ and $\bar{e}_{i\cdot} \sim \mathcal{N}(0, 1/n_p)$. Then

$$D_{ir} = [\sigma_s(\varepsilon_{s_i} - \varepsilon_{s_r}) - \sigma_p(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})]^2 . \tag{12}$$

Upon expanding and taking the expectations, the

$$\begin{aligned} E(D_{ir}) &= 2\sigma_s^2 v - 2\sigma_s^2 c(h_{ir}) + 2\frac{\sigma_p^2}{n_p}, \\ &= \sigma_s^2 R_{ir} + 2\frac{\sigma_p^2}{n_p}, \end{aligned} \tag{13}$$

for $i \neq r$ since the e_{ij}'s are independent of the ε_{s_i}'s, and zero otherwise. This implies

$$\begin{aligned} E(D) &= \sigma_s^2 R + CST, \\ \Rightarrow E(D - CST) &= \sigma_s^2 R. \end{aligned}$$

Therefore regression estimate $\widehat{\sigma_s^2} = (R'R)^{-1}R'(D - CST)$ is unbiased because

$$E((R'R)^{-1}R'(D - CST)) = (R'R)^{-1}R'R\sigma_s^2 = \sigma_s^2.$$

It remains to show that $\mathrm{Var}(\widehat{\sigma_s^2}) \to 0$ as n_p and n_s both go to infinity. Observe that the

$$\mathrm{Var}(\widehat{\sigma_s^2}) = (R'R)^{-1}R'\mathrm{Var}(D)R(R'R)^{-1'}.$$

The vector D will contain n_s zeros. If fact, $\mathrm{Var}(D)$ is an $n_s^2 \times n_s^2$ matrix whose ii^{th} rows and columns are zero. It will have $2n_s^3 - n_s^2$ zero entries in total. Expressions for the elements of $\mathrm{Var}(D) = (\mathrm{Cov}(D_{ir}, D_{i'r'}))_{n_s^2 \times n_s^2}$ must be found in order to explore the asymptotic behavior. In fact,

$$\mathrm{Cov}(D_{ir}, D_{i'r'}) \begin{cases} = 0 & \text{if } i = r \\ = 0 & \text{if } i' = r' \\ = \mathrm{Var}(D_{ir}) = 2\left(\sigma_s^2 R_{ir} + 2\sigma_p^2/n_p\right)^2 & \text{if } i = i',\ r = r',\ i \neq r \\ = \sigma_s^4(2g^2 + R_{ir}R_{ir'}) & \\ \quad +2\sigma_s^2\sigma_p^2(R_{ir} + R_{ir'} + 2g)/n_p + 6\sigma_p^4/n_p^2 & \\ \quad -(\sigma_s^2 R_{ir} + 2\sigma_p^2/n_p)(\sigma_s^2 R_{ir'} + 2\sigma_p^2/n_p) & \text{if } i = i',\ r \neq r',\ i \neq r \\ = \sigma_s^4(2g^2 + R_{ir}R_{i'r'}) & \\ \quad +2\sigma_s^2\sigma_p^2(R_{ir} + R_{i'r'})/n_p + 4\sigma_p^4/n_p^2 & \\ \quad -(\sigma_s^2 R_{ir} + 2\sigma_p^2/n_p)(\sigma_s^2 R_{i'r'} + 2\sigma_p^2/n_p) & \text{if } i \neq i' \end{cases} \tag{14}$$

The details for these expressions are given below.

First, $\mathrm{Var}(D_{ir})$ represents the non-zero diagonal entries in $\mathrm{Var}(D)$. It is calculated using (13) and the fact that

$$E(D_{ir}^2) = 3\left[\sigma_s^2 R_{ir} + \frac{2\sigma_p^2}{n_p}\right]^2.$$

This expectation can be calculated by brute force or by noting that the distribution of D_{ir} is $(\sigma_s^2 R_{ir} + 2\sigma_p^2/n_p)\chi_1^2$ with mean $\sigma_s^2 R_{ir} + 2\sigma_p^2/n_p$ and variance $2(\sigma_s^2 R_{ir} + 2\sigma_p^2/n_p)^2$.

Second, consider the term $\mathrm{Cov}(D_{ir}, D_{ir'}) = E(D_{ir}D_{ir'}) - E(D_{ir})E(D_{ir'})$ for $r \neq r'$. Using (12),

$$\begin{aligned} D_{ir}D_{ir'} &= \sigma_s^4(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\varepsilon_{s_i} - \varepsilon_{s_{r'}})^2 - 2\sigma_s^3\sigma_p(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\varepsilon_{s_i} - \varepsilon_{s_{r'}})(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot}) \\ &\quad + \sigma_s^2\sigma_p^2(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2 - 2\sigma_s^3\sigma_p(\varepsilon_{s_i} - \varepsilon_{s_r})(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})(\varepsilon_{s_i} - \varepsilon_{s_{r'}})^2 \\ &\quad + 4\sigma_s^2\sigma_p^2(\varepsilon_{s_i} - \varepsilon_{s_r})(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})(\varepsilon_{s_i} - \varepsilon_{s_{r'}})(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot}) \\ &\quad - 2\sigma_s\sigma_p^3(\varepsilon_{s_i} - \varepsilon_{s_r})(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2 + \sigma_s^2\sigma_p^2(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\varepsilon_{s_i} - \varepsilon_{s_{r'}})^2 \\ &\quad - 2\sigma_2\sigma_p^3(\varepsilon_{s_i} - \varepsilon_{s_{r'}})(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2 + \sigma_p^4(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2. \end{aligned}$$

Recall that the e's are independent of the ε's. Therefore, the expectations of the 2nd, 4th, 6th and 8th term above are all zero, leaving only

$$\begin{aligned} E(D_{ir}D_{ir'}) &= \sigma_s^4 E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\varepsilon_{s_i} - \varepsilon_{s_{r'}})^2 + \sigma_s^2\sigma_p^2 E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2 \\ &\quad + 4\sigma_s^2\sigma_p^2 E(\varepsilon_{s_i} - \varepsilon_{s_r})(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})(\varepsilon_{s_i} - \varepsilon_{s_{r'}})(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot}) \\ &\quad + \sigma_s^2\sigma_p^2 E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\varepsilon_{s_i} - \varepsilon_{s_{r'}})^2 + \sigma_p^4 E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2. \end{aligned} \tag{15}$$

The presence of dependent random variables in (15) complicates matters. To ease calculations, temporarily define $M = \bar{e}_{i\cdot} - \bar{e}_{r'\cdot}$ and $N = \bar{e}_{i\cdot} - \bar{e}_{r\cdot}$ so that $M \sim \mathcal{N}(0, 2/n_p)$ and $N \sim \mathcal{N}(0, 2/n_p)$ with $\mathrm{Cov}(M, N) = 1/n_p$. Also define $T = \varepsilon_{s_i} - \varepsilon_{s_{r'}}$ and $W = \varepsilon_{s_i} - \varepsilon_{s_r}$ so that $T \sim \mathcal{N}(0, R_{ir'})$ and $W \sim \mathcal{N}(0, R_{ir})$ with $\mathrm{Cov}(T, W) = g(0, h_{ir'}, h_{ri}, h_{rr'})$. Using multivariate normal theory, see Appendix A.4, $M|N \sim \mathcal{N}(N/2, 3/(2n_p))$ and

$$T|W \sim \mathcal{N}(g(0, h_{ir'}, h_{ri}, h_{rr'})W/R_{ir},\ R_{ir'} - g^2(0, h_{ir'}, h_{ri}, h_{rr'})/R_{ir}).$$

Now write

$$E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2 = E(N^2M^2)$$

$$\begin{aligned} &= E(N^2E(M^2|N)) = E\left(N^2\left(\frac{3}{2n_p}+\frac{N^2}{4}\right)\right) \\ &= \frac{3}{2n_p}E(N^2)+\frac{1}{4}E(N^4) \\ &= \frac{6}{n_p^2}. \end{aligned}$$

The value of $E(N^4)$ is obtained using the well-known fact that the *kth* moment of a random variable is equal to the *kth* derivative of its moment generating function evaluated at zero. See Appendix B.1 for the derivatives of the moment generating function of a normal random variable. Similarly,

$$\begin{aligned} E(\varepsilon_{s_i}-\varepsilon_{s_r})^2(\varepsilon_{s_i}-\varepsilon_{s_{r'}})^2 &= E(W^2T^2) = E(W^2E(T^2|W)) \\ &= 2g^2(0,h_{ir'},h_{ri},h_{rr'})+R_{ir'}R_{ir}. \end{aligned}$$

Next,

$$\begin{aligned} E(\varepsilon_{s_i}-\varepsilon_{s_r})^2(\bar{e}_{i\cdot}-\bar{e}_{r'\cdot})^2 &= E(W^2)E(M^2) = 2R_{ir}/n_p. \\ E(\varepsilon_{s_i}-\varepsilon_{s_{r'}})^2(\bar{e}_{i\cdot}-\bar{e}_{r\cdot})^2 &= 2R_{ir'}/n_p. \\ E(\varepsilon_{s_i}-\varepsilon_{s_r})(\bar{e}_{i\cdot}-\bar{e}_{r\cdot})(\varepsilon_{s_i}-\varepsilon_{s_{r'}})(\bar{e}_{i\cdot}-\bar{e}_{r'\cdot}) &= E(WNTM) \\ &= E(WT)E(NM) \\ &= g(0,h_{ir'},h_{ri},h_{rr'})/n_p, \end{aligned}$$

by writing $E(WT) = E(WE(T|W))$ and $E(NM) = E(NE(M|N))$. Line (15) can now be expressed as

$$\begin{aligned} E(D_{ir}D_{ir'}) &= \sigma_s^4(2g^2(0,h_{ir'},h_{ri},h_{rr'})+R_{ir}R_{ir'}) \\ &\quad +\sigma_s^2\sigma_p^2(2R_{ir}+2R_{ir'}+4g(0,h_{ir'},h_{ri},h_{rr'}))/n_p+6\sigma_p^4/n_p^2, \end{aligned}$$

implying

$$\begin{aligned} \mathrm{Cov}(D_{ir}D_{ir'}) &= \sigma_s^4(2g^2(0,h_{ib},h_{ri},h_{rr'})+R_{ir}R_{ir'}) \\ &\quad +\sigma_s^2\sigma_p^2(2R_{ir}+2R_{ir'}+4g(0,h_{ir'},h_{ri},h_{rr'}))/n_p+6\sigma_p^4/n_p^2 \\ &\quad -(\sigma_s^2R_{ir}+2\sigma_p^2/n_p)(\sigma_s^2R_{ir'}+2\sigma_p^2/n_p). \end{aligned}$$

Lastly, consider $\mathrm{Cov}(D_{ir}, D_{i'r'})$ for $i \neq i'$. The calculation mirrors that of the previous covariance term with slight differences. The value of

$$\begin{aligned} E(D_{ir}D_{i'r'}) &= \sigma_s^4 E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\varepsilon_{s_{i'}} - \varepsilon_{s_{r'}})^2 + \sigma_s^2\sigma_p^2 E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\bar{e}_{i'\cdot} - \bar{e}_{r'\cdot})^2 \\ &+ \sigma_s^2\sigma_p^2 E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\varepsilon_{s_{i'}} - \varepsilon_{s_r})^2 + \sigma_p^4 E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\bar{e}_{i'\cdot} - \bar{e}_{r'\cdot})^2. \quad (16) \end{aligned}$$

Each of the expectations in (16) can be calculated as before using multivariate normal theory. In fact,

$$\begin{aligned} E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\varepsilon_{s_{i'}} - \varepsilon_{s_{r'}})^2 &= 2g^2(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'}) + R_{ir}R_{i'r'} \\ E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\bar{e}_{i'\cdot} - \bar{e}_{r'\cdot})^2 &= 2R_{ir}/n_p \\ E(\varepsilon_{s_{i'}} - \varepsilon_{s_{r'}})^2(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2 &= 2R_{i'r'}/n_p \\ E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\bar{e}_{i'\cdot} - \bar{e}_{r'\cdot})^2 &= 4/n_p^2. \end{aligned}$$

Thus, for $i \neq i'$ the

$$\begin{aligned} \mathrm{Cov}(D_{ir}, D_{i'r'}) &= \sigma_s^4(2g^2(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'}) + R_{ir}R_{i'r'}) \\ &+2\sigma_s^2\sigma_p^2(R_{ir} + R_{i'r'})/n_p + 4\sigma_p^4/n_p^2 \\ &-(\sigma_s^2 R_{ir} + 2\sigma_p^2/n_p)(\sigma_s^2 R_{i'r'} + 2\sigma_p^2/n_p). \end{aligned}$$

Consequently, $\mathrm{Var}(D)$ is given by (14).

Now notice that $R'R = \sum_{i=1}^{n_s}\sum_{r=1}^{n_s} R_{ir}^2$, and that R_{ir} can be written in terms of g. Specifically, $R_{ir} = g(0, h_{ir}, h_{ir}, 0)$. Consider

$$\begin{aligned} \mathrm{Var}(\widehat{\sigma_s^2}) &= det\,\mathrm{Var}(\widehat{\sigma_s^2}) \\ &= det\,(R'R)^{-1} det\,(R'\mathrm{Var}(D)R)\, det\,(R'R)^{-1} \\ &= \frac{tr\,(R'\mathrm{Var}(D)R)}{[det\,(R'R)]^2} \\ &= \frac{tr\,(RR'\mathrm{Var}(D))}{(R'R)^2}, \end{aligned}$$

since $R'\mathrm{Var}(D)R$ is a 1×1. Only the diagonal elements of $RR'\mathrm{Var}(D)$ are required for the trace. A typical diagonal element is characterized by

$$\sum_{i'=1}^{n_s}\sum_{r'=1}^{n_s} R_{ir}R_{i'r'}\mathrm{Cov}(D_{ir}, D_{i'r'}),$$

which implies

$$tr(RR'\mathrm{Var}(D)) = \sum_{i=1}^{n_s}\sum_{r=1}^{n_s}\sum_{i'=1}^{n_s}\sum_{r'=1}^{n_s} R_{ir}R_{i'r'}\mathrm{Cov}(D_{ir}, D_{i'r'}).$$

Finally, the value of $\mathrm{Cov}(D_{ir}, D_{i'r'}) \to 2\sigma_s^4 g^2(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'})$ as $n_p \to \infty$. Hence the $\mathrm{Var}(\widehat{\sigma_s^2})$ goes to zero as n_p and n_s go to infinity provided

$$\frac{\sum_{i=1}^{n_s}\sum_{r=1}^{n_s}\sum_{i'=1}^{n_s}\sum_{r'=1}^{n_s} g(0, h_{ir}, h_{ir}, 0)g(0, h_{i'r'}, h_{i'r'}, 0)g^2(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'})}{[\sum_{i=1}^{n_s}\sum_{r=1}^{n_s} g^2(0, h_{ir}, h_{ir}, 0)]^2} \to 0. \tag{17}$$

■

Notice that the consistency condition requires that both n_s and n_p go to infinity. It is difficult to interpret this condition further without imposing a structure on how the number of sites grows to infinity. The following discussion assumes a particular placement of an increasing number of sites.

Transect Example: Suppose the sample area is in the form of a long continuous strip. Locations are equi-distant, unit spaced say, and sites are being added at the end of the line. In this scenario, it is possible to simplify some of the expression in the consistency condition (17). Recall that $g(0, h_{ir}, h_{ir}, 0) = 2v - 2c(h_{ir})$. After some simplification,

$$\sum_i\sum_r g^2(0, h_{ir}, h_{ir}, 0) = (4)2n_s\sum_{i=1}^{n_s-1}(1 - \frac{i}{n_s})(v - c(i))^2.$$

To approximate this sum, let $x_i = i/n_s$ so that

$$\sum_i\sum_r g^2(0, h_{ir}, h_{ir}, 0) \approx 2n_s^2\int_0^1 (1-x)(v - c(n_s x))^2 dx.$$

The covariance structure is such that $c(h) \to 0$ as $h \to \infty$. Therefore, $\sum\sum g^2 = O(n_s^2 v^2)$ which implies that the denominator of (17) is of order $n_s^4 v^4$. Thus, (17) will hold if the numerator is $O(n_s^4)$.

The consistency condition is checked numerically for other examples introduced in Section 4.4.1 under the topic of asymptotics.

Table 5: 100 Simulations of *optsigs*

n_s	n_p	True σ_s^2	Sample Mean	Standard Error	Number of Negative Estimates
10	2	49	51.27	42.28	7
10	5	49	53.86	39.11	4
25	2	49	52.06	26.52	0
25	5	49	53	26.54	0

The program *optsigs* performs STEP 4 of the algorithm. Occasionally, negative estimates for σ_s^2 are obtained. In classical analysis of variance components there exist well-developed theories to handle this situation. We will not explore them here. To analyze the performance of this program alone, suppose all the parameters except σ_s^2 are set to their real values in Example 2.1.1. Results of the 100 repetitions for each of the n_s, n_p combinations are shown in Table 5. While the variance is still quite large, the mean is well estimated and the number of infeasible estimates for σ_s^2 seems to decrease with increasing n_s and n_p.

Algorithm Simulation

All five steps of the algorithm are combined into one program called *optall*. Negative estimates of σ_s^2 cause the program to exit the iteration (step 5) immediately. The maximum number of allowable iterations to convergence is set to 75. The initial values for σ_s^2 and α are 81 and .1 respectively. This program is run 100 times for various combinations of n_s and n_p. In the case of $n_s = 15$ five additional sites are added with the folowing locations and predictor values:

Site	Predictor
s_{11}= (-3,12)	$x_{11} = 89$
s_{12}= (1,5)	$x_{12} = 505$
s_{13}= (6,2)	$x_{13} = 33$
s_{14}= (2,9)	$x_{14} = 185$
s_{15}= (0,-5)	$x_{15} = 250$

Table 6: 100 Simulations of *optall*

		n_s=10 n_p=2			n_s=15 n_p=5		
Φ	True value	Sample Mean	Standard Error	Number of Infeasible Estimates	Sample Mean	Standard Error	Number of Infeasible Estimates
β_0	2	1.6	3.83		1.89	2.78	
β_1	15	15	.01		15	.006	
σ_p^2	16	15	6.25		16.33	2.94	
σ_s^2	49	109	195.4	9	77.45	90.64	1
α	.25	.39	.44	60	.41	.41	29

One simulation did not converge for $n_s = 10$. Simulation results shown in Table 6 suggest adequate estimates for b and σ_p^2. The number of infeasible estimates for σ_s^2 drops from 9 in the case of $n_s = 10$ and $n_p = 2$ to 1 for $n_s = 15$ and $n_p = 5$. The number of boundary estimates for α drops from 60 to 29 with increasing n_s and n_p. Six estimates of Φ have boundary values for alpha and negative estimes for σ_s^2 when $n_s = 10$ and $n_p = 2$, comparing to 0 in the case of $n_s = 15$ and $n_p = 5$.

2.2.4 Simulation Comparisons

The goal of this section is to compare the estimation procedures outlined in the previous two subsections. This is done through small sample simulations programmed in *S-plus*. Observations y_{ij} coming from a $\mathcal{N}(\mathrm{Xb}, \mathrm{V})$ distribution are required in order to estimate the parameter Φ. The realizations are generated by the program *simulatey* working with our Example of Section 2.1.1 and setting $\Phi = (2, 15, 16, 49, .25)$. These are then used in the simulations.

The results are based on 100 simulations with a small sample size of n_s=10 and n_p= 2. All three programs use the same generated set of observations, locations and predictor variables. The same initial values for θ are used in both the programs *optall* and *nummax* while *anovsig* runs on the real values for all the known parameters. The program *optall* runs our preferred estimation procedure. *Nummax* performs a numerical maximization with the likelihood of the y as a function of all 5 parameters

wheras *anovasig* applies the analysis of variance technique to obtain estimates of σ_p^2 and σ_s^2.

From the simulation results with $n_s = 10$ and $n_p = 2$, Table 8 gives a summary of the number of infeasible estimates for σ_s^2 and boundary estimates for α. One advantage of *optall* is that it yields many more feasible estimates of σ_s^2 and fewer boundary values of α than *nummax*. An investigation of Table 7 shows comparable estimates between *optall* and *nummax* for b and σ_p^2. It is interesting to note that *optall* and *nummax* have approximately the same order of standard deviation for these estimates. Although *nummax* performs somewhat better in the estimation of σ_s^2 (among feasible cases) than does *optall*, the reverse is true for α. We propose the usage of the *optall* algorithm rather than the optimization of the likelihood function based on its greater stability in the estimation of σ_s^2 and α.

Table 7: Simulations Comparisons

a) Using all 99 estimates from *optall* and all 100 estimates from *nummax*

		Optall		Nummax		Anovasig	
Φ	True value	Avg estimate	Std error	Avg estimate	Std error	Avg estimate	Std error
β_0	2	1.6	3.8	1.6	3.6		
β_1	15	15	.01	15	.0089		
σ_p^2	16	15	6.25	16	7	28	11
σ_s^2	49	109	195	16	38	-.0034	.00027
α	.25	.39	.442	.58	.47		

b) Summary statistics for $\hat{\Phi}$ involving only feasible estimates for σ_s^2 using 90 estimates from *optall* and 59 estimates from *nummax*.

		Optall		Nummax	
Φ	True value	Avg estimate	Std error	Avg estimate	Std error
β_0	2	1.6	3.9	1.5	3.6
β_1	15	15	.01	15	.0089
σ_p^2	16	15	6.43	13	5.69
σ_s^2	49	122	201	27	46.73
α	.25	.42	.446	.93	.25

Table 8: Comparison of Infeasible and Boundary Estimates

	optall	*nummax*
% of infeasible σ_s^2	9	41
% of boundary values for α	60	78
% of both infeasible/boundary values for σ_s^2 and α	6	20
Average of feasible σ_s^2 estimates	122.36	27.44
Standard error of feasible σ_s^2 estimates	200.63	46.72

2.3 Technical Results

Consider the model

$$y_{ij} = \beta_0 + \beta_1 x_i + \sigma_s \varepsilon_{s_i} + \sigma_p e_{ij} ,$$

for $i = 1, \ldots, n_s$ and $j = 1, \ldots, n_p$. Let $\mu_i = \beta_o + \beta_1 x_i$, $\xi_i = \sigma_s \varepsilon_{s_i}$ and $\zeta_{ij} = \sigma_p e_{ij}$ so that

$$y_{ij} = \mu_i + \xi_i + \zeta_{ij}, \tag{18}$$

with independent errors $\zeta_{ij} \sim \mathcal{N}(0, \sigma_p^2)$ and correlated spatial errors $\xi_i \sim \mathcal{N}(0, \sigma_s^2 v)$. Denote the $\mathrm{Cov}(\xi_i, \xi_{i'}) = \sigma_s^2 c(h_{ii'})$. Assume further that the ξ's are independent of the ζ's. Lemma 2.2 follows the lines of Searle (1971). However, the dependent errors produce additional terms.

Lemma 2.2 *Define* $SSA = n_p \sum_{i=1}^{n_s} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ *and* $SSE = \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2$. *Then*

$$\begin{aligned} E(SSE) &= (n_s n_p - n_s)\sigma_p^2, \\ E(SSA) &= n_p \sum_i (\mu_i - \bar{\mu}_\cdot)^2 + \left[n_s n_p v - n_p n_s^{-1} \sum_i \sum_{i'} c(h_{ii'}) \right] \sigma_s^2 + (n_s - 1)\sigma_p^2. \end{aligned}$$

Proof: From (18), $\bar{y}_{i\cdot} = \mu_i + \xi_i + \bar{\zeta}_{i\cdot}$, and $\bar{y}_{\cdot\cdot} = \bar{\mu}_\cdot + \bar{\xi}_\cdot + \bar{\zeta}_{\cdot\cdot}$. Substituting these expressions into SSA and SSE gives

$$\begin{aligned} SSA &= n_p \sum_i \left[(\mu_i - \bar{\mu}_\cdot) + (\xi_i - \bar{\xi}_\cdot) + (\bar{\zeta}_{i\cdot} - \bar{\zeta}_{\cdot\cdot}) \right]^2, \\ SSE &= \sum_i \sum_j (\zeta_{ij} - \bar{\zeta}_{i\cdot})^2. \end{aligned}$$

Expanding the square and taking the expectations yields

$$\begin{aligned} E(SSA) &= n_p \sum_i \left[(\mu_i - \bar{\mu}_\cdot)^2 + E(\xi_i - \bar{\xi}_\cdot)^2 + E(\bar{\zeta}_{i\cdot} - \bar{\zeta}_{\cdot\cdot})^2 \right], \\ E(SSE) &= \sum_i \sum_j E(\zeta_{ij}^2) - n_p E\left[\sum_i (\bar{\zeta}_{i\cdot})^2 \right], \end{aligned}$$

since both ζ and ξ have mean zero and are independent of one another. Now

$$\begin{aligned} E(\xi_i - \bar{\xi}_\cdot)^2 &= \sigma_s^2 v - 2\sigma_s^2 n_s^{-1} \textstyle\sum_{i'=1}^{n_s} c(h_{ii'}) + \sigma_s^2 n_s^{-2} \sum_{r=1}^{n_s} \sum_{s=1}^{n_s} c(h_{rs}), \\ E(\bar{\zeta}_{i\cdot} - \bar{\zeta}_{\cdot\cdot})^2 &= (n_s - 1) n_s^{-1} n_p^{-1} \sigma_p^2, \\ E\left[\textstyle\sum_i (\bar{\zeta}_{i\cdot})^2 \right] &= n_s \sigma_p^2 / n_p. \end{aligned}$$

Finally, incorporating these calculations produces the desired results. ■

Lemma 2.3 *Given that the dependent y_{ij}'s are $\mathcal{N}(\mu_i, \sigma_s^2 v + \sigma_p^2)$, put $\tilde{\sigma}_s^2 = \sum_i \tilde{\sigma}_{s_i}^2 / n_s$, where*

$$\tilde{\sigma}_{s_i}^2 = \frac{1}{v}\left(\frac{\sum_j y_{ij}^2}{n_p} - \mu_i^2 - \sigma_p^2\right).$$

Then

$$\mathrm{Var}(\tilde{\sigma}_s^2) = \frac{1}{n_s^2 v^2}\sum_{i=1}^{n_s}\sum_{r=1}^{n_s} 2(\sigma_s^2 c(h_{ir}))^2 + 4\mu_i\mu_r\sigma_s^2 c(h_{ir}).$$

Proof:

$$\mathrm{Var}(\tilde{\sigma}_s^2) = \frac{1}{n_s^2}\mathrm{Var}(\sum_i \tilde{\sigma}_{s_i}^2) = \frac{1}{n_s^2}\sum_{i=1}^{n_s}\sum_{r=1}^{n_s}\mathrm{Cov}(\tilde{\sigma}_{s_i}^2, \tilde{\sigma}_{s_r}^2).$$

Consider the covariance term alone. Then

$$\begin{aligned}\mathrm{Cov}(\tilde{\sigma}_{s_i}^2, \tilde{\sigma}_{s_r}^2) &= \frac{1}{v^2 n_p^2}\mathrm{Cov}(\sum_j y_{ij}^2, \sum_j y_{rj}^2) \\ &= \frac{1}{v^2 n_p^2}\sum_{j=1}^{n_p}\sum_{t=1}^{n_p}\mathrm{Cov}(y_{ij}^2, y_{rt}^2).\end{aligned}$$

Now the

$$\begin{aligned}\mathrm{Cov}(y_{ij}^2, y_{rt}^2) &= \mathrm{Cov}(\xi_i^2, \xi_r^2) + 2\mu_r\mathrm{Cov}(\xi_i^2, \xi_r) \\ &+ 2\mu_i\mathrm{Cov}(\xi_r^2, \xi_i) + 4\mu_i\mu_r\mathrm{Cov}(\xi_i, \xi_r). \qquad (19)\end{aligned}$$

Applying multivariate normal theory gives the conditional distribution of $\xi_r|\xi_i \sim \mathcal{N}(c(h_{ri})v^{-1}\xi_i, \sigma_s^2 v - \sigma_s^2(c(h_{ir}))^2 v^{-1})$ The middle terms of (19) reduce to zero since

$$\begin{aligned}\mathrm{Cov}(\xi_i^2, \xi_r) &= E(\xi_i^2, \xi_r) = E(\xi_i^2 E(\xi_r|\xi_i)) \\ &= c(h_{ir})v^{-1}E(\xi_i^3) = 0.\end{aligned}$$

The third moment is calculated by the third derivative of the moment generating function evaluated at zero. Similarly, the

$$\begin{aligned}\mathrm{Cov}(\xi_i^2, \xi_r^2) &= E(\xi_i^2, \xi_r^2) - E(\xi_i^2)E(\xi_r^2) \\ &= E(\xi_i^2 E(\xi_r^2|\xi_i)) - (\sigma_s^2 v)^2 \\ &= 2(\sigma_s^2 c(h_{ir}))^2.\end{aligned}$$

Equation (19) becomes

$$\mathrm{Cov}(y_{ij}^2, y_{rt}^2) = 2(\sigma_s^2 c(h_{ir}))^2 + 4\mu_i\mu_r\sigma_s^2 c(h_{ir}).$$

Finally, the

$$\mathrm{Var}(\tilde{\sigma}_s^2) = \frac{1}{n_s^2 v^2}\sum_{i=1}^{n_s}\sum_{r=1}^{n_s} 2(\sigma_s^2 c(h_{ir}))^2 + 4\mu_i\mu_r\sigma_s^2 c(h_{ir}).$$

■

Chapter 3

An Inverse for V

Recall $V = \sigma_p^2(I+\delta A)$ with $A = \Sigma \otimes U$ from line (7). The matrix $V^{-1} = (I+\delta A)^{-1}/\sigma_p^2$ in this form makes analysis of the asymptotic variance of parameter estimates difficult. In this chapter we obtain a formula for V^{-1} whose utility will become apparent in Chapter 4. The formula arises from a consideration of Taylor series expansion.

3.1 Taylor Series Expansion of V^{-1}

In an ordinary random effects model the variance-covariance matrix of observations is block diagonal by construction. That is, the covariance of y_{ij} and $y_{i'j'}$ when $i \neq i'$ is usually taken to be zero. Due to the spatial dependence however, our covariance is not zero as is shown by (6), and the matrix V is not block diagonal. Maximum likelihood estimation would require the inverse of V, but this form at present is difficult to manipulate. Mardia and Marshall (1984) acknowledge that the computation time for calculating V^{-1} increases as the number of sites increase.

Instead, we consider the Taylor series expansion for V^{-1} obtained by applying Appendix A.3. For $V = \sigma_p^2 I + \sigma_s^2 A$ with $A = \Sigma \otimes U$,

$$V^{-1} = (\sigma_p^2(I + \delta A))^{-1} = \frac{1}{\sigma_p^2}\sum_{r=0}^{\infty}(-\delta A)^r, \tag{20}$$

with $A^0 = I$. Before considering this expression, we must first explore the convergence

of the series. In fact, it converges provided

$$\rho(-\delta \mathrm{A}) = \delta\rho(\mathrm{A}) < 1 \tag{21}$$

where $\rho(\mathrm{A})$ is the spectral radius of A. This convergence condition can be expressed in terms of σ_p^2, σ_s^2, α and the spectral radius of C_1 and C_2. Denote by $\lambda_h(M)$, $h = 1, \ldots, m$ the eigenvalues of a $m \times m$ matrix M and by $\lambda_{(h)}(M)$ the ordered eigenvalues. Recall that covariograms C_1 and C_2 are positive definite so that for $0 < \alpha < 1$, $\Sigma = \alpha\mathrm{C}_1 + (1-\alpha)\mathrm{C}_2$ is positive definite. The following inequality holds for the ordered eigenvalues of the sum of a symmetric matrix and one that is positive definite. See Theorem A.1.

$$\begin{aligned}
\lambda_{(i)}(\Sigma) &= \lambda_{(i)}(\alpha\mathrm{C}_1 + (1-\alpha)\mathrm{C}_2) > \lambda_{(i)}(\alpha\mathrm{C}_1) &&= \alpha\lambda_{(i)}(\mathrm{C}_1)\ , \\
\lambda_{(i)}(\Sigma) &= \lambda_{(i)}((1-\alpha)\mathrm{C}_2 + \alpha\mathrm{C}_1) > \lambda_{(i)}((1-\alpha)\mathrm{C}_2) &&= (1-\alpha)\lambda_{(i)}(\mathrm{C}_2)\ \forall i.
\end{aligned}$$

Therefore,

$$\lambda_{(i)}(\Sigma) > \max(\alpha\lambda_{(i)}(\mathrm{C}_1), (1-\alpha)\lambda_{(i)}(\mathrm{C}_2))\ \ \forall i, \tag{22}$$

$$\Rightarrow \rho(\Sigma) > \max(\alpha\rho(\mathrm{C}_1), (1-\alpha)\rho(\mathrm{C}_2)). \tag{23}$$

Line (22) holds for all i and, in particular, for the spectral radius since the eigenvalues of a positive definite matrix are positive. Now the n_p eigenvalues of U consist of $(n_p - 1)$ zeros and n_p. Thus, according to Theorem A.2, the eigenvalues of $\mathrm{A} = \Sigma \otimes \mathrm{U}$ are $(0, 0, \ldots, 0, n_p\lambda_1(\Sigma), \ldots, n_p\lambda_{n_s}(\Sigma))$. Notice that A is positive semi-definite. Specifically, there are $n_s n_p - n_s$ zeros and n_s values of the form $n_p\lambda_i(\Sigma)$. Combining these results with the inequalities (21) and (23) yields

$$\delta n_p \max(\alpha\rho(\mathrm{C}_1), (1-\alpha)\rho(\mathrm{C}_2)) < \delta n_p\rho(\Sigma) = \delta\rho(A) < 1. \tag{24}$$

Rewriting the convergence condition in terms of σ_p^2, σ_s^2 and $\alpha \in [0, 1]$ gives sufficient conditions for the Taylor series expansion to converge.

Lemma 3.1 *If*

$$\sigma_s^2 < \frac{\sigma_p^2}{n_p \max(\alpha\rho(\mathrm{C}_1), (1-\alpha)\rho(\mathrm{C}_2))} \tag{25}$$

then the Taylor series expansion in (20) converges.

Proof: Rearranging (24) gives the desired result. ■

The convergence condition requires that σ_s^2 be much smaller than σ_p^2. Since we are centering about the matrix 0, the Taylor approximation works best for Σ close to the zero matrix, i.e. σ_s^2 is very small. The approximation should be better if centered at B closer to Σ. In an effort to improve this constraint, various centerings for the Taylor series expansion are investigated. Write $V = \sigma_p^2(B + V/\sigma_p^2 - B)$, for B non-singular. Now the Taylor expansion centered about B is given by

$$ {}_BV^{-1} = \frac{1}{\sigma_p^2}\{B^{-1}\sum_{r=0}^{\infty}(-B)^{-r}(\frac{1}{\sigma_p^2}V - B)^r\}. $$

If B^{-1} commutes with V then the above expression can be written as

$$ {}_BV^{-1} = \frac{1}{\sigma_p^2}B^{-1}\sum_{r=0}^{\infty}\left[(-B)^{-1}(\frac{1}{\sigma_p^2}V - B)\right]^r = \frac{1}{\sigma_p^2}B^{-1}\sum_{r=0}^{\infty}[I - \frac{1}{\sigma_p^2}B^{-1}V]^r \qquad (26) $$

which converges if and only if $\rho(-B^{-1}(V/\sigma_p^2 - B)) < 1$. This condition can be simplified to

$$ 0 < \lambda_h(B^{-1}V) < 2\sigma_p^2, \quad h = 1, \ldots, n_s n_p. \qquad (27) $$

There are, of course, many possible choices for B. The following theorem explores (26) for B of a particular form.

Theorem 3.1 *Let* $V = \sigma_p^2 I_{n_s n_p} + \sigma_s^2(\Sigma_{n_s} \otimes U_{n_p})$ *with* σ_p^2 *and* σ_s^2 *known. Suppose* B *is such that* $B^{-1} = I + \kappa(f(\Sigma) \otimes U)$ *where* κ *is a constant and* f *is a polynomial in* Σ *which commutes with* Σ. *Then for* $i = 1, \ldots, n_s$, ${}_BV^{-1}$ *given in (26) converges provided*

$$ \frac{\sigma_s^2}{\sigma_p^2} < \frac{1 - \kappa n_p f(\lambda_i(\Sigma))}{[1 + \kappa n_p f(\lambda_i(\Sigma))]\ n_p \lambda_i(\Sigma)} $$

when $1 + \kappa n_p f(\lambda_i(\Sigma)) > 0$.

Proof: It is necessary to show that the convergence condition (27) reduces to that in the theorem. Using the properties of Kronecker products, express $B^{-1}V$ as

$$ \begin{aligned} B^{-1}V &= \sigma_p^2 I + \sigma_s^2(\Sigma \otimes U) + \kappa\sigma_p^2(f(\Sigma) \otimes U) + n_p\kappa\sigma_s^2(f(\Sigma)\Sigma \otimes U) \\ &= \sigma_p^2 I + \left[\underbrace{(\sigma_s^2\Sigma + \kappa\sigma_p^2 f(\Sigma) + n_p\kappa\sigma_s^2 f(\Sigma)\Sigma)}_{M} \otimes U\right]. \end{aligned} $$

Then by Lemma A.1 and Theorem A.2, $\lambda_t(\mathrm{B}^{-1}\mathrm{V}) = \sigma_p^2 + \lambda_i(M)\lambda_j(\mathrm{U})$ for $j = 1, \ldots, n_p$, $i = 1, \ldots, n_s$, and $t = 1, \ldots, n_s n_p$. By construction of U and M their eigenvalues are

$$\begin{aligned} \lambda_j(\mathrm{U}) &= (\overbrace{0, \ldots, 0}^{n_p - 1}, n_p) \\ \lambda_i(M) &= \sigma_s^2 \lambda_i(\Sigma) + \kappa \sigma_p^2 f(\lambda_i(\Sigma)) + n_p \kappa \sigma_s^2 f(\lambda_i(\Sigma)) \lambda_i(\Sigma). \end{aligned}$$

Therefore the eigenvalues of $\mathrm{B}^{-1}\mathrm{V}$ consist of σ_p^2 repeated $n_s(n_p\text{-}1)$ times and n_s values of the form $\sigma_p^2 + n_p \lambda_i(M)$. These must all satisfy (27). Obviously, $0 < \sigma_p^2 < 2\sigma_p^2$ so we require only that $0 < \sigma_p^2 + n_p \lambda_i(M) < 2\sigma_p^2, \forall i$. This reduces to $\lambda_i(M) < \sigma_p^2 / n_p$. Substituting the expression for the eigenvalues of M and manipulating algebraically yields the desired condition. ■

Corollary 3.1 *Let* $\mathrm{B} = \mathrm{I} + \delta(\mathrm{I} \otimes \mathrm{U})$. *Then the convergence condition for* ${}_B\mathrm{V}^{-1}$ *in (26) becomes* $\delta < 1/(\rho(A) - 2n_p)$.

Proof: For B of the given form, $\mathrm{B}^{-1} = I + -\delta/(1 + \delta n_p)(\mathrm{I} \otimes \mathrm{U})$. Set $\kappa = -\delta/(1 + \delta n_p)$ and $f(\Sigma) = I$. Since $f(\lambda_i(\Sigma)) = 1 \ \forall i$ the result follows by application of Theorem 3.1. ■

The convergence conditions can be rewritten as

$$\begin{aligned} \text{Equation (21)}: &\quad \delta < 1/(n_p \rho(\Sigma)), \\ \text{Condition in Corollary 3.1}: &\quad \delta < 1/(n_p(\rho(\Sigma) - 2)). \end{aligned}$$

In this form, it is clear that to use Corollary 3.1, $\rho(\Sigma)$ must be greater than 2 as $\delta > 0$. These conditions can also be seen as a function of C_1 and C_2 by using (23). The spectral radius for Example 2.1.1 with $n_s = 10, n_p = 2$ and $\alpha = .25$ is $\rho(A) = 3.5194$. The convergence conditions become $\delta < .284$ and $\delta < -2.081$ for the uncentered and centered Taylor series about $\mathrm{B} = \mathrm{I} + \delta(\mathrm{I} \otimes \mathrm{U})$, respectively. Therefore, it is not possible to use the Taylor series expansion centered about $\mathrm{B} = \mathrm{I} + \delta(\mathrm{I} \otimes \mathrm{U})$ for the inverse of V in this example.

3.2 Approximating V^{-1}

Here we will explore candidates for approximations to V^{-1}. If a good approximation were found, it could be used in the likelihood function instead of V^{-1} to obtain a more tractable expression. To find an approximation of V^{-1} we consider truncating the series in (20). One basic requirement of the approximate inverse is that it be positive definite. Under what conditions is this so? To answer this question, we again manipulate eigenvalues.

First consider the first order series expansion $V_1^{-1} = (I - \delta A)/\sigma_p^2$. Using Lemma A.1, the eigenvalues of this matrix are $\sigma_p^{-2} + (-\sigma_s^2/\sigma_p^4)\lambda_h(A)$ where $h = 1, \ldots, n_s n_p$. To ensure positive definiteness, the eigenvalues must be positive and are set greater than zero. In particular, when setting the eigenvalue $\sigma_p^{-2} + (-\sigma_s^2/\sigma_p^4)\rho(A)$ greater than zero, the inequality (25) re-emerges. Namely, the positive definite condition for V_1^{-1} is identical to the convergence condition for V^{-1}.

Next, look at the second order series expansion $V_2^{-1} = (I - \delta A + (\delta A)^2)/\sigma_p^2$. In this case, the eigenvalues are $\sigma_p^{-2} + (-\sigma_s^2/\sigma_p^4)\lambda_h(A) + (\sigma_s^4/\sigma_p^6)\lambda_h(A)$ where $h = 1, \ldots, n_s n_p$. Setting these values greater than zero yields

$$(\lambda_h(A)\sigma_s^2)^2 - \lambda_h(A)\sigma_s^2\sigma_p^2 + (\sigma_p^2)^2 = (\lambda_h(A)\sigma_s^2 - \sigma_p^2)^2 + \lambda_h(A)\sigma_s^2\sigma_p^2 > 0, \quad \forall h.$$

This condition is always satisfied for A positive semi-definite. Therefore, V_2^{-1} is always positive definite.

Now we investigate the first order Taylor expansion centered about B, denoted ${}_B V_1^{-1}$, in the case where B^{-1} and V commute. It is given by

$${}_B V_1^{-1} = \left(2B^{-1} - B^{-2}(I + \delta A)\right)/\sigma_p^2. \tag{28}$$

This will be positive definite if and only if the eigenvalues of (28) are positive. Finally, consider the second order Taylor expansion centered about B, denoted ${}_B V_2^{-1}$ with B^{-1} and V commuting. The conditions for positive definiteness here require that the eigenvalues of

$${}_B V_2^{-1} = \left(3B^{-1} - 3B^{-2}(I + \delta A) + B^{-3}(I + \delta A)^2\right)/\sigma_p^2 \tag{29}$$

be positive.

Now suppose $B^{-1} = I + \kappa(f(\Sigma) \otimes U)$ where κ is a constant and the polynomial $f(\Sigma)$ commutes with Σ. For this choice of B^{-1}, the matrix in (28) is positive definite if the eigenvalues,

$$1 + \{-n_p\delta\lambda_i - \kappa^2 n_p^2 f^2(\lambda_i) - 2\kappa\delta n_p^2 f(\lambda_i)\lambda_i - \kappa^2\delta n_p^3 f^2(\lambda_i)\lambda_i\}$$

are positive for all $i = 1, \ldots, n_s$. Denote $\lambda_i \equiv \lambda_i(\Sigma)$ here for simplicity. Similarly, a sufficient condition for the matrix in (29) to be positive definite is that the eigenvalues

$$\begin{aligned} &1 + \{-\delta n_p\lambda_i + \delta^2 n_p^2\lambda_i^2 + 3\delta n_p^3\kappa^2 f^2(\lambda_i)\lambda_i + n_p^3\kappa^3 f^3(\lambda_i) \\ &+2\delta n_p^4\kappa^3 f^3(\lambda_i)\lambda_i + \delta^2 n_p^3[3\kappa f(\lambda_i) + 3\kappa^2 n_p f^2(\lambda_i) + \kappa^3 n_p^2 f^3(\lambda_i)]\lambda_i^2\} \end{aligned}$$

be positive for all $i = 1, \ldots, n_s$.

Example: Consider only the first four locations and predictor variables in the Example of Section 2.1.1 with 2 observations per site, that is n_s=4 and n_p=2. Even for this scenario, V is an 8x8 positive definite matrix, whose inverse as a function of three parameters is difficult to calculate. The two exponential covariogram models in (8) were used to generate the following V.

$$V = \begin{pmatrix} \sigma_p^2+\bullet & \bullet & \spadesuit & \spadesuit & \heartsuit & \heartsuit & \diamondsuit & \diamondsuit \\ \bullet & \sigma_p^2+\bullet & \spadesuit & \spadesuit & \heartsuit & \heartsuit & \diamondsuit & \diamondsuit \\ \spadesuit & \spadesuit & \sigma_p^2+\bullet & \bullet & \clubsuit & \clubsuit & \clubsuit & \clubsuit \\ \spadesuit & \spadesuit & \bullet & \sigma_p^2+\bullet & \clubsuit & \clubsuit & \clubsuit & \clubsuit \\ \heartsuit & \heartsuit & \clubsuit & \clubsuit & \sigma_p^2+\bullet & \bullet & \triangle & \triangle \\ \heartsuit & \heartsuit & \clubsuit & \clubsuit & \bullet & \sigma_p^2+\bullet & \triangle & \triangle \\ \diamondsuit & \diamondsuit & \clubsuit & \clubsuit & \triangle & \triangle & \sigma_p^2+\bullet & \bullet \\ \diamondsuit & \diamondsuit & \clubsuit & \clubsuit & \triangle & \triangle & \bullet & \sigma_p^2+\bullet \end{pmatrix}$$

with

$$\begin{aligned} \star &= \sigma_s^2(.8\alpha + .2) \\ \spadesuit &= \sigma_s^2(-.11166\alpha + .1633) \\ \heartsuit &= \sigma_s^2(-.1373\alpha + .1447) \\ \diamondsuit &= \sigma_s^2(.1029\alpha - .1029) \\ \clubsuit &= \sigma_s^2(-.1295\alpha + .1581) \\ \triangle &= \sigma_s^2(-.0628\alpha + .1729). \end{aligned}$$

Using $\alpha = .25, \sigma_p^2 = 16,$ and $\sigma_s^2 = 10$ and the exponential models in (8), the exact and approximate inverses for V are calculated and shown in Table 9. The matrix ${}_B\mathrm{V}_2^{-1}$ is centered about $B = \mathrm{I} + \delta(\mathrm{I} \otimes U)$. These matrices are all positive definite. The lower triangular entries are given by symmetry. Inspection of Table 9 reveals that ${}_B\mathrm{V}_2^{-1}$ is closer to the true inverse of V than the other approximations, even though it fails the radius of convergence test.

3.3 A Formula for V^{-1}

The asymptotic variance $\mathcal{I}^{-1}$ in Chapter 4 (see expression (32)) involves the inverse of the matrix $\mathrm{V}(\theta) = \sigma_p^2(\mathrm{I} + \delta\mathrm{A})$ with $\mathrm{A} = \Sigma \otimes \mathrm{U}$. In order to investigate this variance the Taylor series for V^{-1} will now be developed. Line (20) can be expressed as

$$\begin{aligned}
\sigma_p^2\mathrm{V}^{-1} &= \sum_{r=0}^{\infty}(-\delta A)^r \\
&= \mathrm{I} + \sum_{r=1}^{\infty} n_p^{r-1}(-\delta\Sigma)^r \otimes U \\
&= \mathrm{I} + \frac{1}{n_p}\left[\sum_{r=0}^{\infty}(-\delta n_p\Sigma)^r - \mathrm{I}_{n_s}\right] \otimes U \\
&= \mathrm{I} + \frac{1}{n_p}\left[(\mathrm{I}_{n_s} + n_p\delta\Sigma)^{-1} - \mathrm{I}_{n_s}\right] \otimes U \\
&= \mathrm{I} + P \otimes U, \qquad (30)
\end{aligned}$$

with $P = \frac{1}{n_p}\left[(\mathrm{I}_{n_s} + n_p\delta\Sigma)^{-1} - \mathrm{I}_{n_s}\right]$. The $n_s \times n_s$ matrix P is symmetric and dependent on both n_p and n_s. Certainly, the above argument fails for large n_p since $\delta n_p \rho(\Sigma) > 1$, violating (24). The Taylor approximations are limited in that the radius of convergence shrinks to zero. However, the form suggested by (30) makes sense for every n_s and n_p; and Lemma 3.2 verifies that (30) is indeed the inverse of V. Expression (30) will be used for $\sigma_p^2\mathrm{V}^{-1}$ in $\mathcal{I}^{-1}$.

Lemma 3.2 *Given* $\mathrm{V} = \sigma_p^2(\mathrm{I} + \delta(\Sigma \otimes U))$, *let* $\sigma_p^2\mathrm{V}^{-1} = \mathrm{I} + (P \otimes U)$ *where* $P = \frac{1}{n_p}\left[(\mathrm{I}_{n_s} + n_p\delta\Sigma)^{-1} - \mathrm{I}_{n_s}\right]$. *Then* $\mathrm{V}\mathrm{V}^{-1} = \mathrm{I} = \mathrm{V}^{-1}\mathrm{V}$.

$$V^{-1} = \begin{pmatrix} .053 & -.010 & -.002 & -.002 & -.002 & -.002 & -.001 & -.001 \\ \bullet & .053 & -.002 & -.002 & -.002 & -.002 & -.001 & -.001 \\ \bullet & \bullet & .053 & -.010 & -.002 & -.002 & -.002 & -.002 \\ \bullet & \bullet & \bullet & .053 & -.002 & -.002 & -.002 & -.002 \\ \bullet & \bullet & \bullet & \bullet & .053 & -.010 & -.002 & -.002 \\ \bullet & \bullet & \bullet & \bullet & \bullet & .053 & -.002 & -.002 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .053 & -.010 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .053 \end{pmatrix}$$

$$V_1^{-1} = \begin{pmatrix} .047 & -.016 & -.005 & -.005 & -.004 & -.004 & -.003 & -.003 \\ \bullet & .047 & -.005 & -.005 & -.004 & -.004 & -.003 & -.003 \\ \bullet & \bullet & .047 & -.016 & -.005 & -.005 & -.005 & -.005 \\ \bullet & \bullet & \bullet & .047 & -.005 & -.005 & -.005 & -.005 \\ \bullet & \bullet & \bullet & \bullet & .047 & -.016 & -.006 & -.006 \\ \bullet & \bullet & \bullet & \bullet & \bullet & .047 & -.006 & -.006 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .047 & -.016 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .047 \end{pmatrix}$$

$$V_2^{-1} = \begin{pmatrix} .057 & -.005 & .001 & .001 & .001 & .001 & .002 & .002 \\ \bullet & .057 & .001 & .001 & .001 & .001 & .002 & .002 \\ \bullet & \bullet & .057 & -.005 & .002 & .002 & .001 & .001 \\ \bullet & \bullet & \bullet & .057 & .002 & .002 & .001 & .001 \\ \bullet & \bullet & \bullet & \bullet & .057 & -.005 & .001 & .001 \\ \bullet & \bullet & \bullet & \bullet & \bullet & .057 & .001 & .001 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .057 & -.005 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .057 \end{pmatrix}$$

$$_BV_2^{-1} = \begin{pmatrix} .052 & -.011 & -.002 & -.002 & -.001 & -.001 & -.001 & -.001 \\ \bullet & .052 & -.002 & -.002 & -.001 & -.001 & -.001 & -.001 \\ \bullet & \bullet & .057 & -.005 & -.001 & -.001 & -.001 & -.001 \\ \bullet & \bullet & \bullet & .052 & -.001 & -.001 & -.001 & -.001 \\ \bullet & \bullet & \bullet & \bullet & .052 & -.011 & -.002 & -.002 \\ \bullet & \bullet & \bullet & \bullet & \bullet & .052 & -.002 & -.002 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .052 & -.011 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & .052 \end{pmatrix}$$

Table 9: V^{-1} and Various Approximations

Proof:

$$\begin{aligned} \mathrm{V}\mathrm{V}^{-1} &= (\mathrm{I} + \delta(\Sigma \otimes U))(\mathrm{I} + (P \otimes U)) \\ &= \mathrm{I} + P \otimes U + \delta(\Sigma \otimes U) + \delta(\Sigma \otimes U)(P \otimes U) \\ &= \mathrm{I} + P \otimes U + \delta\Sigma \otimes U + \delta n_p \Sigma P \otimes U \\ &= \mathrm{I} + (P + \delta\Sigma + \delta n_p \Sigma P) \otimes U \end{aligned}$$

Now using the expression for P, note that

$$\begin{aligned} P + \delta\Sigma + \delta n_p \Sigma P &= \frac{1}{n_p}(\mathrm{I} + \delta n_p \Sigma)^{-1} - \frac{1}{n_p}\mathrm{I} + \delta\Sigma \\ &\quad + \delta n_p \Sigma \left(\frac{1}{n_p}(\mathrm{I} + \delta n_p \Sigma)^{-1} - \frac{1}{n_p}\mathrm{I} \right) \\ &= \frac{1}{n_p}(\mathrm{I} + \delta n_p \Sigma)(\mathrm{I} + \delta n_p \Sigma)^{-1} - \frac{1}{n_p}\mathrm{I} \\ &= 0. \end{aligned}$$

Therefore, $\mathrm{V}\mathrm{V}^{-1} = I$. Similarly, for $\mathrm{V}^{-1}\mathrm{V}$. ■

This chapter has established a compact formula for V^{-1}. The next chapter will attempt to exploit this formula in the analysis of the likelihood function. To our knowledge, this formula is new. It can be applied to situations other than those of this thesis. For example, the Σ need not be of the form $\alpha \mathrm{C}_1 + (1 - \alpha)\mathrm{C}_2$. This formula can also be used to accelerate numeric computations in the case of balanced data. The inverse of V now involves finding the inverse of an $n_s \times n_s$ matrix rather than an $n_s n_p \times n_s n_p$ matrix. There is no immediate gain to the work of Mardia and Marshall (1984) who consider only one observation per site. This formula could be applied, however, when extending their model in the case of multiple responses.

Chapter 4

Asymptotics

The aim of this chapter is to find the asymptotic covariance matrix of the maximum likelihood estimators of the parameters and to determine how it is affected by the number of samples per site, the number of sites and their spacing, the covariograms, and the predictor variables. The formula for V^{-1} of Chapter 3 will be used here. A brief discussion on determining sample sizes based on this chapter is found in Section 5.2

4.1 The Asymptotic Covariance Matrix

The asymptotic covariance matrix of the estimators can be computed using derivatives of the likelihood with respect to the parameters. Maintaining a similar notation to that of Mardia and Marshall (1984), the partial second derivatives of the likelihood are given by

$$\begin{aligned}
L_{bb} := \frac{\partial^2 L}{\partial b^2} &:= -X' V^{-1} X \\
L_{b\theta}(l^{th} \text{ column}) &:= X' V^{-1} \frac{\partial V}{\partial \theta_l} V^{-1} Xb - X' V^{-1} \frac{\partial V}{\partial \theta_l} V^{-1} y \\
L_{\theta\theta}((ml)^{th} \text{ column}) &:= \frac{1}{2} tr \left[-V^{-1} \frac{\partial V}{\partial \theta_m} V^{-1} \frac{\partial V}{\partial \theta_l} + V^{-1} \frac{\partial^2 V}{\partial \theta_m \partial \theta_l} \right] - \frac{1}{2} W' V^{ml} W \\
\text{for } l, m = 1, 2, 3.
\end{aligned}$$

where $\mathrm{V}^{ml} = \mathrm{V}^{-1}\left[\frac{\partial V}{\partial\theta_m}\mathrm{V}^{-1}\frac{\partial V}{\partial\theta_l} + \frac{\partial V}{\partial\theta_l}\mathrm{V}^{-1}\frac{\partial V}{\partial\theta_m} - \frac{\partial^2 \mathrm{V}}{\partial\theta_m\partial\theta_l}\right]\mathrm{V}^{-1}$ and $\mathrm{W} = \mathrm{y} - \mathrm{Xb}$. Denote the matrix of second deriviatives of the likelihood function by

$$\mathrm{L}^{(2)} = \begin{pmatrix} \mathrm{L_{bb}} & \mathrm{L_{b\theta}} \\ \mathrm{L_{b\theta'}} & \mathrm{L_{\theta\theta}} \end{pmatrix}. \tag{31}$$

Now define the *information matrix* $\mathcal{I}$ to be the negative of the expected value of (31), that is $\mathcal{I} = -E(L^{(2)})$ (see Appendix B.2). When taking the expectation, notice that

$$\begin{aligned} E(L_{b\theta}) &= E(\mathrm{X}'\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_l}\mathrm{V}^{-1}\mathrm{Xb} - \mathrm{X}'\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_l}\mathrm{V}^{-1}\mathrm{y}) \\ &= \mathrm{X}'\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_l}\mathrm{V}^{-1}\mathrm{Xb} - \mathrm{X}'\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_l}\mathrm{V}^{-1}\mathrm{Xb} \\ &= 0. \end{aligned}$$

Furthermore, $E(\mathrm{W}'\mathrm{V}^{ml}\mathrm{W}) = tr(\mathrm{V}^{ml}\mathrm{V})$, so that

$$E(L_{\theta\theta}) = -\frac{1}{2}tr(\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_m}\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_l}).$$

Under certain conditions, the *asymptotic covariance matrix* is given by the inverse of the information matrix as described by Thisted (1988) on page 159. Hence, the asymptotic covariance matrix of the maximum likelihood estimates of the parameter Φ is given by

$$\mathcal{I}^{-1} = \begin{pmatrix} (\mathrm{X}'\mathrm{V}^{-1}\mathrm{X})^{-1}{}_{2\times 2} & 0 \\ 0 & (\frac{1}{2}t_{ml})^{-1}{}_{3\times 3} \end{pmatrix} \tag{32}$$

with $t_{ml} = tr(\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_m}\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial\theta_l})$. By the properties of traces, note that $t_{ml} = t_{lm}$. Furthermore, under reasonable smoothness conditions, the maximum likelihood estimators are asymptotically normal. The value $[-L^{(2)}(\hat{\Phi})]^{-1}$ is known as the *observed* Fisher information while $(\mathcal{I}(\hat{\Phi}))^{-1}$ is the *estimated* Fisher information. Note that although the estimate of $\mathrm{b} = (\beta_0, \beta_1)$ is asymptotically independent of that of $\theta = (\sigma_p^2, \sigma_s^2, \alpha)$ the variances both involve V^{-1}.

4.1.1 Numeric Estimation of the Asymptotic Variance

Estimates of the asymptotic variance can be based on either the estimated Fisher information, $[\mathcal{I}(\hat{\Phi})]^{-1}$, or the observed Fisher information, $[-L^{(2)}(\hat{\Phi})]^{-1}$. The standard

errors are the square roots of the diagonal elements in either case. While numeric estimation is always possible, it offers no insight to experimental design. Some of these issues are discussed in Chapter 5.

4.1.2 Expanded Expressions of the Asymptotic Variance

In (32), consider first the upper left 2×2 submatrix $(X'V^{-1}X)^{-1}$, whose diagonal represents the asymptotic behavior of the variance of $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively. Using (30), it becomes

$$\frac{\sigma_p^2}{F}\begin{pmatrix} n_p\sum x_i{}^2 + n_p^2\sum\sum x_i x_{i'} p_{ii'} & -n_p\sum x_i - n_p^2\sum\sum x_i p_{ii'} \\ -n_p\sum x_i - n_p^2\sum\sum x_i p_{ii'} & n_s n_p + n_p^2\sum\sum p_{ii'} \end{pmatrix} \tag{33}$$

with

$$\begin{aligned} F &= (n_s n_p + n_p^2\sum\sum p_{ii'})(n_p\sum x_i{}^2 + n_p^2\sum\sum x_i x_{i'} p_{ii'}) \\ &\quad -(n_p\sum x_i + n_p^2\sum\sum x_i p_{ii'})^2. \end{aligned}$$

The $p_{ii'}$'s are the elements of the matrix

$$P = \frac{1}{n_p}[(I + n_p\delta\Sigma)^{-1} - I], \tag{34}$$

where $\delta = \sigma_s^2/\sigma_p^2$.

Next consider the lower right 3×3 submatrix $2(t_{ml})^{-1}$, whose diagonal represents the asymptotic variances of $\widehat{\sigma_p^2}, \widehat{\sigma_s^2}$ and $\widehat{\alpha}$ respectively. The inverse of this symmetric matrix is given by

$$2\begin{pmatrix} \frac{t_{22}t_{33}-t_{23}^2}{T} & \frac{t_{12}t_{33}-t_{13}t_{23}}{-T} & \frac{t_{12}t_{23}-t_{13}t_{22}}{T} \\ \bullet & \frac{t_{11}t_{33}-t_{13}^2}{T} & \frac{t_{11}t_{23}-t_{12}t_{13}}{-T} \\ \bullet & \bullet & \frac{t_{11}t_{22}-t_{12}^2}{T} \end{pmatrix},$$

with $T = t_{11}t_{22}t_{33} - t_{11}t_{23}^2 - t_{12}^2t_{33} + 2t_{12}t_{13}t_{23} - t_{13}^2t_{22}$. Furthermore, each t_{ml} can be expressed in terms of P, given in (34), as follows:

$$t_{11} = \sigma_p^{-4}\{n_s n_p + 2n_p tr(P) + n_p^2 tr(P^2)\}$$

$$
\begin{aligned}
t_{12} &= \sigma_p^{-4}\{n_p tr(\Sigma) + 2n_p^2 tr(P\Sigma) + n_p^3 tr(P^2\Sigma)\} \\
t_{13} &= \sigma_s^2\sigma_p^{-4}\{n_p tr(\frac{\partial\Sigma}{\partial\alpha}) + 2n_p^2 tr(P\frac{\partial\Sigma}{\partial\alpha}) + n_p^3 tr(P^2\frac{\partial\Sigma}{\partial\alpha})\} \\
t_{22} &= \sigma_p^{-4}\{n_p^2 tr(\Sigma^2) + 2n_p^3 tr(\Sigma^2 P) + n_p^4 tr(P\Sigma P\Sigma)\} \\
t_{23} &= \sigma_s^2\sigma_p^{-4}\{n_p^2 tr(\Sigma\frac{\partial\Sigma}{\partial\alpha}) + n_p^3 tr(\Sigma P\frac{\partial\Sigma}{\partial\alpha}) + n_p^3 tr(P\Sigma\frac{\partial\Sigma}{\partial\alpha}) + n_p^4 tr(P\Sigma P\frac{\partial\Sigma}{\partial\alpha})\} \\
t_{33} &= \sigma_s^4\sigma_p^{-4}\{n_p^2 tr((\frac{\partial\Sigma}{\partial\alpha})^2) + 2n_p^3 tr((\frac{\partial\Sigma}{\partial\alpha})^2 P) + n_p^4 tr(P\frac{\partial\Sigma}{\partial\alpha}P\frac{\partial\Sigma}{\partial\alpha})\}.
\end{aligned}
\tag{35}
$$

The above expressions are found by expanding the trace terms. For example, with $V = \sigma_p^2 \mathrm{I} + \sigma_s^2 \Sigma \otimes U$ and $\sigma_p^2 \mathrm{V}^{-1} = \mathrm{I} + P \otimes U$, t_{11} is calculated using the properties of Kronecker products as follows.

$$
\begin{aligned}
tr(\mathrm{V}^{-1}\frac{\partial V}{\partial\sigma_p^2}\mathrm{V}^{-1}\frac{\partial V}{\partial\sigma_p^2}) &= \sigma_p^{-4} tr\left((\mathrm{I} + P \otimes U)(\mathrm{I} + P \otimes U)\right) \\
&= \sigma_p^{-4} tr(\mathrm{I} + 2(P \otimes U) + n_p(P^2 \otimes U)) \\
&= \sigma_p^{-4}\{n_s n_p + 2n_p tr(P) + n_p^2 tr(P^2).
\end{aligned}
$$

The expressions in (35) can be further expanded on substituting $\Sigma = \alpha \mathrm{C}_1 + (1-\alpha)\mathrm{C}_2$ and $\frac{\partial\Sigma}{\partial\alpha} = \mathrm{C}_1 - \mathrm{C}_2$, but this is not pursued here.

4.2 Letting n_p go to Infinity

The asymptotic behaviour of $\mathcal{I}^{-1}$ as n_p goes to infinity will be determined by the leading power of n_p in $\mathcal{I}^{-1}$. In an attempt to retain only the leading coefficients of the terms with the highest power of n_p, Lemma A.2 can be applied to the matrix P, so that the matrix P is asymptotically like $(-1/n_p)\mathrm{I}$, with each $p_{ii'} = -1/n_p$ for $i \neq i'$ and 0 otherwise. If this approximation were used, the asymptotic variances of $\hat{\beta}_o$ and $\hat{\beta}_1$ would become zero. Similarly, all the t_{ml}'s except t_{11} would also become zero. To obtain the constants will require a better asymptotic representation of P than $(-1/n_p)\mathrm{I}$. In fact, to find the limits of the asymptotic variances, all the terms involving n_p must be retained until the very end. This is done using vector notation and expanding expressions. The next two lemmas establish limits of the asymptotic variances for all the estimated parameters.

Expression (33) can be rewritten using vector notation. Let $\mathrm{x}' = (x_1, x_2, \ldots, x_{n_s})$ and $1'_{n_s} = (1, 1, \ldots, 1)$ be an $1 \times n_s$ vector of ones. After some simplification, and since P is symmetric, (33) becomes

$$\frac{\sigma_p^2}{F}\begin{pmatrix} n_p\mathrm{x}'(\mathrm{I}+n_p\delta\Sigma)^{-1}\mathrm{x} & -n_p 1'_{n_s}(\mathrm{I}+n_p\delta\Sigma)^{-1}\mathrm{x} \\ -n_p 1'_{n_s}(\mathrm{I}+n_p\delta\Sigma)^{-1}\mathrm{x} & n_p 1'_{n_s}(\mathrm{I}+n_p\delta\Sigma)^{-1}1_{n_s} \end{pmatrix}, \tag{36}$$

with

$$F = n_p^2 1'_{n_s}(\mathrm{I}+n_p\delta\Sigma)^{-1}[1_{n_s}\mathrm{x}' - \mathrm{x}1'_{n_s}](\mathrm{I}+n_p\delta\Sigma)^{-1}\mathrm{x}.$$

Lemma 4.1 *Assume that* $1_{n_s}\mathrm{x}' - \mathrm{x}1'_{n_s} \neq 0_{n_s}$. *Then as* $n_p \to \infty$,

$$\begin{aligned} \mathrm{Var}\hat{\beta}_0 &\to \frac{\sigma_s^2\ \mathrm{x}'\Sigma^{-1}\mathrm{x}}{1'_{n_s}\Sigma^{-1}[1_{n_s}\mathrm{x}' - \mathrm{x}1'_{n_s}]\Sigma^{-1}\mathrm{x}} \\ \mathrm{Var}\hat{\beta}_1 &\to \frac{\sigma_s^2\ 1'_{n_s}\Sigma^{-1}1_{n_s}}{1'_{n_s}\Sigma^{-1}[1_{n_s}\mathrm{x}' - \mathrm{x}1'_{n_s}]\Sigma^{-1}\mathrm{x}}. \end{aligned}$$

Proof: The condition $1_{n_s}\mathrm{x}' - \mathrm{x}1'_{n_s} \neq 0_{n_s}$ insures that not all $x_i = x_{i'}$. Using Lemma A.2, the entries in (36) are asymptotically equivalent to

$$\begin{aligned} (1,1)\text{-entry} &\sim \mathrm{x}'\Sigma^{-1}\mathrm{x}/\delta &&= O(1), \\ (2,2)\text{-entry} &\sim 1'_{n_s}\Sigma^{-1}1_{n_s}/\delta &&= O(1), \\ F &\sim 1'_{n_s}\Sigma^{-1}[1_{n_s}\mathrm{x}' - \mathrm{x}1'_{n_s}]\Sigma^{-1}\mathrm{x}/\delta^2 &&= O(1). \end{aligned} \tag{37}$$

■

Lemma 4.2 *As* n_p *goes to infinity, the asymptotic variances (in* $\mathcal{I}^{-1}$*) of* $\widehat{\sigma_p^2}$, $\widehat{\sigma_s^2}$, *and* $\hat{\alpha}$ *are given by*

$$\begin{aligned} \mathrm{Var}\widehat{\sigma_p^2} &\to 0 \\ \mathrm{Var}\widehat{\sigma_s^2} &\to 2\Delta_1\sigma_p^4\delta^2/(n_s\Delta_1 - \Delta_2^2) \\ \mathrm{Var}\hat{\alpha} &\to 2n_s\sigma_p^4\delta^2/(\sigma_s^4(n_s\Delta_1 - \Delta_2^2)) \end{aligned}$$

where $\Delta_1 = tr(\Sigma^{-1}\frac{\partial\Sigma}{\partial\alpha}\Sigma^{-1}\frac{\partial\Sigma}{\partial\alpha})$ *and* $\Delta_2 = tr(\Sigma^{-1}\frac{\partial\Sigma}{\partial\alpha})$.

Proof: It is necessary to substitute $P = \frac{1}{n_p}[(\mathrm{I}_{n_s} + n_p\delta\Sigma)^{-1} - \mathrm{I}_{n_s}]$ into the t_{ml} expressions in (35). First all the trace terms involving P are expanded using the properties

of Kronecker products and traces and then Lemma A.2 is applied. Results of these calculations are shown in Table 10. The Maple program *asyvar3.theory*, is used at this point to facilitate the calculation of the asymptotic variances, since retaining all terms involving n_p leads to rather large expressions. The program collects the terms in powers of n_p demonstrating that

$$\begin{aligned} \mathrm{Var}\widehat{\sigma_p^2} &\sim 2\sigma_p^4/n_p n_s \to 0 \\ \mathrm{Var}\widehat{\sigma_s^2} &\to 2\delta^2\sigma_p^4\Lambda_{11}/(n_s\Lambda_{11} - \Lambda_4^2) \\ \mathrm{Var}\hat{\alpha} &\to 2n_s^2\delta^2\sigma_p^4/(\sigma_s^4(n_s\Lambda_{11} - \Lambda_4^2)), \end{aligned}$$

with Λ_{11} and Λ_4 defined as in Table 10. Letting $\Delta_1 = \Lambda_{11}$ and $\Delta_2 = \Lambda_4$ gives the desired result. Here, only the $\mathrm{Var}(\widehat{\sigma_p^2})$ goes to zero. The other two terms are of order one. ■

So in the case of n_p going to infinity, we have a limiting regression experiment with the spatial errors remaining. Regardless of the number of additional data points taken at each site, the spatial error at that point remains, and the estimates of β_0 and β_1 cannot be improved.

4.2.1 Numerical Results for Large n_p

Using expression (30) for the inverse of V and several Maple programs, *asyvar2, asyvar3, short2 and short3*, we calculate the asymptotic variance for increasingly large n_p using Example 2.1.1 with 10 sites. Results for increasing n_p are shown in Table 12. Both the programs *asyvar2* and *short2* calculate the asymptotic variance of $\hat{\beta}_0$ and $\hat{\beta}_1$ but *short2* is much quicker. The first program evaluates the variances by taking the inverse of a product of matrices $\mathrm{X}'\mathrm{V}^{-1}\mathrm{X}$ directly while the second program evaluates the derived expressions in (33). Similarly, *asyvar3* evaluates the variances of $\widehat{\sigma_p^2}$, $\widehat{\sigma_s^2}$, and $\hat{\alpha}$ by computing the trace terms, t_{ml} by their definition. The program *short3* uses the trace expansions in terms of P given in (35) and takes significantly less time. Both sets of programs give almost identical results but in very different amounts of time as is shown in Table 11. Using the quicker programs, the approximate asymptotic variance for all five estimated parameters is calculated for increasingly large n_p as

$$
\begin{aligned}
tr(P) &\sim \Lambda_1/(\delta n_p^2) - n_s/n_p \\
tr(P^2) &\sim n_s/n_p^2 + 2\Lambda_1/(\delta n_p^3) + \Lambda_2/(\delta^2 n_p^4) \\
tr(P\Sigma) &\sim n_s/(\delta n_p^2) - \Lambda_3/n_p \\
tr(P^2\Sigma) &\sim \Lambda_1/(\delta^2 n_p^4) - 2n_s/(\delta n_p^3) + \Lambda_3/n_p^2 \\
tr\left(P\frac{\partial\Sigma}{\partial\alpha}\right) &\sim \Lambda_4/(\delta n_p^2) - \Lambda_5/n_p \\
tr\left(P^2\frac{\partial\Sigma}{\partial\alpha}\right) &\sim \Lambda_6/(\delta^2 n_p^4) - 2\Lambda_4/(\delta n_p^3) + \Lambda_5/n_p \\
tr(\Sigma^2 P) &\sim \Lambda_3/(\delta n_p^2) - \Lambda_7/n_p \\
tr(P\Sigma P\Sigma) &\sim n_s/(\delta^2 n_p^4) - 2\Lambda_3/(\delta n_p^3) + \Lambda_7/n_p^2 \\
tr\left(\Sigma P\frac{\partial\Sigma}{\partial\alpha}\right) &\sim \Lambda_5/(\delta n_p^2) - \Lambda_8/n_p \\
tr\left(P\Sigma P\frac{\partial\Sigma}{\partial\alpha}\right) &\sim \Lambda_4/(\delta^2 n_p^4) - 2\Lambda_5/(\delta n_p^3) + \Lambda_8/n_p \\
tr\left(\left(\frac{\partial\Sigma}{\partial\alpha}\right)^2 P\right) &\sim \Lambda_9/(\delta n_p^2) - \Lambda_{10}/n_p \\
tr\left(P\frac{\partial\Sigma}{\partial\alpha}P\frac{\partial\Sigma}{\partial\alpha}\right) &\sim \Lambda_{11}/(\delta^2 n_p^4) - 2\Lambda_9/(\delta n_p^3) + \Lambda_{10}/n_p^2
\end{aligned}
$$

where

$$
\begin{array}{lll}
\Lambda_1 = tr(\Sigma^{-1}) & \Lambda_2 = tr(\Sigma^{-2}) & \Lambda_3 = tr(\Sigma) \\
\Lambda_4 = tr\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\alpha}\right) & \Lambda_5 = tr\left(\frac{\partial\Sigma}{\partial\alpha}\right) & \Lambda_6 = tr\left(\Sigma^{-2}\frac{\partial\Sigma}{\partial\alpha}\right) \\
\Lambda_7 = tr(\Sigma^2) & \Lambda_8 = tr\left(\Sigma\frac{\partial\Sigma}{\partial\alpha}\right) & \Lambda_9 = tr\left(\Sigma^{-1}\left(\frac{\partial\Sigma}{\partial\alpha}\right)^2\right) \\
\Lambda_{10} = tr\left(\left(\frac{\partial\Sigma}{\partial\alpha}\right)^2\right) & \Lambda_{11} = tr\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\alpha}\Sigma^{-1}\frac{\partial\Sigma}{\partial\alpha}\right) &
\end{array}
$$

Table 10: Trace Expansions

Table 11: Time Requirements in seconds of CPU unless otherwise specified.

n_p	1	5	10	25
asyvar2	5.09	7.74	19.33	161.2
short2	3.02	3.28	3.26	3.06
asyvar3	10.77	323.71	2628.31	> 1 day
short3	7.620	7.670	7.650	< 1 minute

Table 12: Approximate Asymptotic Variance. Theoretical limits were used in the case of $n_p = \infty$.

n_p	$\text{Var}\hat{\beta}_0$	$\text{Var}\hat{\beta}_1$	$\text{Var}\widehat{\sigma_p^2}$	$\text{Var}\widehat{\sigma_s^2}$	$\text{Var}\hat{\alpha}$
1	13.64	.000121	357.69	7419.44	.592
5	10.24	.000056	12.67	5095.93	.204
10	9.73	.000046	5.68	4775.89	.179
15	9.54	.000043	3.65	4663.56	.170
20	9.44	.000041	2.69	4605.75	.166
25	9.39	.000040	2.13	4570.54	.163
50	9.26	.000037	1.04	4498.61	.158
1000	9.14	.000035	.051	4427.68	.152
5000	9.14	.000035	.010	4424.63	.152
∞	9.14	.000035	0	4423.87	.152

is shown in Table 12. The limits derived in the previous two lemmas are used to calculate the asymptotic variance in the case $n_p = \infty$. The variances for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\alpha}$ appear to converge quicker than those of $\widehat{\sigma_p^2}$ and $\widehat{\sigma_s^2}$.

4.3 Eigenvalue Representation

Consider the spectral decompostion of the symmetric matrix V^{-1} using (30). Then

$$V^{-1} = \sum_{h=1}^{n_s n_p} \lambda_h(V^{-1}) e_h e'_h$$

where λ_h's are the eigenvalues and e_h's are the normalized eigenvectors of V^{-1}. First, expressions in terms of eigenvalues can be found for each entry of the 2×2 matrix $X'V^{-1}X$. Then the asymptotic variance of β_0 and β_1 is given by the diagonal entries of the inverse of this matrix. The (1,1) entry of $X'V^{-1}X$, denoted $X'V^{-1}X[1,1]$, is found by multiplying the first row of X' with V^{-1} and the first column of X. Denote $1_{n_s n_p}$ to be a $n_s n_p \times 1$ vector of ones so that

$$\begin{aligned} X'V^{-1}X[1,1] &= 1'_{n_s n_p} V^{-1} 1_{n_s n_p} \qquad (38) \\ &= 1'_{n_s n_p} \sum_h \lambda_h(V^{-1}) e_h e'_h 1_{n_s n_p} \\ &= \sum_h \lambda_h(V^{-1}) 1'_{n_s n_p} e_h e'_h 1_{n_s n_p} \qquad (39) \end{aligned}$$

Now the properties of Kronecker products are used to find the eigenvalues and eigenvectors of $\sigma_p^2 V^{-1} = I + P \otimes U$. The eigenvalues of U are n_p and 0 with multiplicity $n_p - 1$ with corresponding $n_p \times 1$ eigenvectors, u, given by

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} -1 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

The n_s eigenvalues of P are given by $-\delta\lambda_i(\Sigma)/(1 + n_p\delta\lambda_i(\Sigma))$. Denote the n_s eigenvectors of P, and hence of Σ, by s_i. Thus the eigenvalues of $V^{-1}$are 1 with multiplicity $n_s n_p - n_s$, and $1 - n_p\delta\lambda_i(\Sigma)/(1 + n_p\delta\lambda_i(\Sigma))$ for $i = 1, \ldots, n_s$. Notice that V^{-1} is in fact positive definite as required. The eigenvectors of V^{-1} are $e_h = s_i \otimes u_j$. For $\lambda(V^{-1}) = 1$, expression (39) reduces to zero since $1'_{n_s n_p}$ times the corresponding eigenvector gives

zero. For $\lambda_i(V^{-1}) = 1 - n_p\delta\lambda_i(\Sigma)/(1 + n_p\delta\lambda_i(\Sigma))$ the corresponding eigenvector is

$$e_i = (s_{i1}, \ldots, s_{i1}, s_{i2}, \ldots, s_{i2}, \ldots, s_{in_s}, \ldots, s_{in_s})'.$$

Equation (39) then becomes

$$\begin{aligned} X'V^{-1}X[1,1] &= \sum_{i=1}^{n_s} n_p^2 \left(1 - \frac{n_p\delta\lambda_i(\Sigma)}{1 + n_p\delta\lambda_i(\Sigma)}\right) \left(\sum_{t=1}^{n_s} s_{it}\right)^2 \\ &= \sum_i n_p^2 \left(\frac{1}{1 + n_p\delta\lambda_i(\Sigma)}\right) \left(\sum_t s_{it}\right)^2. \end{aligned} \tag{40}$$

Normalization of the eigenvectors of V^{-1} forces each eigenvector of Σ to satisfy $\sum_t (s_{it})^2 = 1/n_p$.

Similarly, the same arguments can be applied to find the other entries.

$$\begin{aligned} X'V^{-1}X[2,2] &= \sum_i n_p^2 \left(\frac{1}{1 + n_p\delta\lambda_i(\Sigma)}\right) \left(\sum_t x_t s_{it}\right)^2. \\ X'V^{-1}X[1,2] &= \sum_i n_p^2 \left(\frac{1}{1 + n_p\delta\lambda_i(\Sigma)}\right) \left(\sum_t x_t s_{it}\right) \left(\sum_t s_{it}\right). \end{aligned} \tag{41}$$

These expressions can now be combined to express the asymptotic variance of $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$(X'V^{-1}X)^{-1} = \frac{1}{F} \begin{pmatrix} \sum_i \frac{1}{1+n_p\delta\lambda_i(\Sigma)} (\sum_t x_t s_{it})^2 & \sum_i \frac{-1}{1+n_p\delta\lambda_i(\Sigma)} \sum_t x_t s_{it} \sum_t s_{it} \\ \sum_i \frac{-1}{1+n_p\delta\lambda_i(\Sigma)} \sum_t x_t s_{it} \sum_t s_{it} & \sum_i \frac{1}{1+n_p\delta\lambda_i(\Sigma)} (\sum_t s_{it})^2 \end{pmatrix},$$

where

$$\begin{aligned} F &= n_p^2 \left(\sum_i \frac{1}{1 + n_p\delta\lambda_i(\Sigma)} \left(\sum_t s_{it}\right)^2\right) \left(\sum_i n_p^2 \frac{1}{1 + n_p\delta\lambda_i(\Sigma)} \left(\sum_t x_t s_{it}\right)^2\right) \\ &\quad - n_p^2 \left[\sum_i \frac{1}{1 + n_p\delta\lambda_i(\Sigma)} \left(\sum_t x_t s_{it}\right) \left(\sum_t s_{it}\right)\right]^2. \end{aligned}$$

If we know how x varies with site and how the eigenvalues and eigenvectors of Σ vary with choice of location, these variances can be minimized with respect to choice of site locations when n_s and n_p are fixed. This is illustrated in Chapter 5.

It is also possible to express the asymptotic variances of the remaining three parameter estimates in terms of eigenvalues and eigenvectors of Σ. As previously noted, since P is a function of Σ, the eigenvalues of P in terms of Σ are $-\delta\lambda_i(\Sigma)/(1+\delta n_p\lambda_i(\Sigma))$. In addition, the eigenvectors of Σ are the same as those of P. Denote $\lambda_i(\Sigma) \equiv \lambda_i$ and again let s_i be the corresponding eigenvectors of Σ. Lemma 4.3 is applied to the trace expressions in (35) to give

$$tr(\Sigma) = \sum_{i=1}^{n_s} \lambda_i \qquad tr(P) = -\delta \sum_{i=1}^{n_s} \lambda_i/(1+\delta n_p\lambda_i)$$
$$tr(\Sigma^2) = \sum_{i=1}^{n_s} \lambda_i^2 \qquad tr(P^2) = \delta^2 \sum_{i=1}^{n_s} \lambda_i^2/(1+\delta n_p\lambda_i)^2$$
$$tr(\Sigma^2 P) = -\delta \sum_{i=1}^{n_s} \lambda_i^3/(1+\delta n_p\lambda_i) \qquad tr(P^2\Sigma) = \delta^2 \sum_{i=1}^{n_s} \lambda_i^3/(1+\delta n_p\lambda_i)^2$$
$$tr(\Sigma P) = -\delta \sum_{i=1}^{n_s} \lambda_i^2/(1+\delta n_p\lambda_i) \qquad tr(P\Sigma P\Sigma) = \delta^2 \sum_{i=1}^{n_s} \lambda_i^4/(1+\delta n_p\lambda_i)^2$$

$$tr(\tfrac{\partial\Sigma}{\partial\alpha}) = \sum_{i=1}^{n_s} \langle s_i', \tfrac{\partial\Sigma}{\partial\alpha} s_i\rangle \qquad tr(P\tfrac{\partial\Sigma}{\partial\alpha}) = -\delta \sum_{i=1}^{n_s} \lambda_i/(1+\delta n_p\lambda_i)\langle s_i', \tfrac{\partial\Sigma}{\partial\alpha} s_i\rangle$$
$$tr(\Sigma\tfrac{\partial\Sigma}{\partial\alpha}) = \sum_{i=1}^{n_s} \lambda_i\langle s_i', \tfrac{\partial\Sigma}{\partial\alpha} s_i\rangle \qquad tr(P^2\tfrac{\partial\Sigma}{\partial\alpha}) = \delta^2 \sum_{i=1}^{n_s} \lambda_i^2/(1+\delta n_p\lambda_i)^2\langle s_i', \tfrac{\partial\Sigma}{\partial\alpha} s_i\rangle$$
$$tr\left(\Sigma P\tfrac{\partial\Sigma}{\partial\alpha}\right) = tr\left(P\Sigma\tfrac{\partial\Sigma}{\partial\alpha}\right) \qquad tr\left(P\Sigma\tfrac{\partial\Sigma}{\partial\alpha}\right) = -\delta \sum_{i=1}^{n_s} \lambda_i^2/(1+\delta n_p\lambda_i)\langle s_i', \tfrac{\partial\Sigma}{\partial\alpha} s_i\rangle$$
$$tr\left(\tfrac{\partial\Sigma}{\partial\alpha}\right)^2 = \sum_{i=1}^{n_s} \langle s_i', \left(\tfrac{\partial\Sigma}{\partial\alpha}\right)^2 s_i\rangle \qquad tr(\left(\tfrac{\partial\Sigma}{\partial\alpha}\right)^2 P) = -\delta \sum_{i=1}^{n_s} \lambda_i/(1+\delta n_p\lambda_i)\langle s_i', \left(\tfrac{\partial\Sigma}{\partial\alpha}\right)^2 s_i\rangle$$
$$tr\left(P\tfrac{\partial\Sigma}{\partial\alpha}P\tfrac{\partial\Sigma}{\partial\alpha}\right) = \delta^2 \sum_{i=1}^{n_s}\sum_{r=1}^{n_s} W_{ir} \qquad tr\left(P\Sigma P\tfrac{\partial\Sigma}{\partial\alpha}\right) = \delta^2 \sum_{i=1}^{n_s} \lambda_i^3/(1+\delta n_p\lambda_i)^2\langle s_i', \tfrac{\partial\Sigma}{\partial\alpha} s_i\rangle$$

where

$$W_{ir} = \frac{\lambda_i\lambda_r}{(1+\delta n_p\lambda_i)(1+\delta n_p\lambda_r)} s_i' \frac{\partial\Sigma}{\partial\alpha} s_r s_r' \frac{\partial\Sigma}{\partial\alpha} s_i.$$

Now the t_{ml}'s and hence the asymptotic variances of $\widehat{\sigma_p^2}$ $\widehat{\sigma_s^2}$, and $\hat{\alpha}$ are a function of the eigenvalues and eigenvectors of Σ. Note that $\frac{\partial\Sigma}{\partial\alpha} = C_1 - C_2$. While it is impractical to write down the expressions for the asymptotic variance in full generality, we compute t_{22} here as an example.

$$t_{22} = \sigma_p^{-4} n_p^2 \sum_{i=1}^{n_s} [\lambda_i^2 - 2\delta n_p\lambda_i^3/(1+\delta n_p\lambda_i) + \delta^2 n_p^2\lambda_i^4/(1+\delta n_p\lambda_i)^2].$$

The eigenvalue representations can offer some insight in examining the behavior as the number of sites increases in a specified fashion. Since site locations affect Σ, the above expressions can also also be used to develop experimental designs.

4.4 Letting n_s go to Infinity

The difficulties in evaluating the asymptotic expressions as n_s goes to infinity are twofold. Firstly, there is the issue of how or where to add more sites. There are two ways in which n_s, the number of sites, can go to infinity. It is possible to increase the number of sites within a fixed domain thereby increasing the density of sites. This is known as *infill asymptotics*. Alternatively, one could increase the number of sites by enlarging the sampling domain. Secondly, it is not immediate from the asymptotic expressions derived in the chapter, how the variance is affected even when an increasing sampling scheme is chosen. For these reasons, this section predominately investigates the asymptotic behavior of the parameter estimates numerically using various sampling schemes. There is, however, some theory on increasing domain asymptotics for spatial regression. This will precede and guide our numerical experiments.

Mardia and Marshall (1984) establish sufficient conditions for the asymptotic normality and weak consistency of $\hat{\Phi}$ in the case of one observation per site with n_s sites by adapting Sweeting's (1980) general result. Denote the information matrix $\mathcal{I} = diag(M_\beta, M_\theta)$ so that $M_\beta = \mathrm{X}'\mathrm{V}^{-1}\mathrm{X}$ and the (m,l) element of M_θ is $t_{ml}/2$. Also let $X[,q]$ denote the qth column of X; and let M[p,m] denote the (p,m)th element of the matrix M. The theorem adapted to our model and notation becomes

Theorem 4.1 *If, as* $n_s \to \infty$,

$$\begin{aligned}
&lim M_\beta^{-1} = 0, \quad lim M_\theta^{-1} = 0, \\
&lim \textstyle\sum_{m,l,p,q=1}^{3} M_\theta^{-1}[p,m] M_\theta^{-1}[q,l] tr(\mathrm{V}\mathrm{V}^{pl}\mathrm{V}\mathrm{V}^{qm}) = 0 \\
&lim \textstyle\sum_{m,p=1}^{3} \sum_{l,q=1}^{2} M_\theta^{-1}[p,m] M_\beta^{-1}[q,l] (X'[,l] \frac{\partial \mathrm{V}^{-1}}{\partial \theta_p} \mathrm{V} \frac{\partial \mathrm{V}^{-1}}{\partial \theta_m} X[,q]) = 0,
\end{aligned}$$

then $\hat{\Phi} \sim \mathcal{N}(\Phi, \mathcal{I}^{-1})$ *where*

$$\mathrm{V}^{pl} = \mathrm{V}^{-1}\left(\frac{\partial \mathrm{V}}{\partial \theta_p}\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial \theta_l} + \frac{\partial \mathrm{V}}{\partial \theta_l}\mathrm{V}^{-1}\frac{\partial \mathrm{V}}{\partial \theta_p} - \frac{\partial^2 \mathrm{V}}{\partial \theta_p \partial \theta_l}\right)\mathrm{V}^{-1}.$$

They go on to find simpler sufficient conditions involving the eigenvalues of V. Let $||.||$ denote the Euclidean matrix norm.

Theorem 4.2 *Suppose that as* $n_s \to \infty$

(*i*) $\lim\max_h \lambda_h(V) = K < \infty$, $\lim\max_h |\lambda_h(\frac{\partial V}{\partial \theta_l})| = K_l < \infty$, $\lim\max_h |\lambda_h(\frac{\partial^2 V}{\partial \theta_l \theta_m})| = K_{lm} < \infty$ *for all* $l, m = 1, 2, 3$

(*ii*) $||\frac{\partial V}{\partial \theta_l}||^{-2} = O(n_s^{-\frac{1}{2}-\epsilon})$, *for some* $\epsilon > 0$ *for* $l = 1, 2, 3$.

(*iii*) $\forall l, m = 1, 2, 3, w_{lm} = \lim\{t_{ml}/(t_{mm}t_{ll})^{1/2}\}$ *exists and* $\mho = (w_{lm})$ *is a non-singular matrix*

(*iv*) $\lim(X'X)^{-1} = 0$,

Then the conditions of Theorem 4.1 hold.

Recall that in our model $V = \sigma_p^2 I + \sigma_s^2(\Sigma \otimes U)$. Using the fact that the Euclidean matrix norm is equal to the square root of the trace norm (see Hoffman (1975) p18-19), conditions (*i*) and (*ii*) reduce to

(*i'*) $\lim\max \lambda_i(\Sigma) < \infty$, $\lim\max |\lambda_i(\frac{\partial \Sigma}{\partial \alpha})| < \infty$, $\lim\max |\lambda_i(\frac{\partial^2 \Sigma}{\partial \alpha^2})| < \infty$

(*ii'*) $1/tr(\Sigma^2) = O(n_s^{-\frac{1}{2}-\epsilon})$, and $1/tr(\frac{\partial \Sigma}{\partial \alpha})^2 = O(n_s^{-\frac{1}{2}-\epsilon})$ for some $\epsilon > 0$.

Mardia and Marshall (1984) establish that an intrinsically stationary process in $\mathbb{R}^d$ sampled from a regular d-dimensional lattice with fixed spacings and increasing domain, satisfies the conditions of Theorem 4.2. Therefore we consider this design in our numerical analysis.

4.4.1 Results Concerning n_s for Specific Designs

We will consider three sampling schemes. Example A and C deal with increasing domain asymptotics while Example B explores infill asymptotics. *S-Plus* programs generate the results.

Example A: Increasing Domain Sampling Scheme

Consider locations on a unit spaced square lattice originating at the origin. A $n \times n$ lattice has n^2 sites with locations coordinates (x, y) $\forall x, y = 0, \ldots, n-1$. Furthermore, regressors are chosen deterministically in such a way that (*iv*) of Theorem 4.2

holds. Consider linearly increasing the covariates as the y-coordinate increases. That is, covariates for locations on the same horizontal level in the plane have the same value, and the value increases linearly the higher the level. This approach satisfies $(X'X)^{-1} \to 0$ empirically for the finite series $n = 2^2, 3^2, \ldots, 12^2$. Therefore, the asymptotic variance of the maximum likelihood estimates for this sampling scheme, is normal with variance $\mathcal{I}^{-1}$ according to Mardia and Marshall (1984). Of special interest is the asymptotic variance of $\widehat{\sigma_p^2}$ and $\widehat{\sigma_s^2}$. Our estimation procedure for all 5 parameters as described in Section 2.2.3 is run 100 times for a 4×4 size grid for a total of 16 sites with 2 observations per site. Two simulations did not converge after 75 iterations. The distributions of estimates for Example A with asymptotic normal density curves are shown in Figure 5. Only the simulations containing nonnegative estimates of σ_s^2 are used in these plots. From 98 simulations, the algorithm yielded 43 feasible estimates.

Example B: Infill Sampling Scheme

The domain for this example is the unit square with the left-lower corner at the origin. Here the sampling domain is fixed and sites are added systematically inside the unit square. The unit square is equally partitioned for each n. The regressors are determined as before so that $(X'X)^{-1} \to 0$ empirically for the finite series $n = 2^2, 3^2, \ldots, 12^2$. Once again, our estimation procedure for all the parameters is run 100 times for a 4×4 size grid with 16 sites and 2 observations per site. One simulation failed to converge. The distributions of estimates for Example B with asymptotic normal density curves are shown in Figure 6. Although the results of Mardia and Marshall (1984) do not apply in the case of infill asymptotics, Figure 6 might suggest some normal distributions. Only the simulations containing nonnegative estimates of σ_s^2 are used in these plots. From 99 simulations, the algorithm yielded 45 feasible estimates.

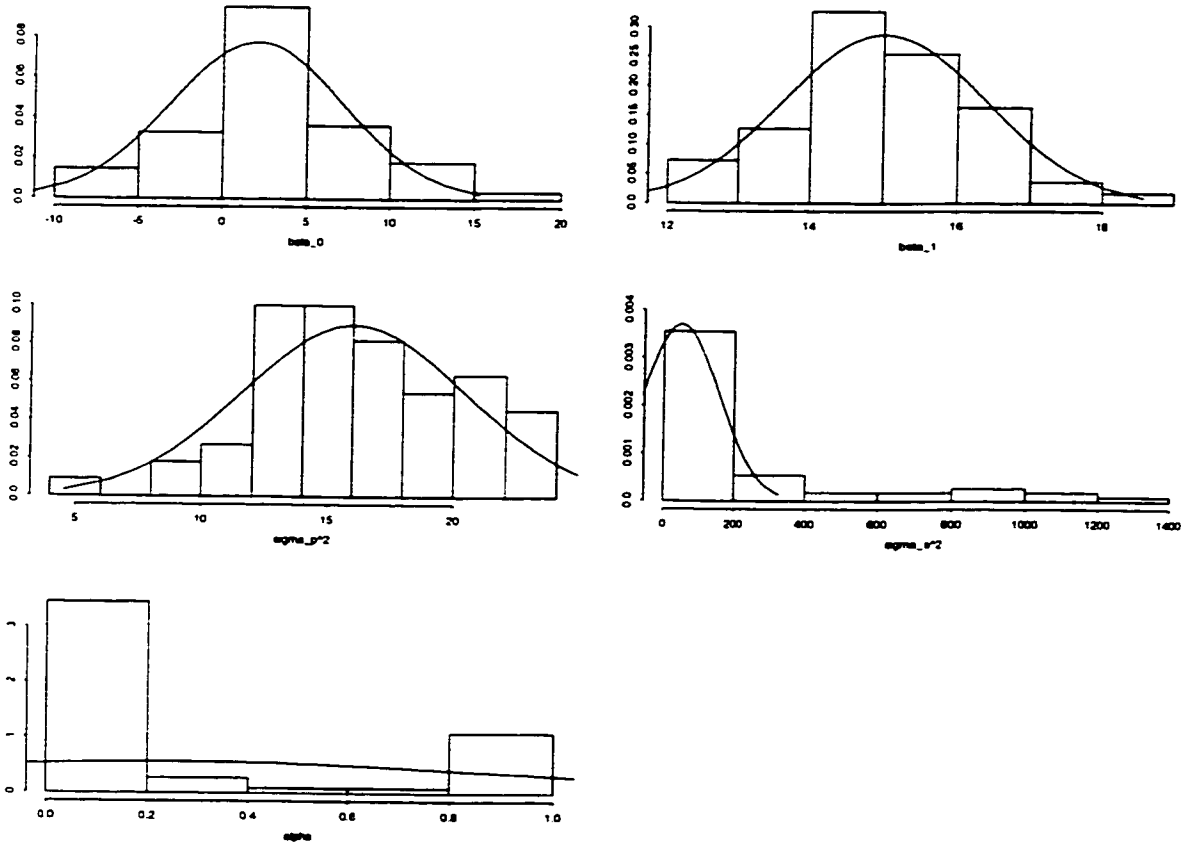


Figure 5: Distributions of estimates for Example A with asymptotic normal density curves. Histogram scaled as a probability density. The sum of the bar height times the bar width equals one.

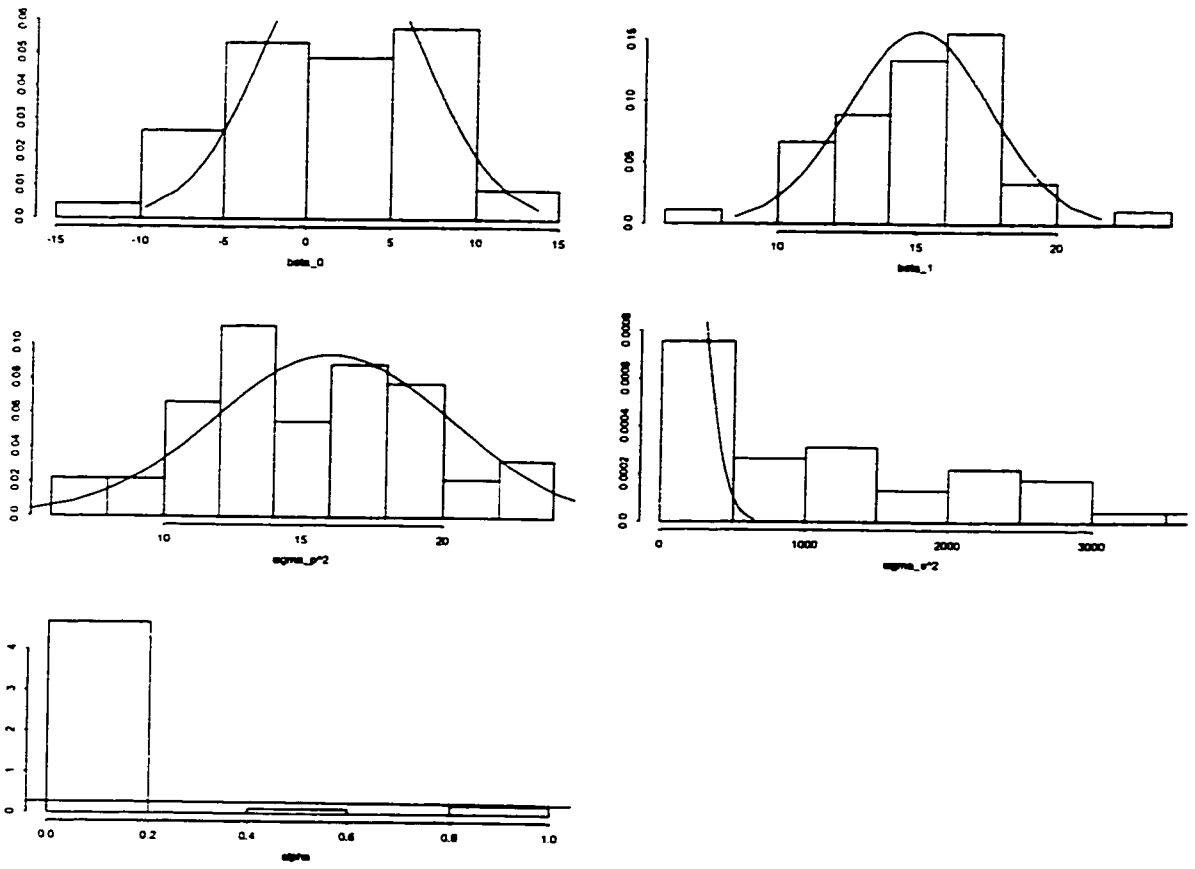


Figure 6: Distributions of estimates for Example B with asymptotic normal density curves. Histogram scaled as a probability density. The sum of the bar height times the bar width equals one.

Example C: Modified Increasing Domain Sampling Scheme

Here the sampling domain is increasing but at a different rate than in Example A. The four corners of the $n \times n$ lattice have coordinates $(0,0)$, $(\sqrt{n},0)$, $(0,\sqrt{n})$ and $(\sqrt{n},\sqrt{n})$. The other locations are filled in to create grids of equal size. Our estimation procedure is run 100 times on this sampling scheme. The distributions of estimates for Example C with asymptotic normal density curves are shown in Figure 7. Again, only the simulations containing nonnegative estimates of σ_s^2 are used in these plots. From 100 simulations 1 did not converge and 45 negative estimates of σ_s^2 are generated.

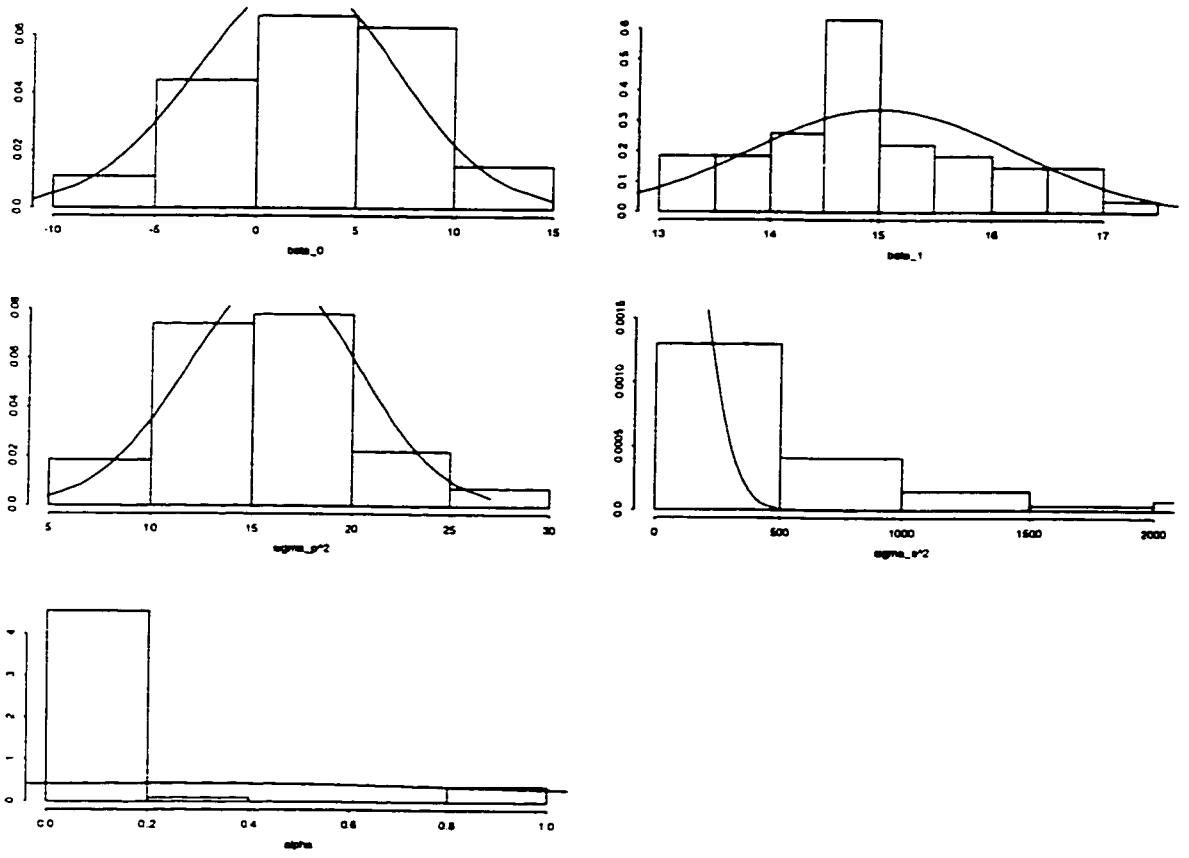


Figure 7: Distributions of estimates for Example C with asymptotic normal density curves. Histogram scaled as a probability density. The sum of the bar height times the bar width equals one.

Remark

While running the estimation procedure for the three examples, σ_s^2 in particular, is estimated. The consistency condition for $\widehat{\sigma_s^2}$ in Theorem 2.1 on page 24 is calculated for $n_s = 4, 9, 16, 25$. Results are shown in Table 13. It appears that the estimate for $\widehat{\sigma_s^2}$ in Example A and more slowly in Example C, may be consistent as the number of sites increases.

Table 13: Consistency Condition for $\widehat{\sigma_s^2}$

n_s	Example A	Example B	Example C
4	.481	.481	.462
9	.417	.477	.436
16	.346	.478	.420
25	.284	.479	.407

Table 14: Asymptotic Variances of $\widehat{\sigma_p^2}$ and $\widehat{\sigma_s^2}$: Calculating $\mathcal{I}^{-1}$ with $n_p = 1$

n_s	Example A		Example B		Example C	
	$\widehat{\sigma_p^2}$	$\widehat{\sigma_s^2}$	$\widehat{\sigma_p^2}$	$\widehat{\sigma_s^2}$	$\widehat{\sigma_p^2}$	$\widehat{\sigma_s^2}$
4	599	294489	599	294489	735	100943
9	99	25384	87	88447	94	30858
16	48	13882	40	54398	42	21555
25	29	9689	23	40946	25	17576
36	19	7541	16	34369	16	15089
144	4.5	2803	3.6	24420	3.8	9029

Table 14 shows the asymptotic variance of the maximum likelihood estimate of σ_p^2 and σ_s^2 for all three sampling schemes. Values are obtained by calculating $\mathcal{I}^{-1}$ with $n_p = 1$ using the programs *exampleA, exampleB, and exampleC.* It appears Example A yields a smaller asymptotic variance for $\widehat{\sigma_s^2}$ than the other two examples. The variance of the spatial error is not greatly improved for infill asymptotics. The rate at which the domain is increasing seems to have an effect on the variances of $\widehat{\sigma_p^2}$ and

$\widehat{\sigma_s^2}$ as is demonstrated by comparing the values for Example A and Example C in Table 14.

4.5 Technical Results

Lemma 4.3 *Let λ_i and ψ_i be the eigenvalues and corresponding $n_s \times 1$ normalized eigenvectors of M for $i = 1, \ldots, n_s$. Let M and N be symmetric matrices. Then the following equations hold for integer $k > 0$.*

$$\begin{aligned} tr(M^k) &= \sum_{i=1}^{n_s} \lambda_i^k \\ tr(M^k N) &= \sum_{i=1}^{n_s} \lambda_i^k \langle \psi_i', N\psi_i \rangle \end{aligned}$$

Similarly,

$$\begin{aligned} tr(N^k) &= \sum_{i=1}^{n_s} \lambda_i^0 \langle \psi_i', N^k \psi_i \rangle \\ tr(N^k M) &= tr(MN^k) = \sum_{i=1}^{n_s} \lambda_i \langle \psi_i', N^k \psi_i \rangle \\ tr(MNMN) &= \sum_{i=1}^{n_s} \sum_{r=1}^{n_s} \lambda_i \lambda_r \psi_r' N \psi_i \psi_i' N \psi_r . \end{aligned}$$

The operation $\langle\ ,\ \rangle$ denotes the familiar dot product.

Proof: Consider the spectral decomposition $M = \sum_i \lambda_i \psi_i \psi_i'$. such that I $= \sum_i \psi_i \psi_i'$, for $i = 1, \ldots, n_s$. Then

$$\begin{aligned} M^k &= \sum \lambda_i^k \psi_i \psi_i' \\ \Rightarrow M^k N &= \sum \lambda_i^k \psi_i \psi_i' N \\ \Rightarrow tr(M^k N) &= \sum \lambda_i^k tr(\psi_i \psi_i' N) \\ &= \sum \lambda_i^k \langle \psi_i', N\psi_i \rangle \quad \text{since N} = \text{N}'. \end{aligned}$$

Since the k^{th} power of a symmetric matrix is symmetric, the above argument can be applied to $tr(NM^k)$ to give

$$tr(N^k M) = tr(MN^k) = \sum \lambda_i \langle \psi_i', N^k \psi_i \rangle .$$

Next,

$$\begin{aligned} tr(N^k) = tr(N^k \mathrm{I}) &= tr(N^k \sum \psi_i \psi_i') \\ &= \sum tr(N^k \psi_i \psi_i') \\ &= \sum tr(\psi_i' N^k \psi_i) \\ &= \sum \lambda_i^0 \langle \psi_i', N^k \psi_i \rangle. \end{aligned}$$

Finally,

$$\begin{aligned} tr(MNMN) &= \sum_i \lambda_i tr(\psi_i \psi_i' NMN) \\ &= \sum_i \lambda_i tr(MN\psi_i \psi_i' N) \\ &= \sum_i \lambda_i \sum_r \lambda_r tr(\psi_r \psi_r' N \psi_i \psi_i' N) \\ &= \sum_i \sum_r \lambda_i \lambda_r tr(\psi_i' N \psi_r \psi_r' N \psi_i) \\ &= \sum_i \sum_r \lambda_i \lambda_r \psi_i' N \psi_r \psi_r' N \psi_i \end{aligned}$$

■

Chapter 5

Experimental Design

In this chapter we develop some rules for experimental design regarding site locations and comment on the number of sites and observations per site required. This is largely based on asymptotics and simulations.

5.1 Location of Sites

Let the number of sites and observations per site be fixed and known. Recall the asymptotic variance of parameter estimates

$$\mathcal{I}^{-1} = \begin{pmatrix} M_b^{-1} & 0 \\ 0 & M_\theta^{-1} \end{pmatrix},$$

where $M_b = (X'\mathrm{V}^{-1}X)_{2\times 2}$ and $M_\theta = (t_{ml}/2)_{3\times 3}$. Suppose a researcher has some knowledge of the parameter values of our model and wishes to construct a sampling scheme which minimizes this variance in some way.

Consider the trace norm of a symmetric matrix T, $||T||^2 = tr(TT') = tr(T^2)$. Now for $\mathcal{I}^{-1}$,

$$\begin{aligned} ||\mathcal{I}^{-1}||^2 &= tr(\mathcal{I}^{-2}) \\ &= tr\begin{pmatrix} M_b^{-2} & 0 \\ 0 & M_\theta^{-2} \end{pmatrix} \\ &= tr(M_b^{-2}) + tr(M_\theta^{-2}) \end{aligned}$$

$$= ||M_b^{-1}||^2 + ||M_\theta^{-1}||^2.$$

Specifically, we can minimize this sum. That is,

$$\begin{aligned} ||M_b^{-1}||^2 &= tr(M_b^{-1}M_b^{-1'}) = tr(M_b^{-2}) \\ &= \sum_h \lambda_h(M_b^{-2}) \\ &= \sum_h \lambda_h^{-2}(M_b). \end{aligned}$$

Similarly,

$$||M_\theta^{-1}||^2 = \sum_h \lambda_h^{-2}(M_\theta).$$

The eigenvalues of M_b and M_θ involve their entries which can in turn be written in terms of the eigenvalues and eigenvectors of Σ as is demonstrated in Section 4.3. Formulas for the eigenvalues of a 2×2 and 3×3 symmetric matrix are found in Appendix A. In fact, the eigenvalues of M_b are

$$\lambda = \frac{M_b[1,1] + M_b[2,2] \pm \sqrt{(M_b[1,1] - M_b[2,2])^2 + 4M_b^2[1,2]}}{2}, \tag{42}$$

where $M_b[1,1]$, $M_b[2,2]$ and $M_b[1,2]$ are given by (40) and (41). Those of M_θ are the roots of the characteristic equation

$$\begin{aligned} \lambda^3 + (-t_{11} - t_{33} - t_{22})\lambda^2 + (-t_{12}^2 + t_{11}t_{22} + t_{11}t_{33} + t_{22}t_{33} - t_{13}^2 - t_{23}^2)\lambda \\ +(-t_{11}t_{22}t_{33} + t_{13}^2t_{22} + t_{12}^2t_{33} + t_{11}t_{23}^2 - 2t_{12}t_{13}t_{23}). \end{aligned} \tag{43}$$

where each t_{ml} can be expressed in terms of eigenvalues and eigenvectors of Σ. Although complex, the sums of (42) and (43) can be minimized numerically, thus creating an 'optimal' sampling scheme for a fixed number of sites and observations per site.

One useful application involves determining the location of one additional site to existing sites. This experiment is applied to our example first with 4 sites and $n_p = 1$.

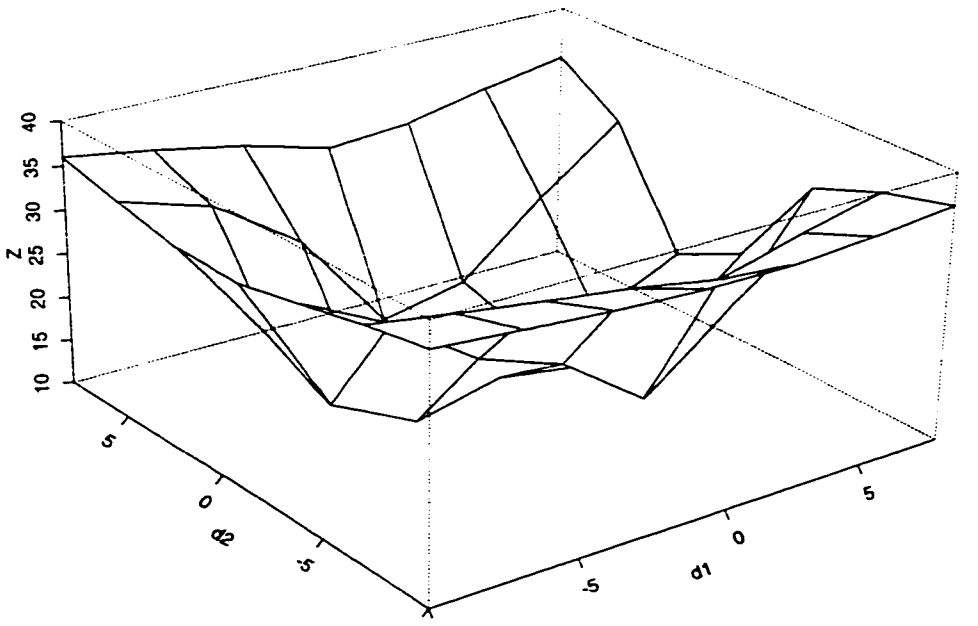


Figure 8: Surface of the norm of the asymptotic variance when adding a 5th site. Norm in thousands.

The problem is to determine the location of the 5th site while minimizing the norm of $\mathcal{I}^{-1}$. Let (d_1, d_2) be the coordinates of the 5th site somewhere on a predetermined grid. The matrices M_b and M_θ are now a function of d_1 and d_2. The regressor for the 5th site is also a function of d_1 and d_2 from a fitted quadratic surface to the existing site locations. A numerical grid search in the *S-Plus* program *experiment* finds the minimum Euclidean norm of $\mathcal{I}^{-1}$. The norm surface is plotted in Figure 8. From these results, the 5th site should be placed approximately at $(-4, -1)$ to minimize the asymptotic variance. At this location, the norm is 11,667 and

$$\mathcal{I}^{-1} = \begin{pmatrix} 21.91 & -.053 & 0 & 0 & 0 \\ -.053 & .0003 & 0 & 0 & 0 \\ 0 & 0 & 552 & -1430 & 1.53 \\ 0 & 0 & -1430 & 11477 & -49 \\ 0 & 0 & 1.53 & -49 & .496 \end{pmatrix}.$$

The choice of grid size and separation influences results.

The experiment is now repeated with 10 sites. The problem is to determine the

location of the 11th site so as to minimize the norm of the asymptotic variance. Here a quadratic surface is fitted to the existing 10 sites to obtain an expresssion for the 11th regressor as a function of d_1 and d_2. The norm surface for the chosen grid is shown in Figure 9. It is difficult to see from this graph that the minimum occurs at (-2,16). Figure 10 magnifies the view in the neighborhood of this point. At this optimal location of the 11th site, the asymptotic norm is 5209 and

$$\mathcal{I}^{-1} = \begin{pmatrix} 8.14 & -.00199 & 0 & 0 & 0 \\ -.002 & .00005 & 0 & 0 & 0 \\ 0 & 0 & 404 & -318 & -7.37 \\ 0 & 0 & -318 & 5173 & -24 \\ 0 & 0 & -7.37 & -24 & .426 \end{pmatrix}.$$

Another idea is to use this this approach iteratively to place one site after another. That is, we start with 4 sites and add the 5th one optimally. Next, we use these 5 sites to add a 6th one optimally and so on, each time fitting the quadratic surface of the regressors. Due to time constraints, grid size and spacing varies for each iterative step. The 6 added sites have coordinates

$$(-4,-1),(10,2),(14,6),(20,8),(21,13),(23,15).$$

Results of this procedure are compared to Example 2.1.1 with $n_s = 10$ and $n_p = 1$. The norm of the asymptotic variance matrix obtained by the iterative technique is 5591 while the norm for Example 2.1.1 works out to be 7170. The asymptotic variances of the parameters are given in Table 15. Thus, it appears that this approach yields not only a smaller norm than that with the haphazard choice of 6 additional sites, but also smaller variances for most of the parameters.

Table 15: Asymptotic variance for two sets of site locations

	β_0	β_1	σ_p^2	σ_s^2	α
with 10 sites of Example 2.1.1	13.8	.0001	453	7156	.62
with 10 sites- 6 of which are added using the iterative approach	9.2	.000003	523	5502	.44

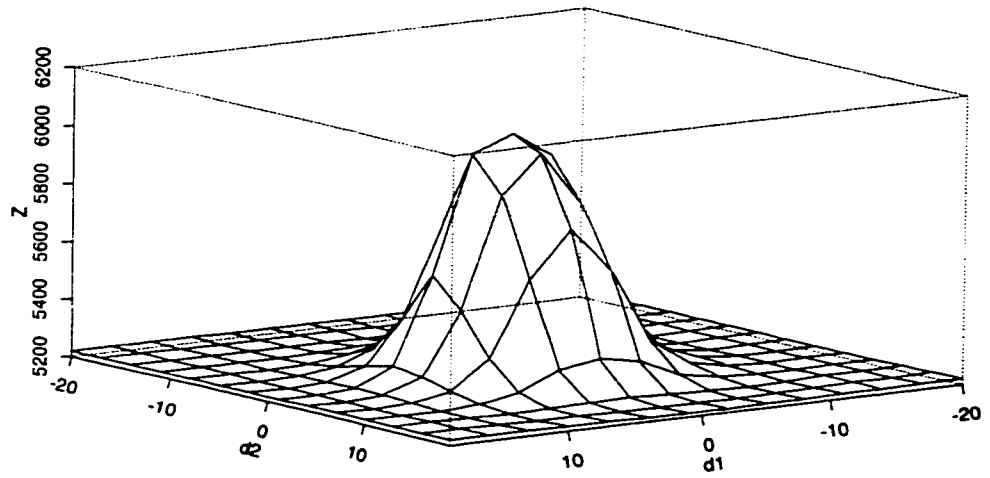


Figure 9: Surface of the norm of the asymptotic variance matrix when adding an 11th site

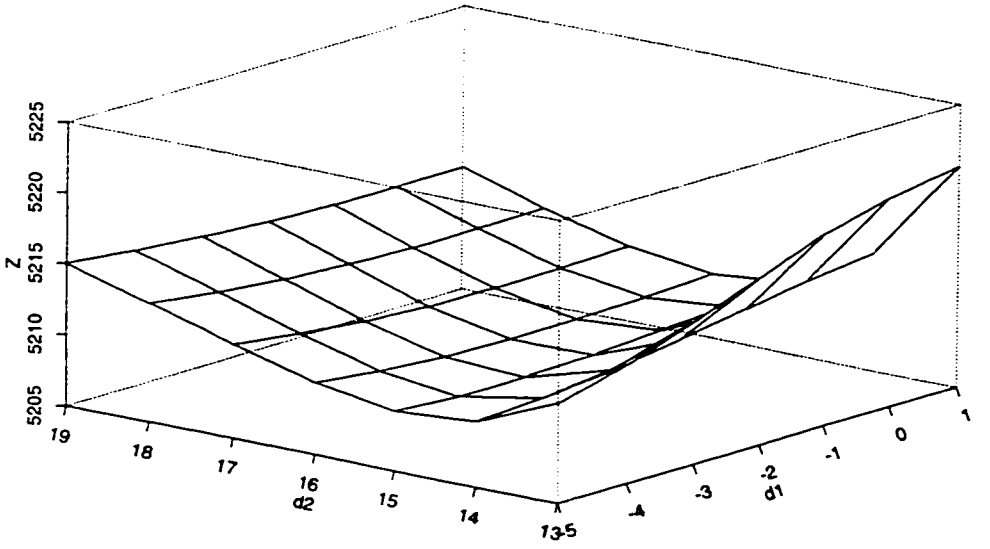


Figure 10: Neighborhood of the minimum of Figure 9. A grid step size of 1 is used for this figure.

5.2 Determining n_s and n_p

This discussion is largely based on the asymptotic results and simulations for specific designs presented in Chapter 4.

The theoretical results of Section 4.2.1 suggest that only the $\mathrm{Var}(\widehat{\sigma_p^2})$ goes to zero as the number of observations per site, n_p, goes to infinity. From Table 12 it appears that a rather large number of observations per site is required to bring the variance of $\widehat{\sigma_s^2}$ to within its asymptotic limit. The limits for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\alpha}$ are attained with fewer observations per site.

Equally as disappointing is the number of sites required to obtain an acceptable variance of $\widehat{\sigma_s^2}$ and $\widehat{\sigma_p^2}$. Although we can only comment on the sampling schemes described in Section 4.4.1, Table 14 suggests a large number of sites is required even for such simple schemes. The asymptotic variance of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\alpha}$ for n_s as large as 144 and $n_p = 1$ varies for the three sampling schemes but is not as large as that of $\widehat{\sigma_s^2}$ or $\widehat{\sigma_p^2}$. Example A with $n_s = 144$ yields the best results with the asymptotic variance for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\alpha}$ being 12.33, .15, and .095 respectively.

Chapter 6

Comments and Future Directions

Up to this point, only the simple linear model described in (4) is used. In this chapter, we investigate and comment on how the results from previous chapters apply or can be adapted to more general models. Whenever possible, modifications are made to previous results. Models for which existing results cannot be applied suggest areas of future research.

6.1 A More General Model

In this section, we expand the model in (4) to allow for more general regression and unbalanced data. The estimation of the parameters and their asymptotic variance under these schemes are discussed. A richer family of covariograms is also considered. A likelihood ratio test statistic is developed to test the hypothesis $H_o : \alpha_t = 0$ when the covariogram matrix $\Sigma = \sum_t \alpha_t C_t$ with $\sum_t \alpha_t = 1$.

Except for specific examples and corresponding conclusions drawn from them, the theoretical results in this thesis are valid for both isotropic and anisotropic covariogram models. The results require only that the spatial process be second-order stationary and that $c(h_{ii'}) = c(h_{i'i})$.

Regression with Varying x_i

In Chapter 2, although there are multiple observations y_{ij} at each site, the x_i is constant at each site. Here, there is a regressor x_{ij} for each observation j at site i. Therefore, the model becomes

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \sigma_s \varepsilon_{s_i} + \sigma_p e_{ij}$$

with $i = 1, \ldots, n_s, j = 1, \ldots, n_p$ and the random variables ε_{s_i} and e_{ij} as before. This implies that only the matrix X in the model $y = \mathrm{X}\mathrm{b} + \sigma_s \varepsilon_s + \sigma_p e$ changes. Therefore, the variance-covariance matrix V, which does not involve X, remains exactly the same.

Estimation: The estimation procedure for the parameters $\beta_0, \beta_1, \sigma_p^2$ and α, is unaffected by the current regression scheme since these parameters are maximized numerically. The estimation procedure of σ_s^2 can be applied with a slight modification. Here define $\mu_{ij} = \beta_0 + \beta_1 x_{ij}$ and

$$D_{ir} = [(\bar{y}_{i\cdot} - \bar{\mu}_{i\cdot}) - (\bar{y}_{r\cdot} - \bar{\mu}_{r\cdot})]^2,$$

so that $D_{ir} \sim (2\sigma_p^2/n_p + \sigma_s^2 R_{ir})\chi_1^2$ as before, and Theorem 2.1 still holds.

Asymptotics: Firstly, we consider the case where n_p goes to infinity in the asymptotic variance of the parameter estimates. The matrix $\mathrm{X}\mathrm{V}^{-1}\mathrm{X}$ is

$$\begin{pmatrix} n_s n_p + n_p^2 \sum_r^{n_s} \sum_i^{n_s} p_{ir} & \sum_i^{n_s} \sum_j^{n_p} x_{ij} + n_p \sum_i^{n_s} \sum_j^{n_p} x_{rj} \sum_i^{n_s} p_{ir} \\ \sum_i^{n_s} \sum_j^{n_p} x_{ij} + \sum_r^{n_s} \sum_i^{n_s} \sum_j^{n_p} p_{ir} x_{ij} & \sum_i^{n_s} \sum_j^{n_p} x_{ij}^2 + \sum_r^{n_s} \sum_t^{n_p} x_{rt} \sum_i^{n_s} \sum_j^{n_p} p_{ir} x_{ij} \end{pmatrix}.$$

The entries now involve sums from 1 to n_p, and it is not evident how these sums behave when letting n_p go to infinity. Therefore, Lemma 4.1 no longer applies for $\hat{\mathrm{b}} = (\hat{\beta}_0, \hat{\beta}_1)$. Fortunately, results pertaining to the asymptotic variance of $\widehat{\sigma_p^2}, \widehat{\sigma_s^2}$ and $\hat{\alpha}$ are unchanged since they involve only V and its partial derivatives. Hence, Lemma 4.2 holds.

In the case of n_s going to infinity, Theorem 4.2, due to Mardia and Marshall (1984), requires that $(\mathrm{X}'\mathrm{X})^{-1} \to 0$. All the other conditions of this theorem are unaffected by the modification.

Multiple Regression

Next, we consider a multiple regression scheme where there are several regressors for each observation at each site. The model is

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij} + \sigma_s \varepsilon_{s_i} + \sigma_p e_{ij} .$$

The matrix X and column vector b are now

$$\mathrm{X} = \begin{pmatrix} 1 & x_{111} & \cdots & x_{p11} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n_sn_p} & \cdots & x_{pn_sn_p} \end{pmatrix}_{n_sn_p \times (p+1)}, \quad b = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{(p+1)\times 1}$$

Estimation: Again the estimation procedure for the parameters is unaffected by the multiple regression scheme. The construction of the estimate of σ_s^2 requires a slight change. In Theorem 2.1 define $u_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij}$. With this adjustment, Theorem 2.1 holds.

Asymptotics: The variance matrix V is unchanged since it is not a function of X nor b. Therefore, the asymptotic results of Lemma 4.2 regarding $\widehat{\sigma_p^2}$, $\widehat{\sigma_s^2}$ and $\hat{\alpha}$ for n_p going to infinity still hold. Lemma 4.1, however, cannot be used to deduce the asymptotic variance of $\hat{\mathrm{b}}$. Again, the entries in $\mathrm{X}'\mathrm{V}^{-1}\mathrm{X}$ involve sums from $j = 1, \ldots, n_p$ whose behavior is not known as n_p goes to infinity.

Results from Mardia and Marshall (1984) regarding asymptotic behavior of parameter estimates as n_s goes to infinity still require $(\mathrm{X}'\mathrm{X})^{-1} \to 0$ for this design matrix X.

Unbalanced Data

In the case of unbalanced data, the number of observations per site can vary. Let k_i be the number of observations at site i. The regression model of Chapter 2 becomes

$$y_{ij} = \beta_0 + \beta_1 x_i + \sigma_s \varepsilon_{s_i} + \sigma_p e_{ij}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k_i$. Let $d = \sum_i k_i$ and denote the $k_i \times k_r$ matrix of ones by $\mathrm{U}_{k_i k_r}$. Then the $d \times d$ covariance matrix of the observations y_{ij} is $V = \sigma_p^2 \mathrm{I}_d + \sigma_s^2 Q$

where

$$Q = \begin{pmatrix} c(h_{11})\mathrm{U}_{k_1k_1} & c(h_{12})\mathrm{U}_{k_1k_2} & \cdots & c(h_{1n_s})\mathrm{U}_{k_1k_{n_s}} \\ c(h_{21})\mathrm{U}_{k_2k_1} & c(h_{22})\mathrm{U}_{k_2k_2} & \cdots & c(h_{2n_s})\mathrm{U}_{k_2k_{n_s}} \\ \vdots & \vdots & \ddots & \vdots \\ c(h_{n1})\mathrm{U}_{k_{n_s}k_1} & c(h_{n2})\mathrm{U}_{k_{n_s}k_2} & \cdots & c(h_{n_sn_s})\mathrm{U}_{k_{n_s}k_{n_s}} \end{pmatrix}_{d\times d}.$$

The matrix Q cannot be expressed as a Kronecker product since the U's are all of different dimensions.

Estimation: The estimation prodecure for all the parameters except σ_s^2 remains unchanged. Theorem 2.1 is appropriately modified. Recall that $R_{ir} = 2v - 2c(h_{ir})$ and $D_{ir} = [\sigma_s(\varepsilon_{s_i} - \varepsilon_{s_r}) - \sigma_p(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})]^2$.

Theorem 6.1 *(Unbalanced case of Theorem 2.1) For $i, i', r, r' = 1 \ldots n_s$, define $D = (D_{ir})$, $R = (R_{ir})$ and $g \equiv g(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'})$ as in Theorem 2.1. Let*

$$CST \equiv \begin{pmatrix} 1/k_1 + 1/k_1 \\ 1/k_1 + 1/k_2 \\ \vdots \\ 1/k_1 + 1/k_{n_s} \\ \vdots \\ 1/k_{n_s} + 1/k_{n_s} \end{pmatrix}_{n_s^2\times 1}.$$

Then the estimate $\widehat{\sigma_s^2} = (R'R)^{-1}R'(D - \sigma_p^2 CST)$ is unbiased with the same consistency condition as in Theorem 2.1 as $n_s \to \infty$ and $k_i \to \infty$ $\forall i$.

Proof: The proof parallels that of Theorem 2.1. Here the pure errors are heteroskedastic so that $\bar{e}_{i\cdot} \sim \mathcal{N}(0, 1/k_i)$. The expressions involving the spatial errors remain unchanged. Now $N = \bar{e}_{i\cdot} - \bar{e}_{r\cdot} \sim \mathcal{N}(0, 1/k_i + 1/k_r)$, so that for $i \neq r$,

$$\begin{aligned} E(D_{ir}) &= \sigma_s^2 R_{ir} + \sigma_p^2(1/k_i + 1/k_r) \\ \Rightarrow E(D) &= \sigma_s^2 R + \sigma_p^2 CST, \end{aligned}$$

and unbiasedness of $\widehat{\sigma_s^2}$ follows.

Now expressions for the entries of $\mathrm{Var}(D) = (\mathrm{Cov}(D_{ir}, D_{i'r'}))_{n_s^2 \times n_s^2}$ which lead to the consistency condition, are found. The

$$\mathrm{Cov}(D_{ir}, D_{i'r'}) = 0 \quad \text{if } i = r \text{ or } i' = r'.$$

The distribution of D_{ir} is a $(\sigma_s^2 R_{ir} + \sigma_p^2(1/k_i + 1/k_r))\chi_1^2$ with variance

$$\begin{aligned} \mathrm{Var}(D_{ir}) &= \mathrm{Cov}(D_{ir}, D_{ir}) \\ &= 2\left(\sigma_s^2 R_{ir} + \sigma_p^2(1/k_i + 1/k_r)\right)^2 \quad \text{if } i \neq r. \end{aligned}$$

Multivariate normal theory is again applied to find the components of $\mathrm{Cov}(D_{ir}, D_{ir'})$ for $r \neq r'$. After much algebraic manipulation

$$\begin{aligned} E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2 &= (1/k_i + 1/k_{r'})(1/k_i + 1/k_r) + 2/k_i^2, \\ E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\varepsilon_{s_i} - \varepsilon_{s_{r'}})^2 &= 2g^2(0, h_{ir'}, h_{ri}, h_{rr'}) + R_{ir'}R_{ir}, \\ E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot})^2 &= R_{ir}(1/k_i + 1/k_{r'}), \\ E(\varepsilon_{s_i} - \varepsilon_{s_{r'}})^2(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2 &= R_{ir'}(1/k_i + 1/k_r), \\ E(\varepsilon_{s_i} - \varepsilon_{s_r})(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})(\varepsilon_{s_i} - \varepsilon_{s_{r'}})(\bar{e}_{i\cdot} - \bar{e}_{r'\cdot}) &= g(0, h_{ir'}, h_{ri}, h_{rr'})/k_i. \end{aligned}$$

Using the equation (15) when $r \neq r'$, the covariance term $\mathrm{Cov}(D_{ir}, D_{ir'}) \to 2\sigma_s^4 g^2$ as $k_i, k_r, k_{r'}$ all go to infinity.

Finally, the following terms are used to evaluate $\mathrm{Cov}(D_{ir}, D_{i'r'})$ for $i \neq i'$.

$$\begin{aligned} E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\varepsilon_{s_{i'}} - \varepsilon_{s_{r'}})^2 &= 2g^2(h_{ii'}, h_{ir'}, h_{ri'}, h_{rr'}) + R_{ir}R_{i'r'}, \\ E(\varepsilon_{s_i} - \varepsilon_{s_r})^2(\bar{e}_{i'\cdot} - \bar{e}_{r'\cdot})^2 &= R_{ir}(1/k_{i'} + 1/k_{r'}), \\ E(\varepsilon_{s_{i'}} - \varepsilon_{s_{r'}})^2(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2 &= R_{i'r'}(1/k_i + 1/k_r), \\ E(\bar{e}_{i\cdot} - \bar{e}_{r\cdot})^2(\bar{e}_{i'\cdot} - \bar{e}_{r'\cdot})^2 &= (1/k_i + 1/k_r)(1/k_{i'} + 1/k_{r'}). \end{aligned}$$

So using (16), the $\mathrm{Cov}(D_{ir}, D_{i'r'}) \to 2\sigma_s^4 g^2$ as the k's all go to infinity.

The rest of the proof follows and the consistency condition is the same as that found in Theorem 2.1. ■

Asymptotics: Since much of the theory involving n_p going to infinity in this thesis requires that V, and consequently V^{-1}, be written in terms of Kronecker products, it not possible to extend any results in this case. For n_s going to infinity, the conditions of Theorem 4.2 remain unchanged for unbalanced data.

Likelihood Ratio Test Statistic

Suppose the matrix Σ can now be written as a linear combination of several covariograms, i.e. $\Sigma = \sum_{t=1}^{m} \alpha_t C_t$ where C_t is a matrix of covariogram models $c_t(h)$ and $\sum_t \alpha_t = 1$. We can formulate a likelihood ratio test statistic (LRT) to test the hypothesis $H_o : \alpha_p = 0$. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and $\hat{\alpha}_o$ represent the vector with the restriction for α_p given in H_o. Denote the unrestricted maximum likelihood estimate by $\hat{\Phi} = (\hat{\beta}_0, \hat{\beta}_1, \widehat{\sigma_p^2}, \widehat{\sigma_s^2}, \hat{\alpha})$ and the resticted maximum likelihood estimate by $\hat{\Phi}_o = (\hat{\beta}_0, \hat{\beta}_1, \widehat{\sigma_p^2}, \widehat{\sigma_s^2}, \hat{\alpha}_o)$. Then the LRT is

$$\Lambda(y) = \frac{L(\hat{\Phi}_o|y)}{L(\hat{\Phi}|y)}.$$

The numerator is the value that maximizes the likelihood (or log-likelihood) function under the parameter restriction and the denominator is the unrestricted maximum of the likelihood function. In practice, $-2log\Lambda$ is generally calculated. Recall that $\hat{\theta} = (\widehat{\sigma_p^2}, \widehat{\sigma_s^2}, \hat{\alpha})$. Using the log-likelihood equation (9), and simplifying gives

$$-2log\Lambda = ln|\mathrm{V}(\hat{\theta})| - ln|\mathrm{V}(\hat{\alpha_o})| + \frac{1}{\widehat{\sigma_p^2}}\{(y - \mathrm{X}\hat{\mathrm{b}})'[(P(\hat{\alpha_o}) - P(\hat{\alpha})) \otimes U](y - \mathrm{X}\hat{\mathrm{b}})\}.$$

Using expression (34) for P, it is possible to simplify the LRT further to

$$\begin{aligned} -2log\Lambda &= ln|V(\hat{\theta})| - ln|V(\hat{\alpha_o})| \\ &+ \frac{1}{\widehat{\sigma_p^2} n_p}\left\{(y - \mathrm{X}\hat{\mathrm{b}})'\left[\left((\mathrm{I} + \hat{\delta}\Sigma(\hat{\alpha_o}))^{-1} - (\mathrm{I} + \hat{\delta}\Sigma(\hat{\alpha}))^{-1}\right) \otimes U\right](y - \mathrm{X}\hat{\mathrm{b}})\right\}. \end{aligned}$$

It is difficult to identify the distribution of this LRT. Numerical methods may be used to to estimate its distribution with the rejection region of $\{y : -2log\Lambda \geq k\}$, for $0 \leq k \leq \infty$.

6.2 Incorporating Topography

To increase the applicability of the model given in (4), it is possible to incorporate topography either through the regressors, the choice of anisotropic covariogram models or both. One possible area of future research is to study the effect of incorporating topography in this fashion.

In certain situations is it reasonable to generate predictor variables from topographic maps. This is facilitated by GIS software which can display and query data with relative ease. The following is an application which incorporates topography.

Consider a mountain range formed by layers of varying sedimentary strata. The mineralization of rock may have occured differently throughout the horizontal layers (as a result of pressure) perhaps requiring the use of anisotropic covariogram models. Furthermore, the regressors can be a function of strata so that topography enters into the model in two ways.

One possible model is the following. For example, consider an anisotropic covariogram for the spatial error. Since the strata are horizontal, the spatial dependence in the east-west direction is very different than the one in the north-south direction. The covariogram, therefore, is a function of direction as well as distance. The matrix Σ can be an unknown linear combination of two such anisotropic covariogram matrices. Of course, this requires that the observations be classified according to strata and that the topography be fairly well understood.

Another possibility is to take a linear combination of a different kind. Let $c_1 \equiv c_1(||h||_1)$. This isotropic covariogram reflects the covariance of the spatial error simply as a function of distance. Suppose another covariogram, $c_2 \equiv c_2(||h||_2)$, reflects the covariance as a function of sedimentary layer distance only. Then Σ can be seen as a linear combination of these two covariograms.

Appendix A

Algebra and Matrix Theory

This appendix gives selected results from various references. The general references for eigenvalues and Kronecker products are Schott (1997) and Harville (1997). The section on matrix series convergence takes results from Graham (1987). The conditional multivariate normal and definitions of quadratic forms can be found in Johnson and Wichern (1988).

A.1 Eigenvalues

Notation: Denote $\lambda_h(M),\ h = 1, \ldots, m$ to be the eigenvalues of an $m \times m$ matrix M and $\lambda_{(h)}(M)$ to be the ordered eigenvalues.

Definition A.1 *Let A be an $m \times m$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. The spectral radius of A, denoted $\rho(\mathrm{A})$ is defined to be*

$$\rho(\mathrm{A}) = \max_{1 \le h \le m} |\lambda_h|.$$

Lemma A.1 *If the $m \times m$ matrix A has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ then the matrix $(\mathrm{A} + \gamma \mathrm{I})$ has eigenvalues $\lambda_1 + \gamma, \lambda_2 + \gamma, \ldots, \lambda_m + \gamma$.*

Theorem A.1 *Let A be an $m \times m$ symmetric matrix and B be an $m \times m$ positive definite matrix. Then the following inequality holds for the ordered eigenvalues.*

$$\lambda_{(h)}(A + B) > \lambda_{(h)}(A) \quad for \quad h = 1, \ldots, m.$$

Results for a 2×2 symmetric matrix

Let $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. Then the eigenvalues of A are given by

$$\lambda = \frac{(a+c) \pm \sqrt{(a-c)^2 + 4b^2}}{2}.$$

Results for a 3×3 symmetric matrix

Let $T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{12} & t_{22} & t_{23} \\ t_{13} & t_{23} & t_{33} \end{pmatrix}$. Then the eigenvalues of T are the roots of the characteristic equation

$$\lambda^3 + (-t_{11} - t_{33} - t_{22})\lambda^2 + (-t_{12}^2 + t_{11}t_{22} + t_{11}t_{33} + t_{22}t_{33} - t_{13}^2 - t_{23}^2)\lambda + (-t_{11}t_{22}t_{33} + t_{13}^2 t_{22} + t_{12}^2 t_{33} + t_{11}t_{23}^2 - 2t_{12}t_{13}t_{23}).$$

A.2 Kronecker Products

Let A be an $m \times n$ matrix and B be a $p \times q$ matrix. The *Kronecker product* of A and B, denoted $A \otimes B$ is the $mp \times nq$ matrix

$$\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \vdots & \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}.$$

Properties of Kronecker Products

Let A,B and C be any matrices α and β be any two scalars. Then

1. $\alpha \otimes A = A \otimes \alpha = \alpha A$,
2. $(\alpha A) \otimes (\beta B) = \alpha\beta(A \otimes B)$,
3. $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

4. $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$ if A and B are of the same size,

5. $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$ if B and C are of the same size,

6. $(A \otimes B)' = A' \otimes B'$,

7. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$,

8. $(A \otimes B)(C \otimes D) = AC \otimes BD$.

9. $tr(A \otimes B) = tr(A)tr(B)$.

Theorem A.2 *Let A be an $m \times m$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ and let B be a $p \times p$ matrix with eigenvalues $\nu_1, \nu_2, \ldots, \nu_p$. Then the set of mp eigenvalues of $\mathrm{A} \otimes \mathrm{B}$ is given by $\{\lambda_h \nu_t : h = 1, \ldots, m;\ t = 1, \ldots, p\}$. Furthermore, suppose λ is an eigenvalue of A with corresponding eigenvector x,; and ν is an eigenvalue of B with corresponding eigenvector y. Then $x \otimes y$ is the eigenvector corresponding to the eigenvalue $\lambda\nu$.*

A.3 Matrix Series Convergence

Theorem A.3 *Let A be an $m \times m$ matrix. Setting $\mathrm{A}^0 = \mathrm{I}$, $[\mathrm{I} - \mathrm{A}]^{-1} = \sum_{r=0}^{\infty} \mathrm{A}^r$ converges if and only if $\rho(\mathrm{A}) < 1$.*

Definition A.2 *Two matrices $A(k)$ and $B(k)$, are asymptotically equivalent if $A^{-1}B$ goes to a constant matrix as k goes to infinity.*

Lemma A.2 *For matrices A and B and scalar k, the matrix $(A + kB)^{-1}$ is asymptotically equivalent to $(kB)^{-1}$.*

Proof: Using the definition of asymptotically equivalent matrices,

$$\begin{aligned} \left((A + kB)^{-1}\right)^{-1} (kB)^{-1} &= (A + kB)k^{-1}B^{-1} \\ &= k^{-1}AB^{-1} + I \\ &\longrightarrow I \end{aligned}$$

A.4 Conditional Multivariate Normal

Let $X \sim \mathcal{N}(\mu, \Sigma)$ with Σ nonsingular. Assume the partition

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

so that

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right).$$

Then the conditional distribution of X_1 given X_2, denoted $X_1|X_2$, is normal with

$$\begin{aligned} E(X_1|X_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \\ \text{Var}(X_1|X_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

A.5 Quadratic Forms

Let x be an $m \times 1$ real vector and $A = (a_{ij})$ an $m \times m$ symmetric matrix. Then

i. A is *positive definite* if $x'Ax = \sum_{i,j=1}^{m} a_{ij}x_i x_j > 0 \ \forall x \neq 0$,

ii. A is *positive semi-definite* if $x'Ax \geq 0 \ \forall x \neq 0$, and $x'Ax = 0$ for some $x \neq 0$,

iii. A is *negative definite* if $x'Ax < 0 \ \forall x \neq 0$,

iv. A is *negative semi-definite* if $x'Ax \leq 0 \ \forall x \neq 0$, and $x'Ax = 0$ for some $x \neq 0$,

v. A is *conditional negative definite* if $x'Ax \leq 0$, with x satisfying $\sum_i x_i = 0$.

Appendix B

Other Tools

A reference for probability theory results in this appendix is Billingsley (1986). The reference for Section B.2 is Bickel and Doksum (1977).

B.1 Moment Generating Function (MGF)

Let X be a random variable with moment generating function $M(t)$.

k^{th} Moments

The k^{th} moment of a random variable X is the k^{th} derivative of it's moment generating function evaluated at zero. Symbolically,

$$E(X^k) = M^{(k)}(0) = \frac{d^k}{dt^k} M(t)\Big|_{t=0}.$$

Derivatives of the Normal MGF

Let X be a normal random variable with mean μ and variance σ^2. Then

$$\begin{aligned}
M(t) &= e^{\mu t + \sigma^2 t^2/2} \\
M^{(1)}(t) &= (\mu + \sigma^2 t)\, e^{\mu t + \sigma^2 t^2/2} \\
M^{(2)}(t) &= [\sigma^2 + (\mu + \sigma^2 t)^2]\, e^{\mu t + \sigma^2 t^2/2} \\
M^{(3)}(t) &= [3\sigma^2(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^3]\, e^{\mu t + \sigma^2 t^2/2}
\end{aligned}$$

$$M^{(4)}(t) = [3(\sigma^2)^2 + 6\sigma^2(\mu + \sigma^2 t)^2 + (\mu + \sigma^2 t)^4]\, e^{\mu t + \sigma^2 t^2/2}$$

B.2 Fisher Information Matrix

Let $p(x, \theta)$ be a density function and assume the following two regularity conditions on the family of distributions $\{P_\theta : \theta \in \Theta\}$.

I. The set $A = \{x : p(x, \theta) > 0\}$ does not depend on θ. For all $x \in A$, $\theta \in \Theta$, $\partial/\partial\theta \log p(x, \theta)$ exists and is finite.

II. If T is any statistic such that $E(|T|) < \infty$ for all $\theta \in \Theta$, then the operations of integration and differentiation by θ can be interchanged in $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} T(x)p(x, \theta)dx$. Then the Fisher information number is given by

$$\mathrm{I}(\theta) = E\left(\frac{\partial}{\partial\theta} \log p(x, \theta)\right)^2$$

Furthermore, if the second partial derivative satisfies conditions (I) and (II) then

$$\mathrm{I}(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2} \log p(x, \theta)\right).$$

For the vector parameter case where $\theta = (\theta_1, \ldots, \theta_p)$, the information matrix is

$$I(\theta) = \left[-E\left(\frac{\partial^2 \log p(x, \theta)}{\partial\theta_i \partial\theta_j}\right)\right].$$

B.3 Miscellaneous Definitions

Definition B.1 *The statistic $\hat{\theta}_n$ is a consistent estimator of the parameter θ if and only if for each $c > 0$*

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| < c) = 1.$$

Theorem B.1 *If $\hat{\theta}$ is an unbiased estimator of the parameter θ and $\mathrm{Var}(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta}$ is a consistent estimator of θ.*

Definition B.2 *The estimates $\hat{\theta}_n$ is weakly consistent for θ if*

$$\hat{\theta}_n = \theta + o_p(1),$$

where $o_p(1)$ denotes a random variable that is $o(1)$ in probability.

Appendix C

Splus programs

S-Plus is a package of statistical tools produced by Mathsoft, Inc. These programs were written using *S-Plus* version 3.4, release 1, on an IBM model 355 RS/6000 computer.

Program Name	Function
Optall, section 2.2.3	Optimization procedure for all 5 parameters
→ Opt3	Optimizes β_0, β_1, and σ_p^2
→ Optalpha	Optimizes alpha
→ Optsigs	Optimizes σ_s^2
Nummax, section 2.2.2	Maximizes the log-likelihood of all 5 parameters simultaneously
GSeidal, section 2.2.2	Solves likelihood equations using a Gauss-Seidal Approach
Anovsig, section 2.2.2	Estimates σ_s^2 using an analysis of variance approach
Methmom, section 2.2.2	Calculates the method of moments estimator for σ_s^2
Simulatey	Produces observations y from a $\mathcal{N}(\mathbf{Xb}, \Phi)$ where $\Phi = (2, 15, 16, 49, .25)$
→ covexp.prog	Produces exponential covariogram model
Initvalues	Finds initial values for β_0, β_1, and σ_p^2
ExampleA, section 4.4.1	Calculates the theoretical asymptotic variance of lattice data for an increasing domain sampling scheme
→ Predictor.latticeA	Creates regressors such that those in the same row of the lattice will have the same value
ExampleB, section 4.4.1	Calculates the theoretical asymptotic variance of lattice data for an infill sampling scheme
Example C , section 4.4.1	Calculates the theoretical asymptotic variance of lattice data for a modified increasing domain sampling scheme
→ Consistency	Calculates the consistency condition (17) for the asymptotic variance of $\widehat{\sigma_s^2}$ in all three examples
Experiment, section 5.1	Finds an optimal location when adding one more site.

Optall

```
#this program combines all the other optimization programs
#to give a final estimate of all 5 parameters
library(Matrix)
cat("the random seed is", .Random.seed , "\n")
for (loop in 1:100){
   source("simulatey")
   source("initvalues")
#initial estimate of (beta0,beta1,sigp^2,sigs^2,alpha)
   initial<-c(guess,81,.1)
   beta0<-initial[1]
   beta1<-initial[2]
   sigp2<-initial[3]
   sigs2<-initial[4]
   alpha<-initial[5]
```

```
  source("opt3")
         beta0<-ans$parameters[1]
         beta1<-ans$parameters[2]
         sigp2<-ans$parameters[3]
    it<-1
    tol4<-1
    tol5<-1
    while (tol4 > .01 | tol5 >.001  ) {
         source("optalpha")
         tol5<-abs(alpha-ans5$parameters)
         alpha<-ans5$parameters
         source("optsigs")
         tol4<-abs(sigs2 -ans4$coefficients)
         if (ans4$coefficients < 0) {sigs2<-ans4$coefficients
                                     tol4<-0
                                     tol5<-0} else { sigs2<-ans4$coefficients}
         update<- round(c(beta0,beta1,sigp2,sigs2,alpha),5)
         #cat(it,update," ", round(tol4,8), round(tol5,8), "\n")
         it<-it+1
         if (it>99) {tol4<-0
                     tol5<-0}
         }
    if (it>99) {finalest<-"Did NOT Converge"} else
               {finalest<- round(c(beta0,beta1,sigp2,sigs2,alpha),3)}
    cat(finalest, "   ", it-1, "\n")
}
```

Opt3

```
#this program maximizes the log-likelihood of 5 parameters
#it requires an initial guess for all the parameters
#the program is written in terms of sigma p and sigma s SQUARED.
# we will only use the estimates for beta0, beta1, and sigmap^2
yob<-matrix(c(TAB1),ncol=1)
one<-matrix(rep(1,n*k),ncol=1)
nooftimes<-rep(k,n)
colofx<-matrix(rep(xi,nooftimes),ncol=1)
X<-cbind(one,colofx)
phi<-function(est1,est2,est3,est4,est5){c(est1,est2,est3,est4,est5)}
p1<-function(phi){as.vector(t(phi)%*%c(1,0,0,0,0))}
p2<-function(phi){as.vector(t(phi)%*%c(0,1,0,0,0))}
p3<-function(phi){as.vector(t(phi)%*%c(0,0,1,0,0))}
p4<-function(phi){as.vector(t(phi)%*%c(0,0,0,1,0))}
p5<-function(phi){as.vector(t(phi)%*%c(0,0,0,0,1))}
detm<-function(M){prod(eigen(M)$values)}
b<-function(phi){matrix(c(p1(phi),p2(phi)),ncol=1)}
SIG<-function(phi){p5(phi)*cc1 + (1-p5(phi))*cc2}
U<-matrix(rep(1,k^2),ncol=k)
A<-function(phi){kronecker(SIG(phi),U)}
V<-function(phi){p3(phi) * diag(n*k) + p4(phi)*A(phi)}
IV<-function(phi){solve.Matrix(V(phi)) }
L<-function(phi){n*k*log(2*pi) + log(detm(V(phi))) + t(yob - X%*%b(phi))
                 %*% IV(phi) %*% (yob - X %*% b(phi))}
ans<-nlminb(start = initial ,obj=L,lower = c(NA, NA, .0001, .0001,0),
           upper = c(NA, NA, NA, NA, 1))
}
```

Optalpha

```
#everything is known except alpha
tr<-function(M){sum(diag(M))}
detM<-function(M){prod(eigen(M)$values)}
yob<-matrix(c(TAB1),ncol=1)
cen<-X%*%cbind(c( beta0,beta1))
TS<-yob-cen
U<-matrix(rep(1,k^2),ncol=k)
RIG<-function(alp){ alp*cc1 + (1-alp)*cc2 }
RA<-function(alp){ kronecker(RIG(alp),U) }
#IRIG<-function(alp){solve(RIG(alp))}
RV<-function(alp){(sigp2) *diag(n*k) + (sigs2) *RA(alp)}
IRV<-function(alp){solve(RV(alp)) }
DRV<-sigs2* (kronecker((cc1-cc2),U))
L<-function(alp){ -log(detM(RV(alp))) - t(TS)%*%IRV(alp) %*%TS}
DL<-function(alp){-tr(IRV(alp)%*%DRV) + t(TS)%*%IRV(alp)%*%DRV%*%IRV(alp)%*%TS}
ML<-function(alp){-L(alp)}
MDL<-function(alp){-DL(alp)}
ans5<-nlminb(start=alpha,obj=ML,gradient=MDL,lower=0,upper=1)
```

Optsigs

```
#regression technique  for estimating sigs2
#alpha,beta0,beta1,and sigp2 are known
#source("simulatey")
SIG<-alpha*cc1 + (1-alpha)*cc2
v<-SIG[1]
#recall TAB1 is a table of response(rows) by site(col) created in simulatey
yidot<-as.vector(apply(TAB1,2,mean))
mui<-beta0 + beta1*xi
tidot<-yidot-mui
dotdif<-0
for ( i in 1:n) { for (r in 1:n) { dir<-tidot[i]-tidot[r]
                                   dotdif<-append(dotdif,dir)}
                }
dotdif<- dotdif[-1]
Dir<-dotdif^2
Newir<-Dir - 2*sigp2/k
#used if taking i not equal to r (will get same answer)
#sir<-seq(1,n^2,by=n+1)
#Cir<- c(SIG[-c(sir)])
Cir<-c(SIG)
Rir<-2*v - 2*Cir
ans4<-lm(Newir~-1 + Rir)
```

Nummax

```
#this program maximizes the log-likelihood of 5 parameters
#it requires an initial guess for all the parameters
#the program is written in terms of sigma p and sigma s SQUARED.
library(Matrix)
cat("the random seed is", .Random.seed , "\n")
for (loop in 1:100) {
source("simulatey")
source("initvalues")
#initial estimate of (beta0, beta1,sigp,sigs,alpha)
```

```
initial<-c(guess,81,.1)
yob<-matrix(c(TAB1),ncol=1)
one<-matrix(rep(1,n*k),ncol=1)
nooftimes<-rep(k,n)
colofx<-matrix(rep(xi,nooftimes),ncol=1)
X<-cbind(one,colofx)
phi<-function(est1,est2,est3,est4,est5){c(est1,est2,est3,est4,est5)}
p1<-function(phi){as.vector(t(phi)%*%c(1,0,0,0,0))}
p2<-function(phi){as.vector(t(phi)%*%c(0,1,0,0,0))}
p3<-function(phi){as.vector(t(phi)%*%c(0,0,1,0,0))}
p4<-function(phi){as.vector(t(phi)%*%c(0,0,0,1,0))}
p5<-function(phi){as.vector(t(phi)%*%c(0,0,0,0,1))}
detm<-function(M){prod(eigen(M)$values)}
b<-function(phi){matrix(c(p1(phi),p2(phi)),ncol=1)}
SIG<-function(phi){p5(phi)*cc1 + (1-p5(phi))*cc2}
U<-matrix(rep(1,k^2),ncol=k)
A<-function(phi){kronecker(SIG(phi),U)}
V<-function(phi){p3(phi) * diag(n*k) + p4(phi)*A(phi)}
IV<-function(phi){solve.Matrix(V(phi)) }
L<-function(phi){n*k*log(2*pi) + log(detm(V(phi))) + t(yob - X%*%b(phi))
                 %*% IV(phi) %*% (yob - X %*% b(phi))}
ans<-nlminb(start = initial ,obj=L,lower = c(NA, NA, .0001, .0001,0),
          upper = c(NA, NA, NA, NA, 1))
cat(round(ans$parameters,4),"\n")
}
```

GSeidal

```
#this program solves the maximum likelihood equations(2)
#using a Gauss-Sedal approach
#this is written in terms of sigma p squared and sigma s squared
library(Matrix)
cat("the random seed is", .Random.seed , "\n")
for (loop in 1:100){
   source("simulatey")
   source("initvalues")
   #need only the guess for sigma p squared
   initial<-c(guess[3],81,.1)
   yob<-matrix(c(TAB1),ncol=1)
   one<-matrix(rep(1,n*k),ncol=1)
   nooftimes<-rep(k,n)
   colofx<-matrix(rep(xi,nooftimes),ncol=1)
   X<-cbind(one,colofx)
   XTR<-t(X)
   tr<-function(M){sum(diag(M))}
   theta<-function(est3,est4,est5){c(est3,est4,est5)}
   p1<-function(theta){as.vector(t(theta)%*%c(1,0,0))}
   p2<-function(theta){as.vector(t(theta)%*%c(0,1,0))}
   p3<-function(theta){as.vector(t(theta)%*%c(0,0,1))}
   SIG<-function(theta){p3(theta)*cc1 + (1-p3(theta) )*cc2}
   U<-matrix(rep(1,k^2),ncol=k)
   A<-function(theta){kronecker(SIG(theta),U)}
   V<-function(theta){p1(theta) * diag(n*k) + p2(theta)*A(theta)}
   IV<-function(theta){solve.Matrix(V(theta))}
   DV3<-function(theta){ p2(theta)* kronecker((cc1-cc2),U) }
```

```
    B<-function(theta){ solve(XTR%*%IV(theta)%*%X) %*% XTR%*%IV(theta) %*% yob }
    b<-B(initial)
    anslist<-c(b,initial,NA,NA)
    R<-yob-X%*%b
    #the newton-raphson equations to solve
    f1<-function(theta){ tr(IV(theta)) - t(R)%*%IV(theta)%*%IV(theta)%*% R }
    f2<-function(theta){ tr(IV(theta)%*%A(theta)) - t(R)%*%IV(theta)
                         %*% A(theta)%*%IV(theta)%*% R }
    f3<-function(theta){ tr(IV(theta)%*%DV3(theta)) - t(R) %*%IV(theta)
                         %*%DV3(theta)%*%IV(theta)%*% R }
    FF<-function(theta){t(c(f1(theta),f2(theta),f3(theta)))}
    TFF<-function(theta){t(FF(theta))}
    delta<-.01
    e1<-c(1,0,0)
    e2<-c(0,1,0)
    e3<-c(0,0,1)
    #fake partials
    J11<-function(theta){( f1(theta + delta*e1) - f1(theta) )/delta }
    J12<-function(theta){( f1(theta + delta*e2) - f1(theta) )/delta }
    J13<-function(theta){( f1(theta + delta*e3) - f1(theta) )/delta }
    J21<-function(theta){( f2(theta + delta*e1) - f2(theta) )/delta }
    J22<-function(theta){( f2(theta + delta*e2) - f2(theta) )/delta }
    J23<-function(theta){( f2(theta + delta*e3) - f2(theta) )/delta }
    J31<-function(theta){( f3(theta + delta*e1) - f3(theta) )/delta }
    J32<-function(theta){( f3(theta + delta*e2) - f3(theta) )/delta }
    J33<-function(theta){( f3(theta + delta*e3) - f3(theta) )/delta }
    jacob<-function(theta){matrix(c(J11(theta),J12(theta),J13(theta),
             J21(theta),J22(theta),J23(theta),J31(theta),J32(theta),J33(theta))
             ,ncol=3, byrow=T)}
    tol<-5
    TOL<-5
    theta<-initial
    ans<-c(b,theta,tol,TOL)
    cat(initial,"\n")
    while ( TOL>.1){
        #newton-raphson for estimating theta
        repeat { sln<-solve(jacob(theta),-TFF(theta))
                 theta<-theta + sln
                 tol<-max(abs(sln))
                 if (tol<.001){ break } }
        b<-B(theta)
        TOL<-max(abs(ans[3:5] - theta))
        ans<-c(b,theta,tol,TOL)
        cat(round(ans,3),"\n")
        anslist<-append(anslist,ans)
        }
}
```

Anovsig

```
#this program solves for sigma s squared using the analysis of variance method
cat("the random number is", .Random.seed , "\n")
cat("sigma p^2","sigma s^2", "\n")
for (loop in 1:100){
   source("simulatey")
   source("initvalues")
   #uses the guess for sigp^2 =mse
   yob<-matrix(c(TAB1),ncol=1)
   one<-matrix(rep(1,n),ncol=1)
   xs<-cbind(one,xi)
```

```
    b<-matrix(c(beta0,beta1),ncol=1)
    mu<-xs%*%b
    f<-function(r){r^2}
    SSMU<-sum(f(mu-mean(mu)))
    ybardd<-mean(yob)
    yidot<-0
    index<-1
    for(h in 1:n){ yidots<-sum(yob[index:(h*k)])
                   yidot<-append(yidot,yidots)
                   index<-h*k + 1 }
    yidot<-yidot[-1]
    yidotbar<-yidot/k
    SSA<-k*sum(f(yidotbar - mean(yob)))
    v<-SIG[1]
    #recall that guess[3] is calculated in initvalues
    est3<- guess[3]
    est4<- ( SSA- k*SSMU - (n-1)*est3  ) / ( n*k*v - (k/n)*(sum(SIG)) )
    cat(round(est3,3),round(est4,3),"\n")
}
```

Methmom

```
#this is a method of moments estimator of sigs^2.
#sigp^2 is known (the real value is used)
cat("the random seed is ", .Random.seed ,"\n")
for (loop in 1:100){
   source("simulatey")
   yob<-matrix(c(TAB1), ncol=1)
   v<-SIG[1]
   mu<-beta0 + beta1*xi
   f<-function(r){r^2}
   R<-0
   index<-1
   for (h in 1:n){ momi<- (1/k)*sum(f(yob[index:(h*k)]))
                   R<- append(R,momi)
                   index<- h*k + 1 }
   R<-R[-1]
   estimi<- (1/v)*( R - f(mu) - sigp^2)
   estim<-mean(estimi)
   cat(round(estim,3), "\n")
}
```

Simulatey

```
#this program generates the random variables yij
#one observation per site at each iteration
#for i=1,..n and j=1,..k. where n=number of sites and k=number of observations
#d=dimension of location vector
#the variance of esi=var(esi')= the variogram at infinity =covariogram at 0
#choose your covariogram models and their corresponding parameters.
# Set  choice=1 for 2 exponential cov models
# choice=2 for 2 power-type  cov models
k<-2
beta0<- 2
beta1<- 15
sigp<- sqrt(16)
sigs<- sqrt(49)
alpha<-.25
```

```
#sigp2<-sigp^2
#sigs2<-sigs^2
choice<-1
eda1<-1
eda2<-.2
a1<-.35
a2<-.09
loc<-read.table("/inst_area/cbocci/locations/location10",header=T)
locmat<-data.matrix(loc)
n<-dim(locmat)[1]
d<-dim(locmat)[2]
pred<-read.table("/inst_area/cbocci/predictors/predictor10",header=T)
xi<-data.matrix(pred)
if (choice==1) {source("/inst_area/cbocci/covmodels/covexp.prog")}
     else {source("/inst_area/cbocci/covmodels/covpow.prog")}
z<-rnorm(n,0,1)
mz<-matrix(z,ncol=1)
SIG <- alpha*cc1 + (1-alpha)*cc2
w<-chol(SIG)
#recall cc1 and cc2 come from covexp.prog
#the matrix w is such that w'w=SIG
#create spatial errors Es
Es<-t(w)%*%mz
#generating yij's
track<-1
RN<-rnorm(n*k,0,1)
mydata<-0
for(j in 1:k) {
        for(i in 1:n) {
                y<-beta0 + beta1*xi[i] + sigs*Es[i] + sigp*RN[track]
                track<-track + 1
                mydata<-append(mydata,y)  }
                }
Y<-data.matrix(mydata[-1])
#create a table of response at site i
TAB1<-matrix(data=Y,nrow=k,byrow=T, dimnames=list(paste("response",1:k),
      paste("site",1:n)))
TAB2<-matrix(data=Y,nrow=n,byrow=F,dimnames=list(paste("site",1:n),
      paste("response",1:k)))
```

Covexp.prog

```
#creates two covariograms of the exponential type
#makes C1
covc1<-0
for (i in 1:n){
       for (iprime in 1:n) {
            if (i<iprime){ h<-dist(locmat[c(i,iprime),1:d],metric="euclidean")
                           c1<-eda1*exp(-(a1^2)*h^2) }
                else if (i==iprime) {h<-0
                                     c1<-eda1 }
                else {h<-dist(locmat[c(iprime,i),1:d],metric="euclidean")
                      c1<-eda1*exp(-(a1^2)*h^2)  }
        rc1<-round(c1,6)
        covc1<-append(covc1,rc1) }
                }
cc1<-matrix(covc1[-1],ncol=n, byrow=T)
#makes C2
covc2<-0
for (i in 1:n){
```

```
        for (iprime in 1:n) {
             if (i<iprime){h2<-dist(locmat[c(i,iprime),1:d],metric="euclidean")
                           c2<-eda2*exp(-(a2^2)*h2^2) }
                else if (i==iprime) {h2<-0
                                     c2<-eda2 }
                else {h2<-dist(locmat[c(iprime,i),1:d], metric="euclidean")
                      c2<-eda2*exp(-(a2^2)*h2^2)}
        rc2<-round(c2,6)
        covc2<-append(covc2,rc2) }
}
cc2<-matrix(covc2[-1],ncol=n, byrow=T)
```

Initvalues

```
#this program gives a good guess for the initial values
#of the parameters beta0,beta1, and sigma p squared.
yob<-matrix(c(TAB1),ncol=1)
nooftimes<-rep(k,n)
XS<-matrix(rep(xi,nooftimes), ncol=1)
ordinary<-lm(yob~XS)
guess<- as.vector (c( ordinary$coefficients, deviance(ordinary)/
                    (ordinary$df)))
#this guess will be an initial estimate for beta0, beta1, and sigma p
#squared in that order
```

ExampleA

```
%#~/summer99/nasyvar
#This program calculates  the true asymptotic variance of
#all 5 parameters as n grows for a lattice with an
#increasing domain.  Example A.
k<-1
#k=number of observations per site
for (cnt in 2:12){
n<-cnt^2
#n=total number of sites
eda1<-1
eda2<-.2
a1<-.35
a2<-.09
sigp2<-16
sigs2<-49
alpha<-.25
loc<-0
for (j in 1:cnt-1){
     for (i in 1:cnt-1) {
         site<-c(i,j)
         loc<-append(loc,site) }
    }
locmat<-matrix(loc[-1],nrow=n, byrow=T)
d<-dim(locmat)[2]
source("/inst_area/cbocci/covmodels/covexp.prog")
SIG<-alpha*cc1 + (1-alpha)*cc2
U<-matrix(rep(1,k^2),ncol=k)
A<-kronecker(SIG,U)
Id<-diag(n*k)
V<-sigp2*Id + sigs2*A
IV<-solve(V)
```

```
############### asy var of beta0 and beta1#########
source("/inst_area/cbocci/predictors/predictor.lattice")
BB<-solve(t(X)%*%IV%*%X)
##############asy var of last 3 paramters ########
R<-kronecker(cc1-cc2,U)
t11<-sum(eigen(IV%*%IV)$values)
t12<-sum(eigen(IV%*%IV%*%A)$values)
t13<-sigs2*sum(eigen(IV%*%IV%*%R)$values)
t22<-sum(eigen(IV%*%A%*%IV%*%A)$values)
t23<-sigs2*sum(eigen(IV%*%A%*%IV%*%R)$values)
t33<-sigs2^2*sum(eigen(IV%*%R%*%IV%*%R)$values)
TT<-matrix(c(t11,t12,t13,t12,t22,t23,t13,t23,t33),ncol=3,byrow=T)
ITT<-2*solve(TT)
cat("n=", n,"\n")
print(BB)
print(ITT)
########check consistency condition for asy var of sigma s^2#######
source("consistency")
}
```

ExampleB

```
#This program calculates the true asymptotic variance of
#all 5 parameters as n grows on a unit square lattice
#for an infill sampling scheme
# Example B
k<-1
#k=number of observations per site
for (cnt in 2:12){
n<-cnt^2
#n=total number of sites
eda1<-1
eda2<-.2
a1<-.35
a2<-.09
sigp2<-16
sigs2<-49
alpha<-.25
deltax<-1/(cnt-1)
regr<-0
loc<-0
for (j in 1:cnt-1){
     for (i in 1:cnt-1) {
         site<-c( i*deltax,j*deltax)
         regrr<-site[2]
         loc<-append(loc,site)
         regr<-append(regr,regrr) }
    }
locmat<-matrix(loc[-1],nrow=n, byrow=T)
d<-dim(locmat)[2]
xi<-data.matrix(regr[-1])
#check that X'X^{-1} goes to zero.
one<-matrix(rep(1,n*k),ncol=1)
onek<-matrix(rep(1,k),ncol=1)
coll<-kronecker(xi,onek)
X<-cbind(one,coll)
cond<-solve(t(X)%*%X)
source("/inst_area/cbocci/covmodels/covexp.prog")
```

```
SIG<-alpha*cc1 + (1-alpha)*cc2
U<-matrix(rep(1,k^2),ncol=k)
A<-kronecker(SIG,U)
Id<-diag(n*k)
V<-sigp2*Id + sigs2*A
IV<-solve(V)
############## asy var of beta0 and beta1#########
BB<-solve(t(X)%*%IV%*%X)
#############asy var of last 3 paramters ########
R<-kronecker(cc1-cc2,U)
t11<-sum(eigen(IV%*%IV)$values)
t12<-sum(eigen(IV%*%IV%*%A)$values)
t13<-sigs2*sum(eigen(IV%*%IV%*%R)$values)
t22<-sum(eigen(IV%*%A%*%IV%*%A)$values)
t23<-sigs2*sum(eigen(IV%*%A%*%IV%*%R)$values)
t33<-sigs2^2*sum(eigen(IV%*%R%*%IV%*%R)$values)
TT<-matrix(c(t11,t12,t13,t12,t22,t23,t13,t23,t33),ncol=3,byrow=T)
ITT<-2*solve(TT)
cat("n=", n,"\n")
print(BB)
print(ITT)
########check consistency condition for asy var of sigma s^2#######
source("consistency")
}
```

ExampleC

```
#This program investigates the asy variance of
#all 5 parameters as n grows for a particular
#lattice. Modified increased sampling scheme
#Example C
k<-1
#k=number of observations per site
eda1<-1
eda2<-.2
a1<-.35
a2<-.09
sigp2<-16
sigs2<-49
alpha<-.25
for (cnt in 2:12){
n<-cnt^2
#n=total number of sites
grow<-cnt-1
loc<-0
for (j in seq(0,sqrt(cnt),by=sqrt(cnt)/grow)) {
    for (i in seq(0,sqrt(cnt),by=sqrt(cnt)/grow)) {
        site<-c(i,j)
        loc<-append(loc,site) }
    }
locmat<-matrix(loc[-1],nrow=n, byrow=T)
d<-dim(locmat)[2]
source("/inst_area/cbocci/covmodels/covexp.prog")
SIG<-alpha*cc1 + (1-alpha)*cc2
U<-matrix(rep(1,k^2),ncol=k)
A<-kronecker(SIG,U)
Id<-diag(n*k)
V<-sigp2*Id + sigs2*A
IV<-solve(V)
```

```
############### asy var of beta0 and beta1#########
source("/inst_area/cbocci/predictors/predictor.lattice")
BB<-solve(t(X)%*%IV%*%X)
##############asy var of last 3 paramters ########
R<-kronecker(cc1-cc2,U)
t11<-sum(eigen(IV%*%IV)$values)
t12<-sum(eigen(IV%*%IV%*%A)$values)
t13<-sigs2*sum(eigen(IV%*%IV%*%R)$values)
t22<-sum(eigen(IV%*%A%*%IV%*%A)$values)
t23<-sigs2*sum(eigen(IV%*%A%*%IV%*%R)$values)
t33<-sigs2^2*sum(eigen(IV%*%R%*%IV%*%R)$values)
TT<-matrix(c(t11,t12,t13,t12,t22,t23,t13,t23,t33),ncol=3,byrow=T)
ITT<-2*solve(TT)
cat("n=", n,"k=",k,"\n")
print(BB)
print(ITT)
########check consistency condition for asy var of sigma s^2#######
source("consistency")
}
```

Predictor.latticeA

```
#creates regressors in a systematic way such that
#X'X^{-1} goes to 0 as n goes to infinity
#on an increasing domain sampling scheme
#regressors in the same row of the lattice will have same value
#n,k known
grid<-sqrt(n)
regr<-0
lin<-1
for (l in 1:grid) {
  for (m in 1:grid) {
        regr<-append(regr,lin) }
  lin<-lin + 1 }
xi<-data.matrix(regr[-1])
#check that X'X^{-1} goes to zero.
one<-matrix(rep(1,n*k),ncol=1)
onek<-matrix(rep(1,k),ncol=1)
col1<-kronecker(xi,onek)
X<-cbind(one,col1)
cond<-solve(t(X)%*%X)
```

Consistency

```
#this program calculates the condition for consistency
# for the estimate of sigmas^2
f<-function(r){r^2}
v<-SIG[1,1]
gden<-2*(v-SIG)
denom<-f(sum(f(gden)))
numer<-0
for (i in 1:n){
   for (r in 1:n){
      for (l in 1:n){
         for (m in 1:n){
piece<-4*(v-SIG[i,r])*(v-SIG[l,m])*f(SIG[i,l]-SIG[i,m] -SIG[r,l] +SIG[r,m])
#cat(piece,"\n")
```

```
numer<-numer + piece }}}}
consist<-numer/denom
cat("for n=", n ,"consistency condition is ",consist,"\n")
```

Experiment

```
#given  (n)locations, optimizes the location
#of the  (n+1) by minimizing the asymptotic
# variance of the parameter estimates
library(Matrix)
n<-10
k<-1
# number of observations per site
sigp2<-16
sigs2<-49
alpha<-.25
eda1<-1
eda2<-.2
a1<-.35
a2<-.09
tr<-function(M){sum(diag(M))}
loc<-read.table("/inst_area/cbocci/locations/location10",header=T)
locmat<-data.matrix(loc)
locmat<-locmat[1:n,]
#n<-dim(locmat)[1]
d<-dim(locmat)[2]
pred<-read.table("/inst_area/cbocci/predictors/predictor10", header=T)
xi<-data.matrix(pred)
xi<-xi[1:n,]
source("/inst_area/cbocci/covmodels/covexp.prog")
SIG<-alpha*cc1 + (1-alpha)*cc2
v<-SIG[1,1]
#new site has 2 coordinates
scoord<-function(s1,s2){c(s1,s2)}
d1<-function(scoord){as.vector(t(scoord)%*%c(1,0))}
d2<-function(scoord){as.vector(t(scoord)%*%c(0,1))}
#fit for x11 using all ten sites
if (n==10) {addx<-function(scoord){132.164 + 45.568*d1(scoord)
    - 26.309*d2(scoord) -1.527*d1(scoord)^2 + 4.753*d2(scoord)^2
    + 3.166*d1(scoord)*d2(scoord) }}
#fit for x5 using 4 sites
if (n==4){addx<-function(scoord){274.966 + 36.38*d1(scoord)
   - 58.32*d2(scoord) - 7.341163*d1(scoord)^2}}
Xi<-function(scoord){matrix(c(xi,addx(scoord)))}
one<-matrix(rep(1,(n+1)*k),ncol=1)
nooftimes<-rep(k,n+1)
colofx<-function(scoord){matrix(rep(Xi(scoord),nooftimes),ncol=1)}
X<-function(scoord){cbind(one,colofx(scoord))}
locall<-function(scoord){rbind(scoord,locmat)}
ds11<-function(scoord){dist(locall(scoord))[1:n]}
ds112<-function(scoord){ds11(scoord)^2}
vec1<-function(scoord){exp(-(.35)^2 * ds112(scoord))}
vec2<-function(scoord){.2*exp(-(.09)^2 *ds112(scoord))}
vec<-function(scoord){alpha*vec1(scoord) + (1-alpha)*vec2(scoord)}
NEWSIG<-function(scoord){ cbind(rbind(SIG,vec(scoord)),c(vec(scoord),v))}
U<-matrix(rep(1,k^2),ncol=k)
A<-function(scoord){kronecker(NEWSIG(scoord),U)}
I<-diag((n+1)*k)
V<-function(scoord){sigp2*I + sigs2*A(scoord) }
IV<-function(scoord){solve(V(scoord))}
```

```
Mb<- function(scoord){ t(X(scoord))%*% IV(scoord) %*% X(scoord)}
IMb<-function(scoord){solve(Mb(scoord))}
#NIMb<-function(scoord){norm.Matrix(IMb(scoord),type="F")}
NEWc1<-function(scoord){cbind(rbind(cc1,vec1(scoord)),c(vec1(scoord),v)) }
NEWc2<-function(scoord){cbind(rbind(cc2,vec2(scoord)),c(vec2(scoord),v)) }
DVDa<-function(scoord){sigs2*kronecker( NEWc1(scoord)-NEWc2(scoord),U)}
t11<-function(scoord){tr(IV(scoord) %*% IV(scoord) )}
t12<-function(scoord){tr(IV(scoord) %*% IV(scoord) %*% A(scoord) )}
t13<-function(scoord){tr(IV(scoord) %*% IV(scoord) %*% DVDa(scoord))}
t22<-function(scoord){tr(IV(scoord) %*% A(scoord)  %*% IV(scoord)
     %*% A(scoord) )}
t23<-function(scoord){tr(IV(scoord) %*% A(scoord)  %*% IV(scoord)
     %*% DVDa(scoord) )}
t33<-function(scoord){tr(IV(scoord) %*% DVDa(scoord)%*%IV(scoord)
     %*% DVDa(scoord) )}
Mtheta<-function(scoord){.5*matrix(c(t11(scoord),t12(scoord),t13(scoord),
        t12(scoord),t22(scoord),t23(scoord),t13(scoord),t23(scoord),
        t33(scoord)),ncol=3)}
IMtheta<-function(scoord){solve.Matrix(Mtheta(scoord))}
############################
#Construct asymptotic variance and norm of it
o23<-matrix(rep(0,6),ncol=3)
o32<-matrix(rep(0,6),ncol=2)
top<-function(scoord){cbind(IMb(scoord),o23)}
bot<-function(scoord){cbind(o32,IMtheta(scoord))}
asyvar<-function(scoord){rbind(top(scoord),bot(scoord))}
normasyvar<-function(scoord){norm.Matrix(asyvar(scoord),type="F")}
######################
#grid search to find minimum
grid<-seq(-20,20,by=3)
nlist<-0
for (i in grid) {
  for (j in grid) { acc<-normasyvar(c(i,j))
    cat(acc,"\n")
    nlist<-append(nlist,acc)
  }}
nlist<-nlist[-1]
VALUES<-matrix(c(nlist),ncol=length(grid),byrow=T)
dimnames(VALUES)<-list(grid,grid)
print(VALUES)
#persp(grid,grid,VALUES)
```

Appendix D

Maple Programs

Maple is a symbolic mathematics package produced by Waterloo Maple Software and the University of Waterloo. Programs described here were written using Maple V, release 3, on an IBM model 390 RS/6000 computer.

Program Name	Function
Inverses, section 3.2	Calculates the inverse and approximate inverses of V^{-1}
Asyvar2, section 4.2.1	Investigates the asymptotic variance of β_o, β_1 as n_p goes to infinity
Short2 section 4.2.1	Same as Asyvar2 but uses derived expressions to speed up calculations
Asyvar3 section 4.2.1	Investigates the asymptotic variance of σ_p^2, σ_s^2, and α as n_p goes to infinity
Short3 section 4.2.1	Same as Asyvar3 but uses derived expressions to speed up calculations

Inverses

```
#This program creates covariograms 1 and 2 of the exponential type.
# n=4  sites k=2 obs/site
#calculates various inverses
#with(linalg):
#with(student):
n:=4;
k:=2;
site1:=vector([-4,0]):
site2:=vector([0,3]):
site3:=vector([2,-2]):
site4:=vector([5,1]):
matsite:=stack(site1,site2,site3,site4);
eda1:=1;
eda2:=.2;
pow1:=.35:
pow2:=.09:
```

```
Digits := 4:
for i from 1 to n do
      for ipr from 1 to n do
              h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
              cov1(i,ipr):= evalf(eda1*exp(-(pow1^2)*(h(i,ipr))^2));
      od;
od;
c1:=matrix(n,n,(r,s)->cov1(r,s));
#print(c1);
for i from 1 to n do
      for ipr from 1 to n do
              h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
              cov2(i,ipr):= evalf(eda2*exp(-(pow2^2)*(h(i,ipr))^2));
      od;
od;
c2:=matrix(n,n,(l,m)->cov2(l,m));
#print(c2);
alpha:=.25:
sigma:=sqrt(16):
rho:=sqrt(10):
delta:=rho^2/sigma^2:
Sigma:= evalm(alpha*c1 + (1-alpha)*c2);
ident:=band([1],n*k):
U:=matrix(k,k,(u,v)->1);
############calculate the direct product manually A=Sigma x U
Z:=band([0], n*k):
r:=1:
w:=0:
for i from 1 to n do
  s:=1:
  t:=0:
      for ipr from 1 to n do
         if Sigma[i,ipr]=0 then
             correct:=band([0],k);
             A:=copyinto(correct,Z,r,s);
         else
             AA(i,ipr):=evalm(Sigma[i,ipr] * U):
             A:=copyinto(AA(i,ipr),Z,r,s):
          fi;
        t:=t+1;
        s:=t*k+1:
      od:
  w:=w+1:
  r:=w*k+1:
od:
#print(A);
################calculate direct product T=IXU
identn:=band([1],n):
Z:=band([0], n*k):
r:=1:
w:=0:
for i from 1 to n do
  s:=1:
  t:=0:
       for ipr from 1 to n do
         if identn[i,ipr]=0 then
            correct:=band([0],k);
            T:=copyinto(correct,Z,r,s);
         else
            TT(i,ipr):=evalm(identn[i,ipr] * U):
            T:=copyinto(TT(i,ipr),Z,r,s):
         fi;
        t:=t+1;
        s:=t*k+1:
       od:
```

```
  w:=w+1:
  r:=w*k+1:
od:
#print(T);
################Inverses###########
inv:=inverse(V);
INV1:=evalm((1/sigma^2)*(ident - (rho^2/sigma^2) * A)) ;
INV2:=evalm((1/sigma^2)*(ident - (rho^2/sigma^2) * A
      + (rho^2/sigma^2)^2 * A^2));
INV3:=evalm((1/sigma^2)*(3*IB - 3*IB&*IB&*(ident + delta*A)
       + IB&*IB&*IB&*(ident + delta*A)^2));
```

Asyvar2

```
#This program investigates the asymptotic variance in k
#of beta0 and beta1
#using V^{-1}=I + PxU and the numerical example
#n=10,  k increasingly large
st:=time():
with(linalg):
with(student):
Digits := 7:
alpha:=.25:
sigp2:=16:
sigs2:=49:
delta:=sigs2/sigp2:
#locations
site1:=vector([-4,0]):
site2:=vector([0,3]):
site3:=vector([2,-2]):
site4:=vector([5,1]):
site5:=vector([4,2]):
site6:=vector([3,6]):
site7:=vector([-1,1]):
site8:=vector([1,1]):
site9:=vector([-2,-3]):
site10:=vector([1,6]):
#predictor variables
x[1]:=12:
x[2]:=100:
x[3]:=435:
x[4]:=215:
x[5]:=309:
x[6]:=400:
x[7]:=72:
x[8]:=75:
x[9]:=59:
x[10]:=17:
matsite:=stack(site1,site2,site3,site4,site5,site6,site7,site8,site9,site10):
pred:=matrix(10,1,(1,m)->x[1]):
n:=10:
#creates covariograms 1 and 2 of the exponential type.
eda1:=1:
eda2:=.2:
pow1:=.35:
pow2:=.09:
for i from 1 to n do
      for ipr from 1 to n do
          h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
          cov1(i,ipr):= evalf(eda1*exp(-(pow1^2)*(h(i,ipr))^2));
      od;
```

```
od;
c1:=matrix(n,n,(r,s)->cov1(r,s)):
for i from 1 to n do
    for ipr from 1 to n do
        h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
        cov2(i,ipr):= evalf(eda2*exp(-(pow2^2)*(h(i,ipr))^2));
    od;
od;
c2:=matrix(n,n,(l,m)->cov2(l,m)):
Sigma:= evalm(alpha*c1 + (1-alpha)*c2):
VAR:=vector(['k','varbo','varb1']):
#########################################################
for k from 1 to 10 do
#calculate L=pred x one
one:=matrix(k,1,(l,m)->1):
Z:=matrix(n*k,1,(l,m)->0):
r:=1:
w:=0:
for i from 1 to n do
   if pred[i,1]=0 then
     correct:=matrix(k,1,(l,m)->0);
     L:=copyinto(correct,Z,r,1);
   else
     LL:=evalm(pred[i,1] * one);
     L:=copyinto(LL,Z,r,1);
    fi;
    w:=w+1;
    r:=w*k + 1;
od:
ones:=matrix(n*k,1,(l,m)->1):
X:=augment(ones,L):
#calculate P
ident:=band([1],n):
P:=evalm( (1/k)*( inverse(ident + delta*k*Sigma) - ident )):
U:=matrix(k,k,(u,v)->1):
#calculate G=PxU
Z:=band([0], n*k):
r:=1:
w:=0:
for i from 1 to n do
  s:=1:
  t:=0:
        for ipr from 1 to n do
         if P[i,ipr]=0 then
           correct:= band([0],k);
           G:=copyinto(correct,Z,r,s);
         else
           GG:=evalm(P[i,ipr] * U);
           G:=copyinto(GG,Z,r,s);
         fi;
         t:=t + 1:
         s:=t*k +1:
        od:
  w:=w+1:
  r:=w*k+1:
od:
II:=band([1],(n*k)):
IV:=evalm( (1/sigp2)*(II + G)):
M:=evalm(transpose(X) &* IV &* X):
ASY:=inverse(M):
varbo:=ASY[1,1]:
varb1:=ASY[2,2]:
v:=vector([k,varbo,varb1]):
VAR:=stack(VAR,v):
od:
```

```
print(VAR);
time() - st;
```

Short2

```
#This program investigates the asymptotic variance in k
#of beta0 and beta1
#using the Taylor Series expansion for a numerical example
#n=10,  k increasingly large
st:=time():
with(linalg):
with(student):
Digits := 7:
alpha:=.25:
sigp2:=16:
sigs2:=49:
delta:=sigs2/sigp2:
#locations
site1:=vector([-4,0]):
site2:=vector([0,3]):
site3:=vector([2,-2]):
site4:=vector([5,1]):
site5:=vector([4,2]):
site6:=vector([3,6]):
site7:=vector([-1,1]):
site8:=vector([1,1]):
site9:=vector([-2,-3]):
site10:=vector([1,6]):
#predictor variables
x[1]:=12:
x[2]:=100:
x[3]:=435:
x[4]:=215:
x[5]:=309:
x[6]:=400:
x[7]:=72:
x[8]:=75:
x[9]:=59:
x[10]:=17:
matsite:=stack(site1,site2,site3,site4,site5,site6,site7,site8,site9,site10):
n:=10:
#creates covariograms 1 and 2 of the exponential type.
eda1:=1:
eda2:=.2:
pow1:=.35:
pow2:=.09:
for i from 1 to n do
    for ipr from 1 to n do
        h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
        cov1(i,ipr):= evalf(eda1*exp(-(pow1^2)*(h(i,ipr))^2));
    od;
od;
c1:=matrix(n,n,(r,s)->cov1(r,s)):
for i from 1 to n do
    for ipr from 1 to n do
        h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
        cov2(i,ipr):= evalf(eda2*exp(-(pow2^2)*(h(i,ipr))^2));
    od;
od;
c2:=matrix(n,n,(l,m)->cov2(l,m)):
Sigma:= evalm(alpha*c1 + (1-alpha)*c2):
```

```
DSDa:=evalm(c1-c2):
VAR:=vector(['k','varbo','varb1']):
#########################################################
for k from 15 to 50 by 5  do
#calculate P
ident:=band([1],n):
P:=evalm( (1/k)*( inverse(ident + delta*k*Sigma) - ident )):
sum1:=0:
for i from 1 to n do
    for r from 1 to n do
        ps1:=x[i]*x[r]*P[i,r]:
        sum1:=sum1 + ps1:
    od:
od:
sum2:=0:
for i from 1 to n do
    for r from 1 to n do
        ps2:=x[i]*P[i,r]:
        sum2:=sum2 + ps2:
    od:
od:
sum3:=0:
for i from 1 to n do
    for r from 1 to n do
        ps3:=P[i,r]:
        sum3:=sum3 + ps3:
    od:
od:
sum4:=0:
for i from 1 to n do
      ps4:=x[i]^2:
      sum4:=sum4 + ps4:
od:
sum5:=0:
for i from 1 to n do
      ps5:=x[i]:
      sum5:=sum5 + ps5:
od:
num11:=sigp2*( k*sum4 + k^2 * sum1):
num22:=sigp2*(n*k + k^2*sum3):
DEN:=(n*k + k^2*sum3)*(k*sum4 + k^2*sum1) - (k*sum5 + k^2*sum2)^2:
varbo:=num11/DEN:
varb1:=num22/DEN:
v:=vector([k,varbo,varb1]):
VAR:=stack(VAR,v):
od:
print(VAR):
time() - st;
```

Asyvar3

```
#This program investigates the asymptotic variance in k
# of sigp2, sigs2, and alpha
# using the V^{-1}=I + PxU and the numerical example
#n=10,  k increasingly large
st:=time():
#gc(0);
#with(linalg):
#with(student):
Digits := 20:
alpha:=.25:
sigp2:=16:
sigs2:=49:
```

```
delta:=sigs2/sigp2:
#locations
site1:=vector([-4,0]):
site2:=vector([0,3]):
site3:=vector([2,-2]):
site4:=vector([5,1]):
site5:=vector([4,2]):
site6:=vector([3,6]):
site7:=vector([-1,1]):
site8:=vector([1,1]):
site9:=vector([-2,-3]):
site10:=vector([1,6]):
matsite:=stack(site1,site2,site3,site4,site5,site6,site7,site8,site9,site10):
n:=10;
#creates covariograms 1 and 2 of the exponential type.
eda1:=1:
eda2:=.2:
pow1:=.35:
pow2:=.09:
for i from 1 to n do
      for ipr from 1 to n do
              h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
              cov1(i,ipr):= evalf(eda1*exp(-(pow1^2)*(h(i,ipr))^2));
              od;
od;
c1:=matrix(n,n,(r,s)->cov1(r,s)):
for i from 1 to n do
      for ipr from 1 to n do
              h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
              cov2(i,ipr):= evalf(eda2*exp(-(pow2^2)*(h(i,ipr))^2));
              od;
od;
c2:=matrix(n,n,(l,m)->cov2(l,m)):
Sigma:= evalm(alpha*c1 + (1-alpha)*c2):
M:=vector(['k','t11','t12','t13','t22','t23','t33']):
VAR:=vector(['k','sigp2','sigs2','alpha']):
############################################################
for k from 1 to 10 do
U:=matrix(k,k,(u,v)->1):
#calculate the direct product manually A=Sigma X U
Z:=band([0], n*k):
r:=1:
w:=0:
for i from 1 to n do
    s:=1:
    t:=0:
        for ipr from 1 to n do
         if Sigma[i,ipr]=0 then
            correct:= band([0],k);
            A:=copyinto(correct,Z,r,s);
 else
            AA:=evalm(Sigma[i,ipr] * U):
            A:=copyinto(AA,Z,r,s):
         fi;
         t:=t + 1;
         s:=t*k+1:
        od:
  w:=w + 1:
  r:=w*k+1:
od:
II:=band([1],(n*k)):
V:=evalm(sigp2*II + sigs2*A):
#derivatives needed in tml expressions
#DVD1:= II:
DVD2:=evalm(A):
```

```
DVD3:=evalm(sigs2* R):
#####################################################
#calculate direct product R=DSigma/Dalpha X U
DSDa:=evalm(c1-c2):
Z:=band([0], n*k):
r:=1:
w:=0:
for i from 1 to n do
  s:=1:
  t:=0:
        for ipr from 1 to n do
         if DSDa[i,ipr]=0 then
           correct:= band([0],k);
           R:=copyinto(correct,Z,r,s);
         else
           RR:=evalm(DSDa[i,ipr] * U);
           R:=copyinto(RR,Z,r,s);
        fi;
        t:=t+1:
        s:=t*k + 1:
       od:
  w:=w + 1:
  r:=w*k + 1:
od:
#########################################################
#calculate P
ident:=band([1],n):
P:=evalm( (1/k)*( inverse(ident + delta*k*Sigma) - ident )):
#calculate the direct product manually G=P X U
Z:=band([0], n*k):
r:=1:
w:=0:
for i from 1 to n do
  s:=1:
  t:=0:
        for ipr from 1 to n do
         if P[i,ipr]=0 then
         correct:= band([0],k);
         G:=copyinto(correct,Z,r,s);
        else
         GG:=evalm(P[i,ipr] * U);
         G:=copyinto(GG,Z,r,s);
        fi;
        t:=t + 1;
        s:=t*k +1:
       od:
  w:=w+1:
  r:=w*k+1:
od:
#########################################################
IV:=evalm( (1/sigp2)*(II + G)):
t11:=trace(IV &* IV):
t12:=trace(IV &* IV &* DVD2):
t13:=trace(IV &* IV &* DVD3):
t22:=trace(IV &* DVD2 &* IV &* DVD2):
t23:=trace(IV &* DVD2 &* IV &* DVD3):
t33:=trace(IV &* DVD3 &* IV &* DVD3):
m[k]:=vector([k,t11,t12,t13,t22,t23,t33]):
M:=stack(M,m[k]):
linfo:=evalm(matrix(3,3,[t11,t12,t13,t12,t22,t23,t13,t23,t33])):
if cond(linfo)<500000 then
   asy:=inverse(linfo);
   for i from 1 to 3 do
     theta[i]:=2*asy[i,i];
   od;
   v[k]:=vector([k,theta[1],theta[2],theta[3]]);
   VAR:=stack(VAR,v[k]);
```

```
else
   print(k,'condition number=', cond(linfo));
fi;
od:
print(M);
print(VAR);
time()-st;
```

Short3

```
#This program investigates the asymptotic variance in k
#of sigp2,sigs2, and alpha
#using V^{-1}=I + PxU and the numerical example
#n=10,  k increasingly large
st:=time():
#gc(0);
#with(linalg):
#with(student):
Digits := 20:
alpha:=.25:
sigp2:=16:
sigs2:=49:
delta:=sigs2/sigp2:
#locations
site1:=vector([-4,0]):
site2:=vector([0,3]):
site3:=vector([2,-2]):
site4:=vector([5,1]):
site5:=vector([4,2]):
site6:=vector([3,6]):
site7:=vector([-1,1]):
site8:=vector([1,1]):
site9:=vector([-2,-3]):
site10:=vector([1,6]):
matsite:=stack(site1,site2,site3,site4,site5,site6,site7,site8,site9,site10):
n:=10;
#creates covariograms 1 and 2 of the exponential type.
eda1:=1:
eda2:=.2:
pow1:=.35:
pow2:=.09:
for i from 1 to n do
    for ipr from 1 to n do
        h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
        cov1(i,ipr):= evalf(eda1*exp(-(pow1^2)*(h(i,ipr))^2));
    od;
od;
c1:=matrix(n,n,(r,s)->cov1(r,s)):
for i from 1 to n do
    for ipr from 1 to n do
        h(i,ipr):= distance(row(matsite,i),row(matsite,ipr));
        cov2(i,ipr):= evalf(eda2*exp(-(pow2^2)*(h(i,ipr))^2));
    od;
od;
c2:=matrix(n,n,(l,m)->cov2(l,m)):
Sigma:= evalm(alpha*c1 + (1-alpha)*c2):
MM:=vector(['k','t11','t12','t13','t22','t23','t33']):
VARR:=vector(['k','sigp2','sigs2','alpha']):
##################################################
for k from 15 to 50 by 5 do
#calculate P
```

```
ident:=band([1],n):
P:=evalm( (1/k)*( inverse(ident + delta*k*Sigma) - ident )):
DSDa:=evalm(c1-c2):
tt11:=(1/sigp2)^2*(n*k + 2*k*trace(P) + k^2 * trace(P&*P)):
tt12:=(1/sigp2)^2 * (k*trace(Sigma) + 2*k^2*trace(P&*Sigma)
      + k^3*trace(P^2&*Sigma)):
tt13:=(1/sigp2)^2 * sigs2 * (k*trace(DSDa) + 2*k^2 * trace(P&*DSDa)
      + k^3* trace(P^2 &* DSDa)):
tt22:=(1/sigp2)^2 * (k^2 * trace(Sigma^2) + 2*k^3 * trace(Sigma^2 &* P)
      + k^4 * trace(P&*Sigma &*P&*Sigma)):
tt23:=(sigs2/sigp2^2) * (k^2 * trace(Sigma&*DSDa) + k^3*trace(Sigma&*P&*DSDa)
      + k^3*trace(P&*Sigma&*DSDa) + k^4*trace(P&*Sigma&*P&*DSDa)):
tt33:=delta^2 *(k^2*trace(DSDa&*DSDa) + 2*k^3*trace(DSDa&*DSDa&*P)
      + k^4*trace(P&*DSDa&*P&*DSDa)):
mm[k]:=vector([k,tt11,tt12,tt13,tt22,tt23,tt33]):
MM:=stack(MM,mm[k]):
DEN:=tt11*tt22*tt33 - tt11*(tt23)^2 - (tt12)^2*tt33 + 2*tt12*tt13*tt23
     - (tt13)^2 * tt22:
llinfo:=matrix(3,3,[tt11,tt12,tt13,tt12,tt22,tt23,tt13,tt23,tt33]):
check:=cond(llinfo):
if check<500000 then
 varsigp2:=2*(tt22*tt33 - (tt23)^2)/DEN:
 varsigs2:=2*(tt11*tt33 - (tt13)^2)/DEN:
 varalpha:=2*(tt11*tt22 - (tt12)^2)/DEN:
 VARR:=stack(VARR,vector([k,varsigp2,varsigs2,varalpha]));
else
print(k,'condition number=',check);
fi;
od:
print(MM);
print(VARR);
time()-st);
```

Asyvar3.theory

```
#This program investigates the asymptotic variance in k
#theoretically by symbolic manipulation
# notation: DSDa represents the partial derivative of Sigma wrt alpha
with(linalg):
#use the following asymptotic approx
tr(P):=A/(k^2*delta) - n/k;
tr(P^2):=n/k^2 + 2*A/(delta*k^3) + B/(k^4*delta^2);
tr(Sigma):=C;
tr(PSigma):=n/(k^2*delta) - C/k;
tr(P^2*Sigma):=A/(k^4*delta^2) - 2*n/(delta*k^3) + C/k^2;
tr(Sigma^2):=D;
tr(Sigma^2*P):=C/(k^2*delta) - D/k;
tr(PSigmaPSigma):=n/(k^4*delta^2) - 2*C/(k^3*delta) + D/k^2;
tr(DSDa):=E;
tr(SigmaDSDa):=F;
tr(PDSDa):=G/(delta*k^2) - E/k;
tr(P^2*DSDa):=H/(k^4*delta^2) - 2*G/(k^3*delta) + E/k^2;
tr(SigmaPDSDa):=E/(k^2*delta) - F/k;
tr(PSigmaDSDa):=E/(k^2*delta) - F/k;
tr(PSigmaPDSDa):=G/(k^4*delta^2) - 2*E/(k^3*delta) + F/k^2;
tr(DSDaDSDa):=J;
tr(DSDaDSDaP):=L/(k^2*delta) - J/k;
tr(PDSDaPDSDa):=M/(k^4*delta^2) -2*L/(k^3*delta) + J/k^2;
```

```
#t_ml expressions
t11:=sigma[p]^(-4)*(n*k + 2*k*tr(P) + k^2 * tr(P^2));
t12:=sigma[p]^(-4) * (k*tr(Sigma) + 2*k^2*tr(PSigma)
      + k^3*tr(P^2*Sigma));
t13:=sigma[p]^(-4) * sigma[s]^2 * (k*tr(DSDa) + 2*k^2 * tr(PDSDa)
      + k^3* tr(P^2*DSDa));
t22:=sigma[p]^(-4)* (k^2 * tr(Sigma^2) + 2*k^3 * tr(Sigma^2*P)
      + k^4 * tr(PSigmaPSigma));
t23:=sigma[p]^(-4)  * sigma[s]^2*(k^2 * tr(SigmaDSDa) + k^3*tr(SigmaPDSDa)
      + k^3*tr(PSigmaDSDa) + k^4*tr(PSigmaPDSDa));
t33:=sigma[p]^(-4)*sigma[s]^4 *(k^2*tr(DSDaDSDa) + 2*k^3*tr(DSDaDSDaP)
      + k^4*tr(PDSDaPDSDa));
den:=t11*t22*t33 - t11*t23^2 - t12^2*t33 + 2*t12*t13*t23 - t13^2*t22:
first:=2*(t22*t33 - t23^2)/den:
second:=2*(t11*t33 - t13^2)/den:
third:=2*(t11*t22 - t12^2)/den:
varsigp2:=collect(first,k);
varsigs2:=collect(second,k);
varalpha:=collect(third,k);
```

Bibliography

[1] Bickel, Peter J., and Doksum, Kjell A. (1977), *Mathematical Statistics: Basic Ideas and Seleted Topics*, Holden-Day, Inc. Oakland, California.

[2] Billingsley, Patrick (1986), *Probability and Measure, 2nd Edition*, John Wiley & Sons, New York.

[3] Brown, P.J., Le, N.D., and Zidek, J.V. (1994), Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics* **22**(4), 489-509.

[4] Chiu, Tom Y.M., Leonard, Tom and Tsui, Kam-Wah (1996), The matrix-logarithmic covariance model. *Journal of the American Statistical Association* **91**, 198-210.

[5] Cook, D.G. and Pocock, S.J. (1983), Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. *Biometrics* **39**, 361-371.

[6] Cressie, Noel A.C. (1993), *Statistics for Spatial Data, Revised Edition*, John Wiley & Sons, New York.

[7] Cressie, N. and Hawkins, D.M. (1980), Robust estimation if the variogram, I. *Journal of the International Association for Mathematical Geology* **12**, 135-146.

[8] Davis, John C.(1986), *Statistics and Data Analysis in Geology, 2nd Edition*, John Wiley & Sons, New York.

[9] Genton, Marc (1998), Highly robust variogram estimation, *Mathematical Geology* **30**(2), 213-221.

[10] Golub, Gene H. and Van Loan, Charles F. (1996), *Matrix Computations, 3rd edition*, John Hopkins University Press, Baltimore, Maryland.

[11] Gomez, C. Salvy, B., and Zimmermann, P. (1995), *Calcul formel : mode d'emploi exemples en Maple*, Masson, Paris.

[12] Graham, Alexander (1987), *Nonnegative Matrices and Applicable Topics in Linear Algebra*, Ellis Horwood Limited, Chichester.

[13] Graybill, Franklin A. (1983), *Matrices with Applications in Statistics, 2nd Edition*, Wadsworth Inc., Belmont, California.

[14] Guttorp, Peter (1997), Analysis of Spokane CO Data, *National Research Center for Statistics and the Environment*, Technical Report Series.

[15] Harville, David A. (1997), *Matrix Algebra From a Statistician's Perspective*, Springer-Verlag, New York.

[16] Hocking, Ronald R. (1985), *The Analysis of Linear Models*, Wadsworth Inc., Belmont, California.

[17] Hoffman, Kenneth. (1975), *Analysis in Euclidean Space*, Prentice Hall, Englewook Cliffs, New Jersey.

[18] Johnson, Richard A. and Wichern, Dean W. (1988), *Applied Multivariate Statistical Analysis, 2nd Edition*, Prentice Hall, Englewood Cliffs, New Jersey.

[19] Journel, A.G. and Huijbregts, C.J. (1978), *Mining Geostatistics*, Academic Press, New York.

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

[21] Matheron, G. (1962), Traite de Geostatistique appliquee, Tome I. *Memoires du Bureau de Recerches Geologiques et Minieres* **14**, Edition Technip, Paris.

[22] Sampson, P.D. and Guttorp, P. (1992), Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* **87**, 108-119.

[23] Schott, James R. (1997), *Matrix Analysis for Statistics*, John Wiley & Sons, New York.

[24] Searle, Shayle R. (1971), *Linear Models*, John Wiley & Sons, New York.

[25] Searle, Shayle R. (1982), *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York.

[26] Smith, Richard L. (1996), Estimating nonstationary spatial correlations, Preprint 1-25.

[27] Thisted, Ronald A. (1988), *Elements of Statistical Computing*, Chapman and Hall, New York.

[28] Venables, W.N. and Ripley, B.D. (1997), *Modern Applied Statistics with S-Plus, 2nd Edition*, Springer-Verlag, New York.