RESAMPLING METHODOLOGY IN SPATIAL PREDICTION AND REPEATED MEASURES TIME SERIES

A Dissertation

by

KRISTA DIANNE RISTER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2010

Major Subject: Statistics

RESAMPLING METHODOLOGY IN SPATIAL PREDICTION AND REPEATED MEASURES TIME SERIES

A Dissertation

by

KRISTA DIANNE RISTER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Soumendra N. Lahiri
Committee Members,	Michael Sherman
	Thomas Wehrly
	David Larson
Head of Department,	Simon Sheather

December 2010

Major Subject: Statistics

ABSTRACT

Resampling Methodology in Spatial Prediction and Repeated Measures Time Series.

(December 2010)

Krista Dianne Rister, B.S., Texas A&M University;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Soumendra N. Lahiri

In recent years, the application of resampling methods to dependent data, such as time series or spatial data, has been a growing field in the study of statistics. In this dissertation, we discuss two such applications.

In spatial statistics, the reliability of Kriging prediction methods relies on the observations coming from an underlying Gaussian process. When the observed data set is not from a multivariate Gaussian distribution, but rather is a transformation of Gaussian data, Kriging methods can produce biased predictions. Bootstrap resampling methods present a potential bias correction. We propose a parametric bootstrap methodology for the calculation of either a multiplicative or additive bias correction factor when dealing with Trans-Gaussian data. Furthermore, we investigate the asymptotic properties of the new bootstrap based predictors. Finally, we present the results for both simulated and real world data.

In time series analysis, the estimation of covariance parameters is often of utmost importance. Furthermore, the understanding of the distributional behavior of parameter estimates, particularly the variance, is useful but often difficult. Block bootstrap methods have been particularly useful in such analyses. We introduce a

new procedure for the estimation of covariance parameters for replicated time series data.

To my brother

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Soumendra N. Lahiri, for his support and guidance throughout the research process. Working with and learning from Dr. Lahiri has truly been an honor.

I would also like to thank my committee members, Dr. Michael Sherman, Dr. Thomas Wehrly, and Dr. David Larson, for their continued support and contributions they have made to my education in the fields of statistics and mathematics. They have all challenged me with difficult and exciting projects, which made this dissertation possible.

Further thanks are due to Kedar Patel, of SanDisk Inc. and the University of California at Berkeley, for his help answering questions about line width roughness parameters, and Gary McManus, of the Oklahoma Climatological Survey at the University of Oklahoma, for answering questions about the use of data from Mesonet.

I am eternally grateful for the love and support of my family, particularly my parents, Edward and Dianne Rister, and my brother, Karl Rister. Their faith in me has been an inspiration and blessing, and has carried me every step of the way.

Finally, I must thank the employees and students of the Department of Statistics at Texas A&M University. Knowing and working with them has been a privilege that I will never forget.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Ever since its introduction by Efron (1979), the bootstrap has had a growing impact on statistical analysis. After Singh (1981) showed the bootstrap failed when the independence assumption on the random variables generating the data was lost, the work of Hall (1985), Künsch (1989) and Liu and Singh (1992), among others, adapted the bootstrap for work with correlated data, allowing for the use of the bootstrap in the temporal and spatial domains.

In this dissertation, we propose new bootstrap methods for bias correction in the area of spatial prediction, or Kriging, and for the estimation of covariance parameters for repeated measures time series. These projects incorporate both parametric and nonparametric bootstrap methods.

A. Trans-Gaussian Kriging

The need for spatial prediction methods comes from various academic disciplines, such as geology, meteorology, and epidemiology. Spatial prediction utilizes observed values at certain locations on a grid or map to predict an unobserved value at one or more other locations. For example, a meteorologist may use temperature readings at a number of spread out weather stations to predict the temperature at a nearby point of interest without an observation station. Kriging is one such type of spatial prediction.

The journal model is *Journal of the American Statistical Association.*

1. Background of Kriging

Kriging is named for D. G. Krige, who developed empirical methods for spatial prediction with mining applications in his master's thesis (Krige 1951). The methods now known as Kriging were developed later by Matheron (1962) and others. For more on the origins and development of Kriging, both the name and the methods, see Cressie (1990). Also, for now we remark that further reading into Kriging methodology may come from a spatial statistics textbook such as Cressie (1993) or Schabenberger and Gotway (2005).

2. Basics of Kriging

Kriging relies on the second-order properties of a spatial process in order to make prediction possible. The assumptions placed on the process differ for various types of Kriging, but a general framework is as follows.

Let $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_n$ be locations on some spatial domain D. For example, if working in the two dimensional space, $D = \mathbb{R}^2$, for each $i = 0, 1, \ldots, n$, $\mathbf{s}_i = (x_i, y_i)$ where x_i and y_i are the Cartesian coordinates. Alternatively, if $\mathbf{s}_i$ is a location on the Earth, then x_i and y_i could represent the longitude and latitude, respectively.

Now, let $Z(\cdot)$ be some spatial process observed on D (e.g. temperature, humidity, concentration of nitrogen in a soil sample, etc.). The goal of Kriging is to predict the value of $Z(\mathbf{s}_0)$ based on $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$. Alternatively, prediction can be made for more than one unobserved location.

If the assumptions regarding the second-order properties of the process $Z(\cdot)$ are met, then Kriging produces a best linear unbiased predictor (BLUP), meaning that Kriging calculates weights $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ such that $\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)$ is unbiased with a minimized mean square prediction error (MSPE).

Kriging when the mean of the spatial process is constant throughout the domain is known as Ordinary Kriging. When the mean is allowed to vary as a function of the location $\mathbf{s}$, a procedure known as Universal Kriging is used. Often, it is easier to estimate the mean function and detrend the data, thus leaving a collection of residuals with mean 0. Ordinary Kriging can then be employed.

3. Shortcomings of Kriging

One of the assumptions of most forms of Kriging is that the observed process $Z(\cdot)$ comes from a Gaussian process, meaning that $Z(\mathbf{s}_0), Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ follow a multivariate normal distribution. In reality, this assumption is not always realistic. While many real world spatial processes are in fact Gaussian, several are not. One type of Kriging which deals with this problem is Trans-Gaussian Kriging. This type of Kriging deals with so-called Trans-Gaussian data, or data that is not from a Gaussian process but may be converted to one using some transformation. Suppose that the observed process $Z(\cdot)$ is not Gaussian, but there exists an invertible function $\phi(\cdot)$ such that

$$Y(\cdot) = \phi^{-1}(Z(\cdot))$$

is a Gaussian process. A common practice is to perform Kriging for the Gaussian process $Y(\cdot)$ and then transform the prediction using ϕ. This practice, however, leads to biased predictions. Trans-Gaussian Kriging (Cressie 1993) attempts to correct for that bias. In this dissertation, we propose and develop a method for bias correction using parametric bootstrap methods. The idea of applying bootstrapping to spatial prediction with Trans-Gaussian data is new. For a similar use of boostrapping in a temporal setting, see Bandyopadhyay and Lahiri (2010).

B. Replicated Time Series Data

In time series analysis, as in the study of spatial statistics, the correlation structure between the observations is of interest. A straightforward example is the AR(1) model. Suppose that Y_t is some random process observed at time t. The AR(1) model states that

$$X_t = \rho X_{t-1} + \epsilon_t,$$

where $E\epsilon_t = 0$, $E\epsilon_t^2 = \sigma^2$, and ϵ_{t_1} and ϵ_{t_2} are independent if $t_1 \neq t_2$. Here, σ^2 represents the variance of the white noise in the process, and ρ represents the strength of correlation to the previous observation in the series. Estimation of these parameters is important for a variety of time series applications, such as prediction of future observations. A common application would be to consider X_t to be the price of a given stock at the end of day t of trading. Naturally, predicting X_{t+k}, the price at the end of day $t + k$ could be very beneficial.

1. Problem Background

In some scientific and engineering fields, the values of the model parameters have practical meaning that is important for the understanding of the field. The manufacturing of semiconductors, for example, has motivated a problem in the study of replicated time series data, regarding the estimation of covariance parameters. In Chapter III, we will briefly discuss the semiconductor application. However, our focus will be on the statistical side of the problem, and a brief description follows here.

Suppose that we have M independent stationary time series such that X_{it} is the observation from the i-th time series taken at time t. Suppose that all M time series

have a common autocovariance function of the form

$$C(h;\theta) = \theta_0 \rho(h;\theta_1,\ldots,\theta_p),$$

where $h \in \mathbb{Z}$ is the difference in time ("lag") between two observations, $\theta_0 > 0$ is the variance of the process, and $\rho(\cdot;\theta_1,\ldots,\theta_p)$ is a correlation function, determining the strength of relationship between two points, depending on h and parameters $\theta_1,\ldots,\theta_p$. Our goal is to not only estimate the parameter values contained in θ, but to use nonparametric bootstrap methods to examine the behavior (specifically, the variance or mean squared error) of those estimators.

2. Basics of the Moving Block Bootstrap

The moving block boostrap (Künsch 1989), or MBB, is a nonparametric resampling algorithm for studying the behavior of a statistic based on dependent data. When the observations are independent, it is possible to resample them individually when bootstrapping. However, with dependent data, resampling observations individually would lose the vitally important correlation structure with nearby observations. The MBB allows for groups, or blocks, of nearby and highly correlated observations to be resampled together. The blocks are then concatenated to form the bootstrap replicates. This resampling procedure is normally performed on the residuals.

Most early work regarding the MBB involved short-range dependent processes. Works such as Lahiri (1993) showed that under long-range dependence, the MBB failed unless further restrictions are put on the statistic in question's asymptotic properties. In this dissertation, we will use the MBB to investigate the variance of covariance parameter estimators.

3. Basics of Subsampling

Subsampling, unlike bootstrap methods, does not involve recreating data sets of the same size as the original data in order to study the properties of a statistic. Instead, the subsampling process involves taking small subsets of the data, and calculating the statistic in question for each subset. In the case of dependent data, these subsets are very much like the blocks of the MBB, but they are not attached to one another to form a data set of the same size as the original. In this dissertation, we will use subsampling to estimate the weights to be used in weighted least squares estimation of covariance parameters.

C. Overview

In Chapter II, a bootstrap method is developed and tested for correcting bias in Kriging prediction when the observed data is not from a Gaussian process, but can be converted to a Gaussian process via a one-to-one transformation of the data. Testing of the method will be discussed with both simulated and real world data. In Chapter III, a bootstrap method is developed and tested for the estimation of covariance parameters for replicated time series data.

CHAPTER II

BOOTSTRAP BASED TRANS-GAUSSIAN KRIGING

A. Introduction

When performing spatial prediction via Kriging for Trans-Gaussian data, a common approach is to perform Kriging on the Gaussian data and then do a simple transformation. However, the resulting naïve predictor will be biased due to the fact that generally, $E(\phi(X)) \neq \phi(E(X))$. To eliminate this bias, Cressie (1993) suggests an additive correction factor. Here, we seek to improve upon the naive predictor through the use of a correction factor to be calculated using bootstrap methods.

B. Problem Description

Let $Z(\cdot)$ be a spatial process on a domain $D \subset \mathbb{R}^d$ and let $\mathbf{Z} \equiv \{Z(\mathbf{s}_i) : i = 1, \ldots, n\}$ be observed, where the sites $\mathbf{s}_1, \ldots, \mathbf{s}_n$ either lie on a regular grid (in which case, we take $D = \mathbb{Z}^d$, the d-dimensional integer grid) or are irregularly spaced (in which case $D = \mathbb{R}^d$). Furthermore, let $\mathbf{s}_0 \in D$ be a site such that $Z(\mathbf{s}_0)$ is unobserved. Suppose that the observable process $Z(\cdot)$ is such that for a known continuous and invertible link function $\phi(\cdot)$, $Z(\mathbf{s}) \equiv \phi(Y(\mathbf{s}))$, and

$$Y(\mathbf{s}) \equiv \mu(\mathbf{s}; \beta) + \epsilon(\mathbf{s}), \quad \mathbf{s} \in D$$

where $\epsilon(\mathbf{s})$ is a zero-mean, second order stationary Gaussian process, with variogram $2\gamma(\cdot; \theta) \equiv E(\epsilon(\cdot) - \epsilon(\mathbf{0}))^2$, and covariogram $C(\cdot; \theta) \equiv E\epsilon(\cdot)\epsilon(\mathbf{0})$. We suppose that $\mu(\cdot; \beta)$ and $2\gamma(\cdot; \theta)$ (and hence $C(\cdot; \theta)$) are known, except for the finite dimensional mean parameters $\beta \in \mathcal{B}$ and covariance parameters $\theta \in \Theta$. Alternatively, $\mu(\cdot; \beta)$ can be identically zero.

1. Examples

Example B.1. *Log-Normal Kriging:* Suppose that the $Z(\cdot)$ variables are positive with probability one. Define

$$Y(\mathbf{s}) = \log Z(\mathbf{s}), \quad \mathbf{s} \in D.$$

Then $Z(\cdot)$ is a Log-Normal process.

Example B.2. *Logit Kriging:* Let $Z(\mathbf{s})$ denote a $(0, 1)$ valued random variables, e.g., a proportion. Consider a health related example. $Z(\mathbf{s})$ might represent the mortality rate due to a certain disease at a given location. In this case, $n(\mathbf{s})$ represents the number of people with the disease, and let π be the probability of dying from the disease. If each patient is independent, then the number of people who die from the disease, $X(\mathbf{s})$, follows a binomial distribution with parameters π and $n(\mathbf{s})$. Then $Z(\mathbf{s}) = X(\mathbf{s})/n(\mathbf{s})$. Let

$$Y(\mathbf{s}) = \text{logit } Z(\mathbf{s}), \quad \mathbf{s} \in D,$$

where logit $Z = \log\Big(Z/[1 - Z]\Big)$, $Z \in (0, 1)$. Then $Z(\cdot)$ is a translated Logit-Normal process.

C. Background

1. Trans-Gaussian Kriging

If possible, we wish to work with the underlying Gaussian process instead of the observed Trans-Gaussian data, which is nonstationary in both the mean and the covariance. The problem is simplified further if we can work with a process with mean 0. Since one assumption is that the form of $\mu(\cdot; \beta)$ is known (e.g., linear model, quadratic model, etc.), it is possible to detrend the data and perform Kriging on the

residuals.

First, consider the "ideal" case where the parameters β and θ are known. Note that in this case, the exact mean of $Y(\cdot)$'s are known, and hence, we may work with the error variables $\epsilon(\mathbf{s}) = Y(\mathbf{s}) - \mu(\mathbf{s};\beta)$. Let $\tilde{Y}(\mathbf{s}_0)$ be the Ordinary Kriging predictor of $Y(\mathbf{s}_0)$ based on $\mathbf{Y} \equiv \{Y(\mathbf{s}_i) : i = 1, \ldots, n\}$, which are observed, when β and θ are known. Specifically, if $\boldsymbol{\epsilon} \equiv \{\epsilon(\mathbf{s}_i) : i = 1, \ldots, n\}$, then the Ordinary Kriging predictor is

$$\begin{aligned} \tilde{Y}(\mathbf{s}_0) &= \mu(\mathbf{s}_0;\beta) + \boldsymbol{\lambda}'\boldsymbol{\epsilon} \\ &= \mu(\mathbf{s}_0;\beta) + \sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{s}_i) \end{aligned} \tag{C.1}$$

where the weights $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$ are calculated for Ordinary Kriging as suggested by Cressie (1993). Let $\boldsymbol{\gamma} \equiv (C(\mathbf{s}_0 - \mathbf{s}_1;\theta), \ldots, C(\mathbf{s}_0 - \mathbf{s}_n;\theta))'$ and Σ be an $n \times n$ matrix such that its (i,j)-th entry is given by $C(\mathbf{s}_i - \mathbf{s}_j;\theta)$. When θ is known (as we assume here), the Kriging weights are

$$\boldsymbol{\lambda} \equiv \left(\gamma + \mathbf{1}\frac{1 - \mathbf{1}\Sigma^{-1}\gamma}{\mathbf{1}\Sigma^{-1}\mathbf{1}}\right)' \Sigma^{-1}. \tag{C.2}$$

The "natural" but biased Trans-Gaussian Kriging predictor of $Z(\mathbf{s}_0)$ in the "ideal" case is then given by

$$\tilde{Z}(\mathbf{s}_0) = \phi\left(\tilde{Y}(\mathbf{s}_0)\right). \tag{C.3}$$

In practice, the parameters β and θ are typically unknown. We suppose that estimators $\hat{\beta}$ and $\hat{\theta}$ of the model parameters are given. Let

$$\hat{\epsilon}(\mathbf{s}) = Y(\mathbf{s}) - \mu\left(\mathbf{s};\hat{\beta}\right).$$

Also, let $\hat{\gamma}$ and $\hat{\Sigma}$ be estimated versions of γ and Σ obtained by replacing θ with $\hat{\theta}$.

Define the estimated predictor of $Y(\mathbf{s}_0)$ by

$$\hat{Y}(\mathbf{s}_0) = \mu\left(\mathbf{s}_0; \hat{\beta}\right) + \sum_{i=1}^{n} \hat{\lambda}_i \hat{\epsilon}(\mathbf{s}_i), \tag{C.4}$$

where $\hat{\boldsymbol{\lambda}} = \left(\hat{\lambda}_1, \ldots, \hat{\lambda}_n\right)$ is given by (C.2) with $\gamma = \hat{\gamma}$, $\Sigma = \hat{\Sigma}$. The naïve Trans-Gaussian predictor of $Z(\mathbf{s}_0)$ is then given by

$$\hat{Z}(\mathbf{s}_0) = \phi\left(\hat{Y}(\mathbf{s}_0)\right). \tag{C.5}$$

2. Bias Corrected Predictors

Note that the predictors in (C.3) and (C.5) are biased for a nonlinear $\phi(\cdot)$ such as in the examples. Often, this bias is significant when $Y(\mathbf{s}_0)$ is in the tails of its marginal univariate normal distribution. Then, depending on the extremity of the transformation performed by $\phi(\cdot)$, any inaccuracy in $\hat{Y}(\mathbf{s}_0)$ can be magnified.

As a result, often a bias corrected version of $\tilde{Z}(\mathbf{s}_0)$ or $\hat{Z}(\mathbf{s}_0)$ is constructed, primarily using the delta method, such as in a predictor similar to the one proposed by Cressie (1993), given by

$$\hat{Z}^C(\mathbf{s}_0) = \hat{Z}(\mathbf{s}_0) + \phi''(\hat{\mu}_{0,Y}) \left(\frac{\hat{\tau}^2(\mathbf{s}_0)}{2} - \hat{m}_Y \right), \tag{C.6}$$

where $\hat{\mu}_{0,Y} = \mu\left(\mathbf{s}_0; \hat{\beta}\right)$ is an estimate of $E(Y(s_0))$, $\hat{\tau}^2(\mathbf{s}_0)$ is an estimate of the Kriging variance, or mean square prediction error (MSPE), of $\tilde{Y}(\mathbf{s}_0)$, given by

$$\hat{\tau}^2(\mathbf{s}_0) = C\left(\mathbf{0}; \hat{\theta}\right) - \hat{\boldsymbol{\lambda}}' \hat{\gamma} + \hat{m}, \tag{C.7}$$

and $\hat{m} = m_Y\left(\hat{\theta}\right)$. Here, $m_Y(\theta)$ is a Lagrangian multiplier, given by

$$m_Y(\theta) = \frac{(1 - \mathbf{1}'\Sigma^{-1}\boldsymbol{\gamma})}{\mathbf{1}\Sigma^{-1}\mathbf{1}}.$$

Cressie (1993) states that in the case where $\mu(\cdot; \beta) \equiv 0$ and θ is known, the accuracy

of this predictor relies on $\sigma_Y^2(\mathbf{s}_0)$, the variance of $Y(\mathbf{s}_0)$, being "small." It can be shown that the correction factor in (C.6) is derived from the formal Taylor expansion of $\phi(\cdot)$, centered at $EY(\mathbf{s}_0)$,

$$\phi(Y(\mathbf{s}_0)) = \sum_i \frac{\phi^{(i)}(\mu(\mathbf{s}_0;\beta))}{i!} \left[Y(\mathbf{s}_0) - \mu(\mathbf{s}_0;\beta)\right]^i, \tag{C.8}$$

where $\phi^{(i)}(\mu(\mathbf{s}_0;\beta))$ denote the i-th derivative of $\phi(\cdot)$ at $\mu(\mathbf{s}_0;\beta)$. See Section II.E for the derivation of (C.8). Notice that the approximate predictor in (C.6) ignores the terms in this summation with $i \geq 3$. This leads to other limitations on the type of function $\phi(\cdot)$ that can be used. A low order polynomial function such as $\phi(y) = y^3$ will be acceptable because the higher order derivatives are zero. However, some families of functions $\phi(\cdot)$ such as exponential functions will have nonzero higher order derivatives, causing the later terms in (C.8) to become nontrivial.

A Bayesian approach to bias correction is presented in De Oliveira et al. (1997). However, here, we concern ourselves with frequentist methods.

In the next section, we construct a bias-corrected predictor based on a parametric spatial bootstrap, improving upon the bias correction method described above. Specifically, for the biased Kriging predictor $\hat{Z}(\mathbf{s}_0)$ in (C.5), we are looking to estimate a value c such that

$$\hat{Z}_M^{BC}(\mathbf{s}_0) \equiv c\hat{Z}(\mathbf{s}_0) \tag{C.9}$$

is approximately unbiased. Here, the subscript M corresponds to a "multiplicative" bias correction. The multiplicative bias correction factor c described in (C.9) works best for nonnegative Z-variables, such as in the case of log-normal Kriging. For Z-variables with an unrestricted range, such as in power transformation models, an alternative is to use an additive correction factor a based on the bootstrap such that

$$\hat{Z}_A^{BC}(\mathbf{s}_0) = \hat{Z}(\mathbf{s}_0) + a \tag{C.10}$$

is approximately unbiased.

D. Bootstrap Based Bias Corrected Prediction

1. Known Parameters

For now, we will work in the setting where we assume we know the value of the covariance parameter θ and the mean parameter β. Note that in the mulitiplicative case that the factor that makes $\tilde{Z}(\mathbf{s}_0)$ exactly unbiased is given by $c \equiv c(\beta, \theta) = EZ(\mathbf{s}_0)/E\tilde{Z}(\mathbf{s}_0)$, while in the additive case, $a \equiv a(\beta, \theta) = EZ(\mathbf{s}_0) - E\tilde{Z}(\mathbf{s}_0)$.

Since c and a depend on unknown parameters, they can not be used in practice. Further, the explicit forms of $c(\beta, \theta)$ and $a(\beta, \theta)$ are often difficult to derive analytically. In this section, we present a bootstrap method for estimating the factors $c(\beta, \theta)$ and $a(\beta, \theta)$ that does not require the user to do such analytical work; it replaces the tedious and intractable analytical derivation by a simple, albeit computer intensive bootstrap algorithm. The same principle also produces an estimate of the MSPE of the Trans-Gaussian predictor with and without bias-correction. Explicit formula for the MSPE of the Trans-Gaussian predictor is not available in the literature, except in some very specific cases (e.g., Log-Normal Kriging and its generalized versions); see Cressie (1993), Shimizu and Iwase (1987), and the references therein.

The parametric bootstrap algorithm for the known parameter case for estimating a and c is as follows:

1. Generate $\epsilon^*(\mathbf{s}_0), \epsilon^*(\mathbf{s}_1), \ldots, \epsilon^*(\mathbf{s}_n)$ from the Gaussian process with covariogram $C(\cdot; \theta)$. Note that this involves the generation $n + 1$ observations.

2. Compute $\tilde{Y}^*(\mathbf{s}_0)$ by replacing $\{\epsilon(\mathbf{s}_i) : i = 1, \ldots, n\}$ in (C.1) with the generated variables $\{\epsilon^*(\mathbf{s}_i) : i = 1, \ldots, n\}$.

3. Compute $\tilde{Z}^*(\mathbf{s}_0) = \phi\left(\tilde{Y}^*(\mathbf{s}_0)\right)$ and $Z^*(\mathbf{s}_0) = \phi\left(Y^*(\mathbf{s}_0)\right)$.

4. Let $\tilde{a}_n = E_* Z^*(\mathbf{s}_0) - E_* \tilde{Z}^*(\mathbf{s}_0)$ and $\tilde{c}_n = E_* Z^*(\mathbf{s}_0) / E_* \tilde{Z}^*(\mathbf{s}_0)$ where E_* denotes the conditional expectation given $\mathbf{Y}$. Then, the bootstrap-based bias-corrected predictors of $Z(\mathbf{s}_0)$ are given by

$$\begin{aligned} \tilde{Z}_A^{BC}(\mathbf{s}_0) &= \tilde{Z}(\mathbf{s}_0) + \tilde{a}_n, \text{ and} \\ \tilde{Z}_M^{BC}(\mathbf{s}_0) &= \tilde{c}_n \cdot \tilde{Z}^B(\mathbf{s}_0). \end{aligned}$$

In practice, $\tilde{a}_n$ and $\tilde{c}_n$ are evaluated by Monte Carlo simulation, where steps 1, 2, and 3 are repeated a large number of times, say B times, resulting in the bootstrap replicates $\{\tilde{Z}^{*j}(\mathbf{s}_0) : j = 1, \ldots, B\}$ and $\{Z^{*j}(\mathbf{s}_0) : j = 1, \ldots, B\}$, for $\tilde{Z}^*(\mathbf{s}_0)$ and $Z^*(\mathbf{s}_0)$, respectively. Then, the Monte-Carlo approximation to $\tilde{a}_n$ and $\tilde{c}_n$ are given by

$$\begin{aligned} \tilde{a}_n^{MC} &\equiv \frac{1}{B}\left[\sum_{j=1}^{B} Z^{*j}(\mathbf{s}_0) - \sum_{j=1}^{B} \tilde{Z}^{*j}(\mathbf{s}_0)\right], \text{ and} \\ \tilde{c}_n^{MC} &\equiv \frac{\sum_{j=1}^{B} Z^{*j}(\mathbf{s}_0)}{\sum_{j=1}^{B} \tilde{Z}^{*j}(\mathbf{s}_0)}, \end{aligned}$$

respectively.

2. Unknown Parameters

For real applications, the value of the covariance parameter θ is unknown, as is the mean structure of the Gaussian process. Therefore, in reality the aforementioned predictors must be calculated using estimates of the parameters.

For concreteness, we propose using Cressie's weights to obtain $\hat{\theta}$, a WLS estimate of the covariance parameter θ, which will then be used in the calculations of $\hat{\gamma}$ and $\hat{\Sigma}$. However, other choices of $\hat{\beta}$ and $\hat{\theta}$ are also possible.

1. From the observed $\{Z(\mathbf{s}_i) : i = 1, \ldots, n\}$, find $\{Y(\mathbf{s}_i) = \phi^{-1}(Z(\mathbf{s}_i)) : i =$

$1, \ldots, n\}$.

2. Construct $\hat{\beta}$, an estimator of the parameter (vector) β, based on $\{Y(\mathbf{s}_i) : i = 1, \ldots, n\}$. Then, use the residuals to calculate Matheron's (1962) method of moments estimator for the variogram and weighted least squares to find $\hat{\theta}$.

3. Generate $\epsilon^*(\mathbf{s}_0), \epsilon^*(\mathbf{s}_1), \ldots, \epsilon^*(\mathbf{s}_n)$ from the Gaussian process with covariogram $C\left(\cdot; \hat{te}\right)$.

4. Compute $\hat{Y}^*(\mathbf{s}_0)$ by replacing $\{\epsilon(\mathbf{s}_i) : i = 1, \ldots, n\}$ in (C.4) with the generated variables $\{\epsilon^*(\mathbf{s}_i) : i = 1, \ldots, n\}$ and replacing β and θ with $\hat{\beta}$ and $\hat{\theta}$, respectively.

5. Compute $\hat{Z}^*(\mathbf{s}_0) = \phi\left(\hat{Y}^*(\mathbf{s}_0)\right)$ and $Z^*(\mathbf{s}_0) = \phi(Y^*(\mathbf{s}_0))$.

6. Let $\hat{a}_n = E_* Z^*(\mathbf{s}_0) - E_* \hat{Z}^*(\mathbf{s}_0)$ and let $\hat{c}_n = E_* Z^*(\mathbf{s}_0) / E_* \hat{Z}^*(\mathbf{s}_0)$ where E_* denotes the conditional expectation given $\mathbf{Y}$. Then, the bootstrap-based bias-corrected predictors of $Z(\mathbf{s}_0)$ are given by

$$\hat{Z}_A^{BC}(\mathbf{s}_0) = \hat{Z}(\mathbf{s}_0) + \hat{a}_n, \text{ and} \tag{D.11}$$

$$\hat{Z}_M^{BC}(\mathbf{s}_0) = \hat{c}_n \cdot \hat{Z}^B(\mathbf{s}_0), \tag{D.12}$$

respectively.

Again, a Monte Carlo simulation is normally used to approximate $\hat{a}_n$ and $\hat{c}_n$. Steps 3, 4, and 5 are repeated B times resulting in the bootstrap replicates $\{\hat{Z}^{*j}(\mathbf{s}_0) : j = 1, \ldots, B\}$ and $\{Z^{*j}(\mathbf{s}_0) : j = 1, \ldots, B\}$. Then, the Monte Carlo approximation to $\hat{c}_n$ is given by

$$\hat{c}_n^{MC} \equiv \frac{\sum_{j=1}^{B} Z^{*j}(\mathbf{s}_0)}{\sum_{j=1}^{B} \hat{Z}^{*j}(\mathbf{s}_0)}.$$

A similar adjustment can be made for unknown parameters in the case of an additive correction factor $\hat{a}_n^{MC}$.

3. Bootstrap MSPE Estimation

Another advantage of the bootstrap procedure is that it allows for estimation of the MSPE of each predictor based on a single data set, because of the large number of bootstrap replicates produced by the resampling procedure. Here, we work in the case of unknown parameters, so we will refer to Steps 1-6 above leading to the predictors in (D.11) and (D.12). An analogous procedure can be easily seen if parameters are known.

For $\hat{Z}(\mathbf{s}_0)$, repeat Steps 3, 4, and 5 B times as described above. The bootstrap estimate of the MSPE is given by $\hat{\text{MSPE}}\left(\hat{Z}(\mathbf{s}_0)\right) = E_*\left[\hat{Z}^*(\mathbf{s}_0) - Z^*(\mathbf{s}_0)\right]^2$, so the Monte Carlo approximation will be

$$\hat{\text{MSPE}}_n^{MC}\left(\hat{Z}(\mathbf{s}_0)\right) = \frac{1}{B}\sum_{j=1}^{B}\left[\hat{Z}^{*j}(\mathbf{s}_0) - Z^{*j}(\mathbf{s}_0)\right]^2.$$

For $\hat{Z}^C(\mathbf{s}_0)$, the procedure is very similar. Because we can treat the estimates $\hat{\beta}$ and $\hat{\theta}$ as the true values of the parameters for the boostrap distribution, the correction does not change for each replicate, and

$$\hat{Z}^{*C}(\mathbf{s}_0) = \hat{Z}^*(\mathbf{s}_0) + \phi''(\hat{\mu}_{0,Y})\left(\frac{\hat{\tau}^2(\mathbf{s}_0)}{2} - \hat{m}_Y\right).$$

Therefore, the bootstrap estimate of the MSPE of $\hat{Z}^C(\mathbf{s}_0)$ is given by

$$\hat{\text{MSPE}}\left(\hat{Z}^C(\mathbf{s}_0)\right) = E_*\left[\hat{Z}^*(\mathbf{s}_0) + \phi''(\hat{\mu}_{0,Y})\left(\frac{\hat{\tau}^2(\mathbf{s}_0)}{2} - \hat{m}_Y\right) - Z^*(\mathbf{s}_0)\right]^2,$$

and the Monte Carlo approximation is

$$\hat{\text{MSPE}}_n^{MC}\left(\hat{Z}^C(\mathbf{s}_0)\right) = \frac{1}{B}\sum_{j=1}^{B}\left[\hat{Z}^{*j}(\mathbf{s}_0) + \phi''(\hat{\mu}_{0,Y})\left(\frac{\hat{\tau}^2(\mathbf{s}_0)}{2} - \hat{m}_Y\right) - Z^{*j}(\mathbf{s}_0)\right]^2,$$

where all quantities are as defined above.

Estimating the MSPE for the bootstrap predictors is more complicated, because

each bootstrap replicate does not create its own estimate of a_n or c_n. Here, a nested bootstrap procedure is necessary. The following procedure must be followed for each bootstrap replicate $\{\epsilon^*(\mathbf{s}_i) : i = 0, \ldots, n\}$:

1. Based on $\{\epsilon^*(\mathbf{s}_i) : i = 0, \ldots, n\}$, calculate $\hat{\theta}^*$ using WLS methods, in a method similar to the one used to calculate $\hat{\theta}$.

2. Generate $\epsilon^{**}(\mathbf{s}_0), \epsilon^{**}(\mathbf{s}_1), \ldots, \epsilon^{**}(\mathbf{s}_n)$, the nested bootstrap replicate, based on a Gaussian process with covariogram $C\left(\cdot; \hat{\theta}^*\right)$.

3. Calculate $\hat{Y}^{**}(\mathbf{s}_0)$ by replacing $\{\epsilon(\mathbf{s}_i) : i = 1, \ldots, n\}$ in (C.4) with $\{\epsilon^{**}(\mathbf{s}_i) : i, \ldots, n\}$ and replacing β and θ with $\hat{\beta}$ and $\hat{\theta}^*$, respectively. Notice that this only requires a bootstrap estimate of θ.

4. Compute $\hat{Z}^{**}(\mathbf{s}_0)$ and $Z^{**}(\mathbf{s}_0)$.

5. Let $\hat{c}_n^* = E_{**}Z^{**}(\mathbf{s}_0)/E_{**}\hat{Z}^{**}(\mathbf{s}_0)$ and $\hat{a}_n^* = E_{**}Z^{**}(\mathbf{s}_0) - E_{**}\hat{Z}^{**}(\mathbf{s}_0)$. In practice, these conditional expectations are calculated based on B_2 nested bootstrap replicates for large B_2. We recommend $B_2 \leq B$.

The bootstrap MSPE estimates of the multiplicative and additive bootstrap predictors are thus given by the conditional expectations

$$\text{MSPE}\left(\hat{Z}_M^{BC}(\mathbf{s}_0)\right) = E_*\left[\hat{c}_n^*\hat{Z}^*(\mathbf{s}_0) - Z^*(\mathbf{s}_0)\right]^2$$

and

$$\text{MSPE}\left(\hat{Z}_A^{BC}(\mathbf{s}_0)\right) = E_*\left[\hat{a}_n^* + \hat{Z}^*(\mathbf{s}_0) - Z^*(\mathbf{s}_0)\right]^2.$$

The Monte Carlo approximations are

$$\text{MSPE}_n^{MC}\left(\hat{Z}_M^{BC}(\mathbf{s}_0)\right) = \frac{1}{B}\sum_{j=1}^{B}\left[\hat{c}_n^{*j}\hat{Z}^{*j}(\mathbf{s}_0) - Z^{*j}(\mathbf{s}_0)\right]^2,$$

and

$$\text{MSPE}_n^{MC}\left(\hat{Z}_A^{BC}(\mathbf{s}_0)\right) = \frac{1}{B}\sum_{j=1}^{B}\left[\hat{a}_n^{*j} + \hat{Z}^{*j}(\mathbf{s}_0) - Z^{*j}(\mathbf{s}_0)\right]^2.$$

Note that this nested bootstrap procedure is computationally very intensive, involving B estimates of θ and the generation of BB_2 bootstrap replicates.

4. Bootstrap Prediction Intervals

In addition to point estimates, prediction intervals are often used in spatial prediction. These intervals are beneficial to point estimates in that they gauge the amount of variability in a predictor, and inform the reader of that variability via a margin of error. Moreover, the intervals give a range of plausible values for an observation, with a given level of certainty or confidence. We start by referring to the estimate of the Kriging variance of $\hat{Y}(\mathbf{s}_0)$, given in (C.7). A $100(1-\alpha)\%$ prediction interval for $Y(\mathbf{s}_0)$ is given by

$$\left(\hat{Y}(\mathbf{s}_0) - z_{\alpha/2}\hat{\tau}(\mathbf{s}_0), \hat{Y}(\mathbf{s}_0) + z_{\alpha/2}\hat{\tau}(\mathbf{s}_0)\right), \tag{D.13}$$

where $z_{\alpha/2}$ is the number such that the probability of a standard normal variable being greater than $z_{\alpha/2}$ is $\alpha/2$. This interval may be interpreted by saying that if many samples are taken, and $\hat{Y}(\mathbf{s}_0)$ and $\hat{\tau}(\mathbf{s}_0)$ is calculated each time, the resulting prediction interval will contain $Y(\mathbf{s}_0)$ $100(1-\alpha)\%$ of the time. When $Y(\cdot)$ is a Gaussian process, a $100(1-\alpha)\%$ prediction interval for $Z(\mathbf{s}_0)$ may simply be calculated by transforming the endpoints of the interval in (D.13), i. e.,

$$\left(\phi\left(\hat{Y}(\mathbf{s}_0) - z_{\alpha/2}\hat{\tau}(\mathbf{s}_0)\right), \phi\left(\hat{Y}(\mathbf{s}_0) + z_{\alpha/2}\hat{\tau}(\mathbf{s}_0)\right)\right). \tag{D.14}$$

The interpretation of this interval is similar to the one for the interval in (D.13), and the intervals will have the same coverage probabilities.

The bootstrap procedure can also be used to produce prediction intervals. When

the $Y(\cdot)$ process is Gaussian, these prediction intervals should perform no better than the interval in (D.14). However, when working with real world data, the assumptions of a Gaussian process may be questionable, and bootstrap prediction intervals will be of some use. Using order statistics, a $100(1-\alpha)\%$ bootstrap-t prediction interval is be given by

$$\left(\phi\left(\hat{Y}(\mathbf{s}_0) + t^*_{(B(\alpha/2))}\hat{\tau}(\mathbf{s}_0)\right), \phi\left(\hat{Y}(\mathbf{s}_0) + t^*_{(B(1-\alpha/2))}\hat{\tau}(\mathbf{s}_0)\right)\right), \tag{D.15}$$

where $t^* = (\hat{Y}^*(\mathbf{s}_0) - Y^*(\mathbf{s}_0))/\hat{\tau}$, B is the number of bootstrap replicates, and $t^*_{(n)}$ is the n-th order statistic of t^*; that is, the n-th observation when the the B values of t^* are placed in increasing order. Also, symmetric bootstrap-t intervals can be given by

$$\left(\phi\left(\hat{Y}(\mathbf{s}_0) + |t^*|_{(B(\alpha/2))}\hat{\tau}(\mathbf{s}_0)\right), \phi\left(\hat{Y}(\mathbf{s}_0) + |t^*|_{(B(1-\alpha/2))}\hat{\tau}(\mathbf{s}_0)\right)\right), \tag{D.16}$$

where the order statistics work similarly as before, but this time are for the absolute value of t^*.

E. Theoretical Results

1. Preliminaries

The observed process is

$$Z(\mathbf{s}) = \phi(Y(\mathbf{s})),$$

where $\phi(\cdot)$ is a known one-to-one fucntion,

$$Y(\mathbf{s}) = \mu(\mathbf{s};\beta) + \epsilon(\mathbf{s}),\ \mathbf{s} \in D,$$

and $\epsilon(\cdot)$ is a second order zero-mean stationary spatial Gaussian process on $D \subset \mathbb{R}^d$, with covariogram $C(\cdot;\theta)$.

Let

$$\hat{Y}(\mathbf{s}_0) = \mu\left(\mathbf{s}_0; \hat{\beta}\right) + \hat{\epsilon}(\mathbf{s}_0)$$

and

$$\tilde{\epsilon}(\mathbf{s}_0) = \boldsymbol{\lambda}'\boldsymbol{\epsilon}$$

be the ideal predictor of $\epsilon(\mathbf{s}_0)$ when $\boldsymbol{\epsilon} = \{\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_n)\}$ are observable, with weights given by (C.2) and MSPE

$$\tau^2(\mathbf{s}_0) = C(\mathbf{0}; \theta) - \boldsymbol{\lambda}'\gamma + m,$$

with

$$m = \frac{1 - \mathbf{1}'\Sigma^{-1}\gamma}{\mathbf{1}\Sigma^{-1}\mathbf{1}}.$$

Here,

$$\begin{aligned} \gamma &= (C(\mathbf{s}_0 - \mathbf{s}_1; \theta), \ldots, C(\mathbf{s}_0 - \mathbf{s}_n; \theta))', \\ \Sigma &= ((C(\mathbf{s}_i - \mathbf{s}_j; \theta))) \end{aligned}$$

for unknown parameters θ. These formulas are taken from Cressie (1993). If $\hat{\theta}$ is an estimate of θ, then

$$\begin{aligned} \hat{\epsilon}(\mathbf{s}_0) &= \hat{\boldsymbol{\lambda}}'\hat{\boldsymbol{\epsilon}}, \\ \hat{Z}(\mathbf{s}_0) &= \phi\left(\hat{Y}(\mathbf{s}_0)\right) \\ &= \phi\left(\mu(\mathbf{s}_0; \hat{\beta}) + \hat{\epsilon}(\mathbf{s}_0)\right), \end{aligned}$$

where $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}\left(\hat{\theta}\right)$ and $\hat{\boldsymbol{\epsilon}}$ is the vector of residuals, with i-th component $\hat{\epsilon}_i = Y(\mathbf{s}_i) - \mu\left(\mathbf{s}_0; \hat{\beta}\right)$, $1 \leq i \leq n$.

2. Derivation of the Taylor Series Based Bias Corrected Predictor

From the Taylor's expansion of $\phi(\cdot)$, we see that

$$\begin{aligned}\phi(Y(\mathbf{s}_0)) &= \phi(\mu(\mathbf{s}_0;\beta)) + [Y(\mathbf{s}_0) - \mu_Y(\mathbf{s}_0,\beta)]\phi'(\mu(\mathbf{s}_0;\beta)) \\ &\quad + \frac{[Y(\mathbf{s}_0) - \mu_Y(\mathbf{s}_0;\beta)]^2}{2}\phi''(\mu(\mathbf{s}_0;\beta)) + \cdots,\end{aligned}$$

and therefore, using the terms upto the second order, we obtain the approximation

$$\begin{aligned}EZ(\mathbf{s}_0) &= E\phi(Y(\mathbf{s}_0)) \\ &\approx \phi(\mu(\mathbf{s}_0;\beta)) + 0 \\ &\quad + \frac{\phi''(\mu(\mathbf{s}_0;\beta))}{2}E[Y(\mathbf{s}_0) - \mu_Y(\mathbf{s}_0;\beta)]^2. \qquad \text{(E.17)}\end{aligned}$$

And similarly,

$$\begin{aligned}E\phi\left(\tilde{Y}(\mathbf{s}_0)\right) &= E\phi\left(\mu(\mathbf{s}_0;\beta) + \tilde{\epsilon}(\mathbf{s}_0)\right) \\ &\approx E\left\{\phi(\mu(\mathbf{s}_0;\beta)) + \phi'(\mu(\mathbf{s}_0;\beta))\tilde{\epsilon}(\mathbf{s}_0)\right. \\ &\quad \left. + \frac{\phi''(\mu(\mathbf{s}_0;\beta))}{2}\tilde{\epsilon}(\mathbf{s}_0)^2\right\}. \qquad \text{(E.18)}\end{aligned}$$

Note that $E\tilde{\epsilon}(\mathbf{s}_0) = 0$. From (E.17) and (E.18),

$$\begin{aligned}&E\left(\phi\left(\tilde{Y}(\mathbf{s}_0)\right) - \phi(Y(\mathbf{s}_0))\right) \\ &\approx \frac{\phi''(\mu(\mathbf{s}_0;\beta))}{2}\left[E\tilde{\epsilon}(\mathbf{s}_0)^2 - E\epsilon(\mathbf{s}_0)^2\right] \\ &= \frac{\phi''(\mu(\mathbf{s}_0;\beta))}{2}\left[E\left(\tilde{\epsilon}(\mathbf{s}_0) - \epsilon(\mathbf{s}_0)\right)^2 + 2E\left(\epsilon(\mathbf{s}_0)(\tilde{\epsilon}(\mathbf{s}_0) - \epsilon(\mathbf{s}_0))\right]\right. \\ &= \frac{\phi''(\mu(\mathbf{s}_0;\beta))}{2}[\tau^2(\mathbf{s}_0) + 2(m - \tau^2(\mathbf{s}_0))] \\ &= -\frac{\phi''(\mu(\mathbf{s}_0;\beta))}{2}[\tau^2(\mathbf{s}_0) - 2m] \\ &= -\phi''(\mu(\mathbf{s}_0;\beta))\left[\frac{\tau^2(\mathbf{s}_0)}{2} - m\right],\end{aligned}$$

since

$$\begin{aligned} E\left(\epsilon(\mathbf{s}_0)(\tilde{\epsilon}(\mathbf{s}_0)-\epsilon(\mathbf{s}_0))\right) &= E\left(\epsilon(\mathbf{s}_0)(\boldsymbol{\lambda}'\boldsymbol{\epsilon}-\epsilon(\mathbf{s}_0))\right) \\ &= \boldsymbol{\lambda}'\boldsymbol{\gamma}-C(\mathbf{0};\theta) \\ &= m-[C(\mathbf{0};\theta)-\boldsymbol{\lambda}'\boldsymbol{\gamma}+m] \\ &= m-\tau^2(\mathbf{s}_0). \end{aligned}$$

This yields the "ideal" and "estimated" Taylor's expansion based bias corrected predictors:

$$\begin{aligned} \tilde{Z}^C(\mathbf{s}_0) &= \phi\left(\tilde{Y}(\mathbf{s}_0)\right)+\phi''(\mu(\mathbf{s}_0;\beta))\left[\frac{\tau^2(\mathbf{s}_0)}{2}-m\right], \\ \hat{Z}^C(\mathbf{s}_0) &= \phi\left(\hat{Y}(\mathbf{s}_0)\right)+\phi''\left(\mu(\mathbf{s}_0;\hat{\beta})\right)\left[\frac{\hat{\tau}^2}{2}-\hat{m}\right]. \end{aligned}$$

where $\hat{\tau}^2(\mathbf{s}_0)$ and $\hat{m}$ are given by, replacing θ with $\hat{\theta}$ in $\tau^2(\mathbf{s}_0) \equiv \tau^2(\mathbf{s}_0;\theta)$ and $m \equiv m(\theta)$.

3. Bias of Predictors

In this section, as well as in the accompanying proofs, we highlight the dependence of various quantities depending on (the distribution of) $Z(\mathbf{s}_0), Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ by using the subscript n. Thus, we shall write $\hat{\theta}=\hat{\theta}_n$, $\hat{\beta}=\hat{\beta}_n$, $\gamma=\gamma_n$, $\Sigma=\Sigma_n$, $\tilde{Z}(\mathbf{s}_0)=\tilde{Z}_n(\mathbf{s}_0)$, $\hat{Z}(\mathbf{s}_0)=\hat{Z}_n(\mathbf{s}_0)$, etc.

Further, we work under the assumption that the following assumption is met.

Condition E.1. Assume that $EY(\mathbf{s}_0)=\mu(\mathbf{s}_0;\beta)$ where $Y(\mathbf{s}_0)=\mu(\mathbf{s}_0;\beta)+\epsilon(\mathbf{s}_0)$ and that $\hat{\beta}_n$ is an estimate of the parameter β. Further, suppose that $\tilde{\epsilon}(\mathbf{s}_0)$ and $\hat{\epsilon}(\mathbf{s}_0)$ are Ordinary Kriging predictors of $\epsilon(\mathbf{s}_0)$ using known and unknown parameters,

respectively. Further, assume that

$$\left|\mu(\mathbf{s}_0;\hat{\beta}_n) - \mu(\mathbf{s}_0;\beta)\right| + |\hat{\epsilon}(\mathbf{s}_0) - \tilde{\epsilon}(\mathbf{s}_0)| \to_p 0$$

Proposition E.1. *Suppose that Condition E.1 holds and that $\phi(\cdot)$ admits a power-series representation given by*

$$\phi(x) = \sum_{k=0}^{\infty} d_k(x - \mu(\mathbf{s}_0))^k,\ x \in \mathbb{R} \tag{E.19}$$

for some $d_0, d_1, \ldots \in \mathbb{R}$. Further, supppose that $E\left[\hat{Z}_n(\mathbf{s}_0)\right]^2 = O(1)$ and that for some $k_1 \in (0,\infty)$,

$$\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} k_j|d_k d_j| 2^{(k+j-2)/2}\Gamma\left(\frac{k+j-1}{2}\right)\left[\sigma_n^{j+k-2} + \sigma_0^{j+k-2}\right] < k_1 < \infty, \tag{E.20}$$

for n large, where $\Gamma(\cdot)$ is the Gamma function and where $\sigma_0^2 = C(\mathbf{0};\theta)$ and $\sigma_n^2 = E\tilde{\epsilon}(\mathbf{s}_0)^2 = \boldsymbol{\lambda}_n'\Sigma_n\boldsymbol{\lambda}_n$. Then:

$$Bias\left(\hat{Z}_n(\mathbf{s}_0)\right) = \sum_{k=1}^{\infty}\frac{\phi^{(2k)}(\mu(\mathbf{s}_0;\beta))}{k!2^k}(\sigma_n^{2k} - \sigma_0^{2k}) + o(1), \tag{E.21}$$

and

$$Bias\left(\hat{Z}_n^C(\mathbf{s}_0)\right) = \sum_{k=2}^{\infty}\frac{\phi^{(2k)}(\mu(\mathbf{s}_0;\beta))}{k!2^k}(\sigma_n^{2k} - \sigma_0^{2k}) + o(1). \tag{E.22}$$

Next we comment about the regularity conditions. Condition E.1 is a mild requirement on the estimators $\hat{\beta}_n$ and $\hat{\theta}_n$. In particular, if $\mu(\mathbf{s}_0;\cdot)$ is continuous at β and $\left\{\hat{\beta}_n\right\}_{n\geq 1}$ is consistent for β, then

$$\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) - \mu(\mathbf{s}_0;\beta) \to_p 0.$$

Similarly, if $\|\Sigma_n(\theta)^{-1}\| = O(1)$, then it can be shown that consistency of $\left\{\hat{\theta}_n\right\}_{n\geq 1}$ and (equi-)continuity of $C(\mathbf{s};\cdot)$ at θ implies that $\hat{\epsilon}(\mathbf{s}_0) - \tilde{\epsilon}_n(\mathbf{s}_0) \to_p 0$.

Condition (E.19) on $\phi(\cdot)$ implies that $\phi(\cdot)$ is infinitely diffentiable, although a

more complicated expression can be derived for functions $\phi(\cdot)$ that are twice differentiable. The power series representation is specifically used for deriving an explicit expression for the bias of the naïve predictor in terms of σ_0^2 and σ_n^2. Note that this condition holds for both Log-normal Kriging and Logit Kriging (Examples B.1 and B.2). The requirement that $E\left[\hat{Z}_n(\mathbf{s}_0)\right]^2 = O(1)$ can be weakened to $E\left|\hat{Z}_n(\mathbf{s}_0)\right|^{1+\delta} = O(1)$ for some $\delta > 0$. Here, we take $\delta = 1$ as it seems to be the most natural condition in this context; the MSPE of $\hat{Z}_n(\mathbf{s}_0)$ is bounded only when $E\left[\hat{Z}_n(\mathbf{s}_0)\right]^2 = O(1)$.

Finally, Condition (E.20) is a sufficient condition for deriving the series representation of Bias $\left(\hat{Z}_n(\mathbf{s}_0)\right)$. Under (E.20), the sum on the right side of (E.21) is absolutely convergent.

Note that under the conditions of Proposition E.1, the bias of the naïve predictor is the difference between the values of an analytic function of σ_n^2 and σ_0^2. For pure-increasing domain asymptotic structure (cf. Cressie (1993), Lahiri (2003)) with either regularly-spaced or irregularly spaced data points, $\{\sigma_n\}_{n\geq 1}$ does not converge to σ_0 and therefore, the bias typically remains bounded away from zero. However, in the presence of an infill component in the spatial sampling design, under mixed or pure infill asymptotics, $\{\sigma_n\}_{n\geq 1}$ may converge to σ_0 (e. g., if $\mathbf{s}_0$ is a limit point of $\mathbf{s}_i$'s) and the bias must go to zero in this case. A similar remark applies to the bias-corrected predictor $\hat{Z}_n^C(\mathbf{s}_0)$ of Cressie (1993).

Next consider the bias properties of the bootstrap based bias-corrected predictors $\hat{Z}_{M,n}^{BC}(\mathbf{s}_0)$ and $\hat{Z}_{A,n}^{BC}(\mathbf{s}_0)$.

Let $f_0(\cdot)$ denote the probability density function of the $N(0,1)$ distribution, i. e.

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{ for } x \in \mathbb{R}.$$

Let $g(b,u) \equiv \int \phi(\mu(\mathbf{s}_0;b) + ux) f_0(x)dx$, for $b \in \mathcal{B}$, $u \in [0,\infty)$. Let $\boldsymbol{\lambda} = \lambda_n(\theta) = (\lambda_{1n}(\theta), \ldots, \lambda_{nn}(\theta))'$.

Condition E.2. Suppose that k_0, k_1, k_2, and k_3 are positive constants such that the following conditions hold.

(i) There exists $\delta_0 > 0$ and $\{\alpha_i\}_{i\geq 1} \subset [0, \infty)$ with $\sum_{i=1}^{\infty} \alpha_i < \infty$ such that for all $t \in \Theta$,

$$|\lambda_{in}| \leq \alpha_i \text{ for all } i = 1, \ldots, n,$$

and for all $\|t - \theta\| < \delta_0$,

$$|\lambda_{in}(t) - \lambda_{in}(\theta)| \leq k_0 |t - \theta| \alpha_i, \text{ for } 1 \leq i \leq n,$$

where $n \geq 1$.

(ii) $\{\tau_n(\cdot)\}_{n\geq 1}$ is equicontinuous at the true value θ.

(iii) $|\mu(\mathbf{s}; t_1) - \mu(\mathbf{s}; t_2)| \leq k_1 |t_1 - t_2|$ for all $\mathbf{s}, t_1, t_2$.

(iv) $\phi(\cdot)$ is differentiable and there exists a monotone function $\Phi^{(1)} : [0, \infty) \to [0, \infty)$ such that

$$E_{\beta,\theta} \Phi^{(1)} \left(|\mu(\mathbf{s}_0; \beta)| + k \sum_{i=1}^{n} \alpha_i |\epsilon(\mathbf{s}_i)| + k \left\| \hat{\beta}_n - \beta \right\| \right)^2 \leq k_2$$

for all n large and all β, θ.

(v) There exists a sequence $\{b_n\}_{n\geq 1} \to 0$ as $n \to \infty$ such that

$$\left(E_{\beta,\theta} \left\| \hat{\beta}_n - \beta \right\|^4 \right)^{1/4} \leq b_n$$

and

$$P_{\beta,\theta} \left(\left\| \hat{\theta}_n - \theta \right\| > \delta \right) \leq k(\delta) b_n$$

for any $\delta > 0$, for all β, θ.

(vi) $C(\mathbf{0}, \theta) < k_3$ for all θ.

Finally, consider the following condition.

Condition E.3. Suppose that $g(\cdot,\cdot)$ and $C(\mathbf{0},\cdot)$ are functions such that the following conditions hold.

(i) $g(\cdot,\cdot)$ and $C(\mathbf{0},\cdot)$ are continuous and $\{g^2(\hat{\beta}_n, C(\mathbf{0},\hat{\theta}_n))\}_{n\geq 1}$ and $\{g^2(\hat{\beta}_n, \tau_n^2(\hat{\theta}_n))\}_{n\geq 1}$ are uniformly integrable.

(ii) $E\left|\hat{Z}_n(\mathbf{s}_0)/E_*(\hat{Z}^*(\mathbf{s}_0))\right|^2 = O(1)$.

Then, we have the following result.

Proposition E.2. *Suppose that Conditions E.1, E.2, and E.3 hold. Then*

$$Bias\left(\hat{Z}^{BC}_{M,n}(\mathbf{s}_0)\right) = o(1), \text{ and} \tag{E.23}$$

$$Bias\left(\hat{Z}^{BC}_{A,n}(\mathbf{s}_0)\right) = o(1). \tag{E.24}$$

Thus, it follows that under the conditions of Proposition E.2, the bootstrap bias-corrected predictors are asympotically unbiased for any limiting configuration of the data sets $\mathbf{s}_1,\ldots,\mathbf{s}_n$. In particular, unlike the naïve predictor $\hat{Z}_n(\mathbf{s}_0)$ or its analytical bias corrected version $\hat{Z}^C_n(\mathbf{s}_0)$, the asymptotic unbiasedness of $\hat{Z}^{BC}_{\cdot,n}(\mathbf{s}_0)$ holds even under pure increasing domain asymptotic structure, when $\sigma_n \not\to \sigma_0$.

By construction, the "ideal" bootstrap based bias-corrected predictors are exactly unbiased (for any given sample size n). The error term $o(1)$ in (E.23) and (E.24) results from estimating the unknown parameters β and θ by $\hat{\beta}_n$ and $\hat{\theta}_n$, respectively. The magnitude of this term in finite samples is primarily determined by the size of the left side of Condition E.1.

F. Simulations

1. Framework

For the purposes of these simulations, we suggested setting D to be some subset of $\mathbb{R}^2$ containing $\mathbf{s}_0$ near its center. Here, $\mathbf{s}_0$ will be set to be the origin and D will be one of the following subsets of $\mathbb{R}^2$:

- $D_1 := \{(i,j) \in \mathbb{Z}^2 : |i|, |j| \leq 3\}$
- $D_2 := \{(i,j) \in \mathbb{R}^2 : |i|, |j| \leq 3 \text{ and } 2i, 2j \in \mathbb{Z}\}$
- $D_3 := \{(i,j) \in \mathbb{Z}^2 : -4 \leq i, j \leq 5\}$
- $D_4 := \{(i,j) \in \mathbb{R}^2 : -4 \leq i, j \leq 5 \text{ and } 2i, 2j \in \mathbb{Z}\}$
- $D_5 := \{(i,j) \in \mathbb{Z}^2 : -9 \leq i, j \leq 10\}$

These choices for D allow for the observance of the asymptotic behavior of the predictor for both an expanding and filling in observed data set. Notice, for example, that as we move from D_1 to D_3 to D_5, the window expands from a 7×7 grid to a 20×20 grid. Also, D_2 is a filled in version of D_1, and D_4 is a filled in version of D_2, allowing for closer neighbors to be used for prediction.

We then generated $\epsilon(\mathbf{s}_0), \ldots, \epsilon(\mathbf{s}_n)$ from a zero mean, second order stationary Gaussian process with one of the following covariograms:

- Matérn: $C(\mathbf{h}) := \exp(-\theta\|\mathbf{h}\|)$ with $\theta = 0.5$
- Exponential: $C(\mathbf{h}) := \exp(-\theta_1|h_1| - \theta_2|h_2|)$ with $\theta = (0.5, 1)$

The first covariogram is a special case of the Matérn (1960) class of covariograms with variance equal to 1 and a smoothness parameter of 0.5, while the second is an

exponential covariogram. These two covariograms give examples of both isotropic and anisotropic models.

Next, we let $\mu(\cdot;\beta) \equiv 1$. Then,

$$\begin{aligned} Y(\mathbf{s}) &= \mu(\mathbf{s};\beta) + \epsilon(\mathbf{s}) \\ &= 1 + \epsilon(\mathbf{s}). \end{aligned}$$

Because of interest in polynomial, exponential, and power functions (particularly power functions, as methods such as the Box-Cox Transformation are often used in practice for Trans-Gaussian data), the following functions were used to transform $Y(\mathbf{s}_0), \ldots, Y(\mathbf{s}_n)$ to $Z(\mathbf{s}_0), \ldots, Z(\mathbf{s}_n)$:

- $\phi(y) := y^3$
- $\phi(y) := e^y$
- $\phi(y) := e^{2y}$

When analyzing the data, we started by estimating the mean function $\mu(\mathbf{s};\beta)$. When the mean is constant for all $\mathbf{s}$, we simply used the sample mean $\bar{Y}$. We then calculated the estimated residuals and performed Trans-Gaussian Kriging.

All simulations were done for 1000 randomly generated datasets and $B = 500$ bootstrap replicates for each sample.

We compared four predictors: the naïve biased predictor $\hat{Z}(\mathbf{s}_0)$, Cressie's predictor $\hat{Z}^C(\mathbf{s}_0)$ with an additive correction factor based on the second derivative of $\phi(\cdot)$, and two bootstrap predictors, $\hat{Z}_A^{BC}(\mathbf{s}_0)$ with an additive correction factor and $\hat{Z}_M^{BC}(\mathbf{s}_0)$ with a multiplicative correction factor. The predictors were compared using the following measurements:

Table 1. Results for $\phi(y) = y^3$, $C(\mathbf{h}) = \exp(-\theta\|\mathbf{h}\|)$ with $\theta = 0.5$

		D_1	D_2	D_3	D_4	D_5
Ratios	$\hat{Z}$	1.449	1.194	1.376	1.182	1.393
	$\hat{Z}^C$	1.008	0.995	0.945	1.006	0.993
	$\hat{Z}_A^{BC}$	1.007	0.993	0.944	1.008	0.996
	$\hat{Z}_M^{BC}$	1.066	1.007	0.978	1.003	0.982
Biases	$\hat{Z}$	-1.217	-0.661	-0.976	-0.664	-1.136
	$\hat{Z}^C$	-0.031	0.022	0.208	-0.027	0.029
	$\hat{Z}_A^{BC}$	-0.029	0.027	0.210	-0.032	0.017
	$\hat{Z}_M^{BC}$	-0.244	-0.027	0.079	-0.011	0.075
MSPEs	$\hat{Z}$	29.607	16.622	28.769	18.055	29.679
	$\hat{Z}^C$	26.470	15.511	27.199	17.471	28.310
	$\hat{Z}_A^{BC}$	26.575	15.518	27.285	17.548	28.422
	$\hat{Z}_M^{BC}$	26.979	15.147	27.295	17.669	29.310

- Ratio of expected values of an observation and its prediction:

$$\text{Ratio} = \frac{EZ(\mathbf{s}_0)}{E\hat{Z}(\mathbf{s}_0)}$$

- Bias of the predictor:

$$\text{Bias} = E\left[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right]$$

- Mean square prediction error (MSPE):

$$\text{MSPE} = E\left[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right]^2$$

An accurate predictor will have a ratio of expect values close to 1, and a bias close to 0. A smaller MSPE represents less variability in the prediction.

In addition to comparing the performance of the bootstrap predictors to the naïve predictor and Cressie's predictor, actual coverage probabilities for normal, bootstrap-t (D.15), and symmetric bootstrap-t (D.16) prediction intervals were calculated.

2. Simulation Results

Overall, the bootstrap bias correction procedure does tend to decrease bias, as does Cressie's predictor, particularly as the observation grids are filled in (moving from D_1 to D_2 and D_3 to D_4). Which of the two bias correction methods works better depends on the model.

Tables 1 and 2 show the results for $\phi(y) = y^3$ with both the isotropic (Table 1) and anisotropic (Table 2) covariograms. Looking at the ratios of expected values and biases, it is easily seen that all three predictors with a correction factor perform better than the naïve predictor. Here, Cressie's predictor in (C.6) performs as well as the bootstrap predictors.

This is due to the "nice" differentiation behavior of the function $\phi(y) = y^3$. The Taylor expansion yields

$$\begin{aligned}\phi(Y(\mathbf{s}_0)) &= \sum_{i=0}^{\infty} \frac{\phi^{(i)}(\mathbf{s}_0)}{i!}(Y(\mathbf{s}_0) - \mu(\mathbf{s}_0;\beta))^i \\ &= \sum_{i=0}^{3} \frac{\phi^{(i)}(\mathbf{s}_0)}{i!}(Y(\mathbf{s}_0) - \mu(\mathbf{s}_0;\beta))^i.\end{aligned}$$

The same also holds for $\phi\left(\hat{Y}(\mathbf{s}_0)\right)$. When the expected value is taken for a bias calculation, the only terms left are when $i = 0$ and $i = 2$, providing the basis for Cressie's predictor.

Finally, we note that the bias and MSPE tend to increase as the observation grid

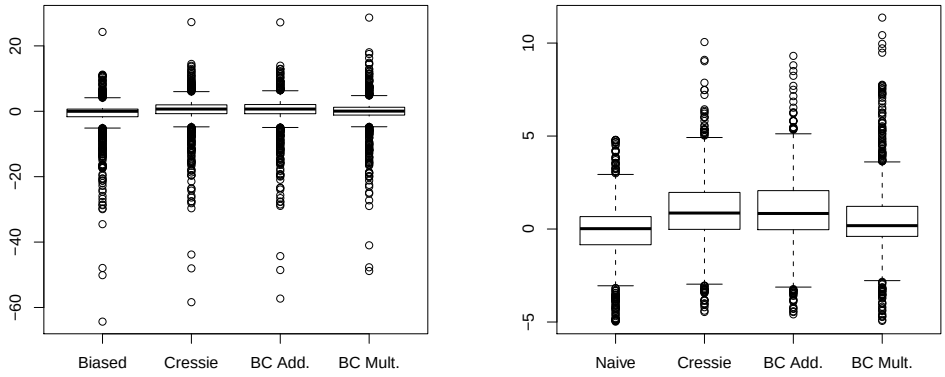
Table 2. Results for $\phi(y) = y^3$, $C(\mathbf{h}) = \exp(-\theta_1|h_1| - \theta_2|h_2|)$ with $\theta = (0.5, 1)'$

		D_1	D_2	D_3	D_4	D_5
Ratios	$\hat{Z}$	1.351	1.133	1.331	1.114	1.386
	$\hat{Z}^C$	0.995	1.006	0.962	1.006	0.986
	$\hat{Z}_A^{BC}$	0.998	1.006	0.961	1.003	0.986
	$\hat{Z}_M^{BC}$	1.027	1.002	0.960	1.009	1.005
Biases	$\hat{Z}$	-0.996	-0.464	-0.926	-0.405	-1.011
	$\hat{Z}^C$	0.019	-0.025	0.145	-0.011	0.052
	$\hat{Z}_A^{BC}$	0.008	-0.024	0.151	-0.015	0.050
	$\hat{Z}_M^{BC}$	-0.099	-0.006	0.154	-0.037	-0.017
MSPEs	$\hat{Z}$	26.308	10.236	36.998	10.718	25.210
	$\hat{Z}^C$	24.395	9.822	35.176	10.521	24.094
	$\hat{Z}_A^{BC}$	24.578	9.855	35.303	10.561	24.156
	$\hat{Z}_M^{BC}$	26.157	9.749	32.397	11.109	23.285

is filled in, and that the MSPEs are comparable for all four predictors, which was seen in all simulations.

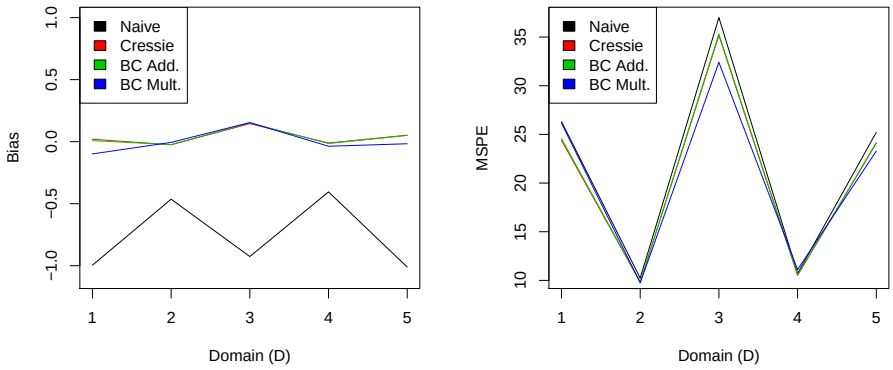
Figure 1 shows the biases for 1000 realizations of all four predictors for the isotropic covariogram. It is easy to see that the bias correction factors reduce the skew in the distribution of the biases, bringing the outliers closer to the boxes. Recall that the most bias arises when $Y(\mathbf{s}_0)$ is far from its mean, and this effect increases as the derivative of $\phi(\cdot)$ becomes more positive or negative.

Figure 2 demonstrates the improvement in bias as a result of the corrected predictors, without a dramatic increase in MSPE. With a transformation such as $\phi(y) = y^3$, Cressie's predictor is as effective as the bootstrap predictors, as we see in Tables 1



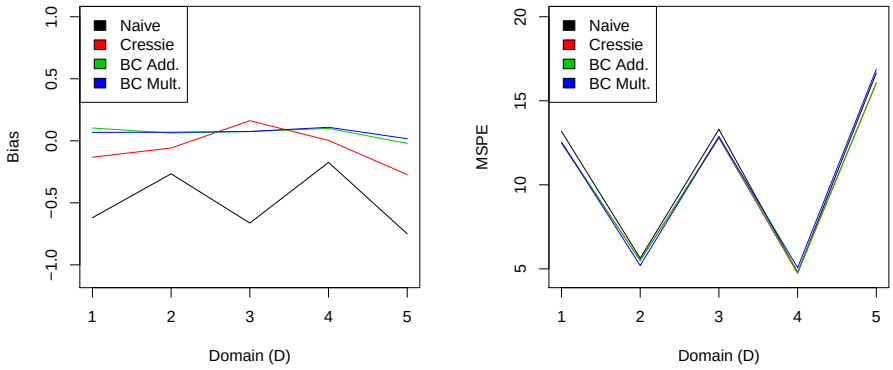
(a) (b)

Figure 1. (a) Biases for 1000 realizations of the four predictors, where $\phi(y) = y^3$, $C(\mathbf{h}) = \exp(-\theta\|\mathbf{h}\|)$, $\theta = 0.5$, and the observation grid is D_1; (b) Bias of observations where original bias is between -5 and 5.



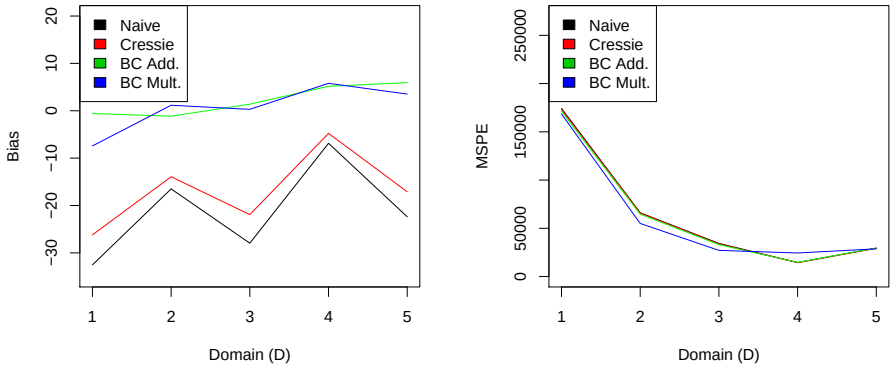
(a) (b)

Figure 2. (a) Biases for each predictor for D_1 through D_5; (b) MSPEs for each predictor for D_1 through D_5. Here, $\phi(y) = y^3$, $C(\mathbf{h}) = \exp(-\theta_1|h_1| - \theta_2|h_2|)$, $\theta = c(0.5, 1)'$.



(a) (b)

Figure 3. (a) Biases for each predictor for D_1 through D_5; (b) MSPEs for each predictor for D_1 through D_5. Here, $\phi(y) = e^y$, $C(\mathbf{h}) = \exp(-\theta_1|h_1| - \theta_2|h_2|)$, $\theta = c(0.5, 1)'$.



(a) (b)

Figure 4. (a) Biases for each predictor for D_1 through D_5; (b) MSPEs for each predictor for D_1 through D_5. Here, $\phi(y) = e^{2y}$, $C(\mathbf{h}) = \exp(-\theta_1|h_1| - \theta_2|h_2|)$, $\theta = c(0.5, 1)'$.

and 2.

However, it is clear from Figures 3 and 4 that Cressie's predictor begins to fail in correcting bias compared to the bootstrap predictors for exponential transformation functions, as implied by comparing the biases of the three predictors in Propositions E.1 and E.2.

Table 3. Results for $\phi(y) = e^y$, $C(\mathbf{h}) = \exp(-\theta\|\mathbf{h}\|)$ with $\theta = 0.5$

		D_1	D_2	D_3	D_4	D_5
Ratios	$\hat{Z}$	1.219	1.095	1.209	1.108	1.254
	$\hat{Z}^C$	1.037	1.012	1.037	1.034	1.096
	$\hat{Z}_A^{BC}$	0.971	0.971	0.972	0.999	1.030
	$\hat{Z}_M^{BC}$	0.986	0.972	0.975	0.993	1.024
Biases	$\hat{Z}$	-0.780	-0.396	-0.730	-0.451	-0.933
	$\hat{Z}^C$	-0.153	-0.052	-0.151	-0.152	-0.404
	$\hat{Z}_A^{BC}$	0.130	0.136	0.122	0.007	-0.133
	$\hat{Z}_M^{BC}$	0.060	0.131	0.107	0.030	-0.107
MSPEs	$\hat{Z}$	15.687	7.684	15.270	9.251	20.070
	$\hat{Z}^C$	14.466	7.484	14.542	8.923	19.205
	$\hat{Z}_A^{BC}$	14.539	7.564	14.694	8.942	19.027
	$\hat{Z}_M^{BC}$	14.265	7.723	14.645	8.636	17.822

This is further illustrated by Tables 3 and 4, which show the results for $\phi(y) = e^y$. Here, while the later terms in the Taylor expansion do not go to zero, they will tend to get smaller as $\phi^{(i)}(y) = e^y$ becomes smaller compared to $i!$. Cressie's predictor is still an improvement over the naïve predictor, but it is not comparable to the bootstrap predictors.

Table 4. Results for $\phi(y) = e^y$, $C(\mathbf{h}) = \exp(-\theta_1|h_1| - \theta_2|h_2|)$ with $\theta = (0.5, 1)'$

		D_1	D_2	D_3	D_4	D_5
Ratios	$\hat{Z}$	1.172	1.063	1.183	1.041	1.195
	$\hat{Z}^C$	1.032	1.013	1.039	0.999	1.063
	$\hat{Z}_A^{BC}$	0.976	0.986	0.983	0.978	1.005
	$\hat{Z}_M^{BC}$	0.984	0.985	0.983	0.976	0.996
Biases	$\hat{Z}$	-0.620	-0.266	-0.663	-0.174	-0.750
	$\hat{Z}^C$	-0.132	-0.058	-0.162	0.003	-0.272
	$\hat{Z}_{Add.}^{BC}$	0.103	0.062	0.075	0.100	-0.021
	$\hat{Z}_{Mult.}^{BC}$	0.068	0.069	0.075	0.109	0.017
MSPEs	$\hat{Z}$	13.179	5.630	13.303	4.787	16.657
	$\hat{Z}^C$	12.470	5.500	12.786	4.741	16.066
	$\hat{Z}_A^{BC}$	12.537	5.482	12.856	4.816	16.044
	$\hat{Z}_M^{BC}$	12.546	5.201	12.874	5.068	16.870

Table 5 shows the results for $\phi(y) = e^{2y}$ with the isotropic covariogram. Here, we see that as the transformation function becomes more extreme, the estimators become less reliable.

The naïve predictor performs poorly, and Cressie's predictor performs significantly worse than the bootstrap predictors. With this transformation function, we can no longer count on the later terms in the in Taylor expansion to be small, particularly when $\phi^{(i)}(y) = 2^i e^{2y} > i!$.

Compared to the earlier models, the accuracy of the bootstrap predictors has begun to deteriote, although they both still outperform the existing predictors. Interestingly, the MSPE of the multiplicative predictor appears to be more volatile than

Table 5. Results for $\phi(y) = e^{2y}$, $C(\mathbf{h}) = \exp(-\theta\|\mathbf{h}\|)$ with $\theta = 0.5$

		D_1	D_2	D_3	D_4	D_5
Ratios	$\hat{Z}$	2.229	1.635	1.940	1.487	2.057
	$\hat{Z}^C$	1.544	1.433	1.443	1.329	1.615
	$\hat{Z}_A^{BC}$	0.694	0.918	0.720	0.878	0.882
	$\hat{Z}_M^{BC}$	0.841	0.856	0.749	0.911	0.902
Biases	$\hat{Z}$	-27.491	-25.928	-21.065	-17.288	-24.726
	$\hat{Z}^C$	-17.580	-20.182	-13.351	-13.070	-18.321
	$\hat{Z}_A^{BC}$	21.923	5.892	16.909	7.351	6.460
	$\hat{Z}_M^{BC}$	9.402	11.235	14.550	5.180	4.214
MSPEs	$\hat{Z}$	21387	74594	29472	17504	19801
	$\hat{Z}^C$	19116	72271	29056	17271	19477
	$\hat{Z}_A^{BC}$	20803	68583	31012	17778	19562
	$\hat{Z}_M^{BC}$	29172	55962	60113	18003	19140

the MSPE of the additive predictor.

For the anisotropic covariogram, similar results are seen in Table 6, which corresponds to Figure 4.

Table 7 gives the coverage probabilities for the prediction intervals corresponding to the simulation in Table 1. We compare the normal prediction interval in (D.14), the bootstrap-t interval in (D.15), and the symmetric bootstrap-t interval in (D.16).

The three prediction interval methods perform reasonably well, with the normal and symmetric bootstrap-t intervals performing best. This is not surprising, given that in these simulations, we know we are in fact dealing with a Gaussian process, so we expect the normal prediction interval to perform very well. The performance of

Table 6. Results for $\phi(y) = e^{2y}$, $C(\mathbf{h}) = \exp(-\theta_1|h_1| - \theta_2|h_2|)$ with $\theta = (0.5, 1)'$

		D_1	D_2	D_3	D_4	D_5
Ratios	$\hat{Z}$	2.379	1.389	2.171	1.182	1.987
	$\hat{Z}^C$	1.875	1.310	1.733	1.120	1.612
	$\hat{Z}_A^{BC}$	1.010	1.020	0.974	0.896	0.884
	$\hat{Z}_M^{BC}$	1.152	0.981	0.994	0.885	0.927
Biases	$\hat{Z}$	-32.542	-16.496	-27.957	-6.868	-22.382
	$\hat{Z}^C$	-26.199	-13.929	-21.924	-4.767	-17.108
	$\hat{Z}_A^{BC}$	-0.576	-1.141	1.390	5.164	5.940
	$\hat{Z}_M^{BC}$	-7.428	1.166	0.300	5.792	3.526
MSPEs	$\hat{Z}$	174284	66089	34371	14490	29520
	$\hat{Z}^C$	173045	65765	33763	14442	29284
	$\hat{Z}_A^{BC}$	171144	64592	33067	14720	29599
	$\hat{Z}_M^{BC}$	168512	55034	27128	24423	28719

the bootstrap intervals indicates that they could be useful with real world data.

G. Oklahoma Climatology Data

For a real world data example, we look at a problem relating weather and energy consumption. Define $Z(\mathbf{s})$ to be the "heating degree days" (HDDs) at a location $\mathbf{s}$, depending on the average daily outside air temperature $T(\mathbf{s})$. HDD depends on a base temperature, often taken to be 65° Fahrenheit, such that

$$Z(\mathbf{s}) = \begin{cases} 65^\circ \text{ F} - T(\mathbf{s}), & \text{for } T(\mathbf{s}) < 65^\circ \text{ F} \\ 0 & \text{otherwise.} \end{cases} . \tag{G.25}$$

Table 7. Prediction interval coverage probabilities for $\phi(y) = y^3$, $C(\mathbf{h}) = \exp(-\theta\|\mathbf{h}\|)$ with $\theta = 0.5$

		D_1	D_2	D_3	D_4	D_5
Normal	90%	0.881	0.892	0.891	0.876	0.886
	95%	0.945	0.944	0.939	0.935	0.949
	99%	0.984	0.982	0.988	0.983	0.989
Bootstrap-t	90%	0.787	0.794	0.794	0.783	0.798
	95%	0.884	0.885	0.894	0.880	0.887
	99%	0.970	0.966	0.975	0.971	0.982
Symmetric	90%	0.878	0.888	0.888	0.880	0.886
Bootstrap-t	95%	0.945	0.944	0.940	0.938	0.940
	99%	0.984	0.977	0.985	0.983	0.990

An interpretation of HDD is as follows: if one location $\mathbf{s}_1$ has an HDD value of 20 and another location $\mathbf{s}_2$ has an HDD value of 40, a building at $\mathbf{s}_2$ will require twice as much energy to heat as a similar building at $\mathbf{s}_1$. As implied by the (G.25), HDDs are calculated on a daily basis. For more on HDDs, see Ristinen and Kraushaar (2006). Cooling degree days (CDDs) is a similar measurement regarding the amount of energy required to cool a building. Generally, HDDs are higher in the winter and CDDs in the summer.

Often, HDDs are accumulated over weeks, months, or seasons, to study energy usage over an extended period of time, but we will focus on a single day of observations. Figure 5 shows the heating degree days for January 1, 2009 at 116 observation sites in Oklahoma, provided by the Oklahoma Climatological Survey (2010), a joint venture between the University of Oklahoma and Oklahoma State University that

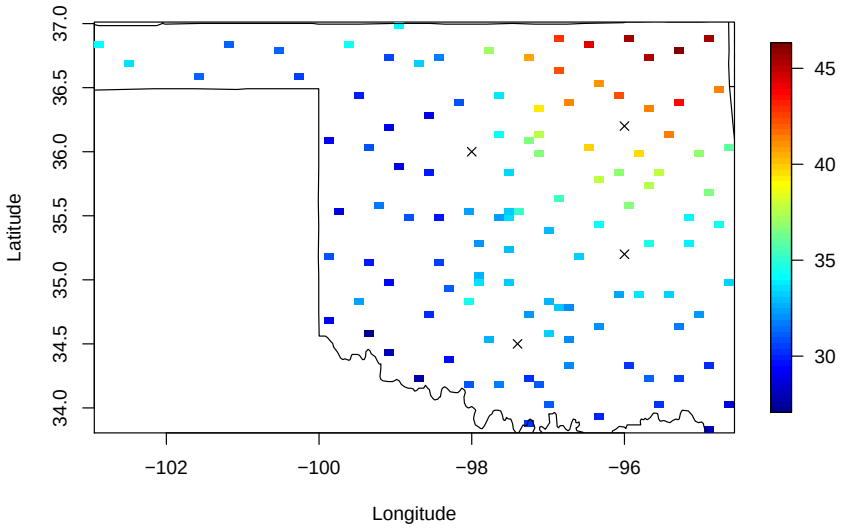
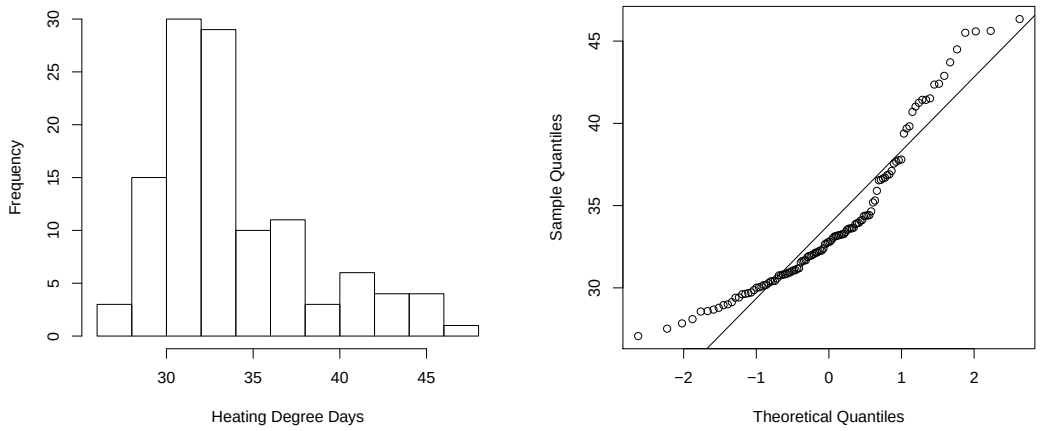


Figure 5. HDD observations in Oklahoma, Jan. 1, 2009, with multiplication symbols (×) representing prediction points

tracks weather throughout the state. Suppose we wish to predict the daily HDD measurement at the points (-98,36), (-96,36.2), (-96,35.2), and (-97.4). These sites are marked by a "×" symbol in Figure 5.

There appears to be a definite relationship between the spatial location and the HDD measurement, with a higher number of HDDs corresponding to the sites in the northeastern part of the state. Further, Figure 6 shows that the distribution HDDs is highly skewed. The next step is to identify a suitable transformation function $\phi(\cdot)$.

We begin by using the transformation suggested by Box and Cox (1964), which has proven to be suitable for positive variables. If negative observations were present, the Box-Cox transformation fails, and another method, such as the one suggested by Yeo and Johnson (2000), should be tried. The Box-Cox transformation function is



(a) (b)

Figure 6. (a) Histogram and (b) normal quantile plot showing that the distribution of HDDs is skewed.

given by:

$$g_\lambda(x_i) = \begin{cases} \frac{x_i^\lambda - 1}{\lambda(GM(\mathbf{x}))^{\lambda-1}} & \text{if } \lambda \neq 0, \\ GM(\mathbf{x}) \log(x_i) & \text{if } \lambda = 0. \end{cases}$$

where $\mathbf{x} = \{x_1, \ldots, x_n\}$ and $GM(\mathbf{x})$ is the geometric mean of $\mathbf{x}$. The geometric mean is a constant that scales the data, and is often omitted. For the HDD data, the optimal value of λ is -3.005236. The omission of the geometric mean for this data set leads to transformed observations with a small variance, causing computational problems later in the analysis. Therefore, we will include the geometric mean in $g_\lambda(\cdot)$.

Define $\mathbf{x} \equiv \mathbf{Z} = \{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$, $Y(\mathbf{s}) = g_\lambda(Z(\mathbf{s})) \equiv \phi^{-1}(Z(\mathbf{s}))$, and $\kappa \equiv (GM(\mathbf{Z}))^{\lambda-1}$. It is easy to check that

$$\begin{aligned} Z(\mathbf{s}) &= \phi(Y(\mathbf{s})) \\ &= g_\lambda^{-1}(Y(\mathbf{s})) \end{aligned}$$

$$= \quad (y\lambda\kappa + 1)^{1/\lambda},$$

where $\lambda = -3.005236$.

The distribution of transformed HDD observations is shown in Figure 7. It is clear that normality is plausible for $Y(\mathbf{s})$.

Table 8. Linear model summary for HDD data

Effect	Coefficient	p-value
Intercept	429250.2	$< 2 \times 10^{-16}$
Latitude	-11.125688	0.01626
Longitude	37.22961	0.00355
Interaction	0.3514275	0.00728

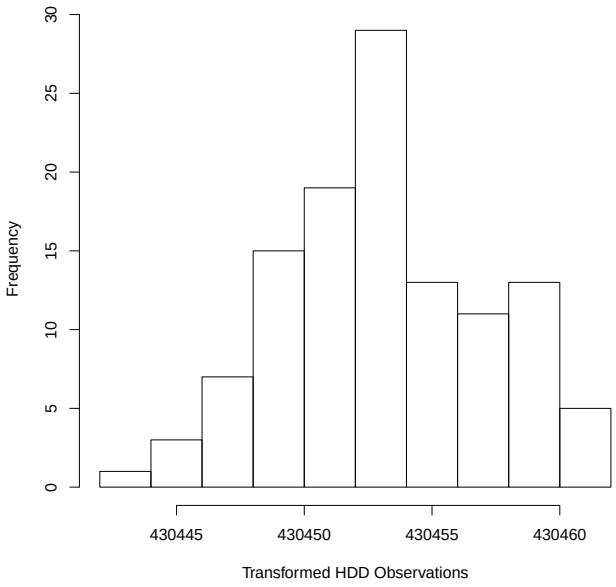
Next, we calculate $\hat{\epsilon}(\mathbf{s}_1), \ldots, \hat{\epsilon}(\mathbf{s}_n)$, the estimated residuals, by estimating the mean parameter β. We assume a mean function depending on the latitude and longitude, such that

$$\begin{aligned} Y(\mathbf{s}) &= \mu(\mathbf{s};\beta) + \epsilon(\mathbf{s}) \\ &= \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + \beta_3 X_1(\mathbf{s}) X_2(\mathbf{s}) + \epsilon(\mathbf{s}), \end{aligned}$$

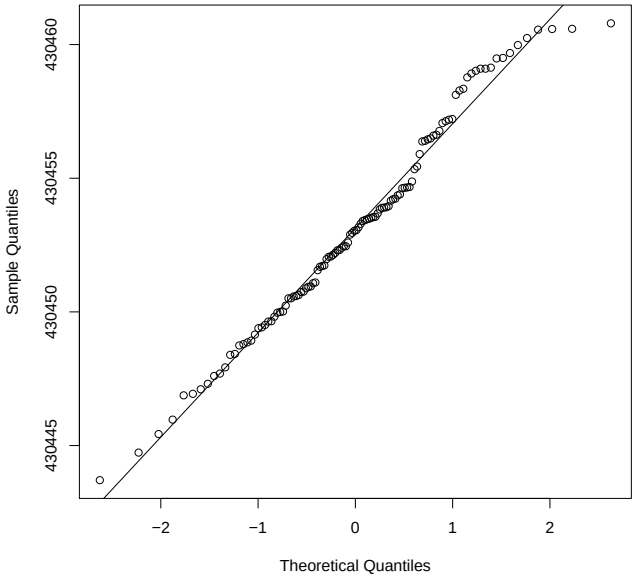
where $X_1(\mathbf{s})$ is the longitude of $\mathbf{s}$, $X_2(\mathbf{s})$ is the latitude, and $\mu(\mathbf{s};\beta) = \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + \beta_3 X_1(\mathbf{s}) X_2(\mathbf{s})$. The estimated coefficients and p-values from a least squares regression model are given in Table 8, with all coefficients significantly different from 0.

For our covariogram, we choose

$$C(\mathbf{h}) = \theta_1 \exp(-\|\mathbf{h}\|/\theta_2),$$



(a)



(b)

Figure 7. (a) Histogram and (b) normal quantile plot showing that the distribution of transformed HDDs is approximately normal.

where $\|\mathbf{h}\|$ is the Euclidean distance between two locations. θ_1 represents the marginal variance (σ^2) of $\epsilon(\mathbf{s})$, whereas θ_2 represents the range of correlation. These parameters were estimated using weighted least squares. This covariogram is a member of the Matérn (1960) class, with fixed smoothness parameter equal to 0.5, which is believed to have considerable use in practice (cf. Stein (1999), Schabenberger and Gotway (2005)).

Table 9. Predictors for HDD data

	(-98,36)	(-96,36.2)	(-96,35.2)	(-97.4,34.5)
$\hat{Z}(\mathbf{s}_0)$	25.15465	28.23200	26.25444	25.41312
$\hat{Z}^C(\mathbf{s}_0)$	25.21641	28.27352	26.30458	25.46030
$\hat{Z}_A^{BC}(\mathbf{s}_0)$	25.02503	28.03837	26.19769	25.38501
$\hat{Z}_M^{BC}(\mathbf{s}_0)$	25.02985	28.03462	26.19745	25.38438

WLS procedures give $\hat{\theta} = (9.603985, 2.683276)'$. We now have all information needed to calculate the predictors in (C.5), (C.6), (C.9), and (C.10). The predictors for each of the four prediction points are given in Table 9. If each predictor is considered to be an estimate of the amount of energy needed to heat a home at a given location, $\hat{Z}^C(\mathbf{s}_0)$ estimates that between 0.14 and 0.25% more energy is required than $\hat{Z}(\mathbf{s}_0)$ does. On the other hand, the bootstrap predictors say that between 0.11 and 0.7% less energy is required than $\hat{Z}(\mathbf{s}_0)$.

H. Proofs

Here, we once again shall write $\hat{\theta} = \hat{\theta}_n$, $\hat{\beta} = \hat{\beta}_n$, $\gamma = \gamma_n$, $\Sigma = \Sigma_n$, $\tilde{Z}(\mathbf{s}_0) = \tilde{Z}_n(\mathbf{s}_0)$, $\hat{Z}(\mathbf{s}_0) = \hat{Z}_n(\mathbf{s}_0)$, etc. Denote the k-th derivative of $\phi(\cdot)$ by $\phi^{(k)}(\cdot)$, $k \geq 1$. We will also use $\phi'(\cdot)$, $\phi''(\cdot)$, $\phi'''(\cdot)$ for $\phi^{(1)}(\cdot)$, $\phi^{(2)}(\cdot)$, and $\phi^{(3)}(\cdot)$, respectively. Let $\mathbb{1}(\cdot)$ denote the

indicator function, with $\mathbb{1}(S) = 0$ or 1 accordingly as a statement S is false or true. Let k, $k(\cdot)$ denote generic constants in $(0, \infty)$ that may depend on some arguments (if any) but not on n.

Proof of Proposition E.1. First, consider (E.21). Let $N \sim N(0,1)$. Then, it is easy to check that for any $k \geq 1$,

$$E|N|^k = \frac{\left(\sqrt{2}\right)^k}{\sqrt{\pi}} \Gamma\left(\frac{k+1}{2}\right),$$

where $\Gamma(\cdot)$ is the Gamma function. Note that by Jensen's inequality, for any $a, b \in (0, \infty)$,

$$\int_0^\infty x^a e^{-x} dx \leq \left(\int_0^\infty x^{a+b} e^{-x} dx\right)^{\frac{a}{a+b}}.$$

Thus,

$$\Gamma(a+1) \leq [\Gamma(a+b-1)]^{\frac{a}{a+b}}$$

for $a > 0$ and $b \geq 0$. Further, for any integer $r \geq 2$, it is easy to check that

$$\Gamma\left(r - \frac{1}{2}\right) \geq \frac{\Gamma(r)\Gamma(3/2)}{r-1}.$$

Hence, by (E.20), it follows that for $x \in \{\sigma_n, \sigma_0\}$,

$$\begin{aligned}
&\sum_{k=1}^{\infty} |d_k| 2^{k/2} \Gamma\left(\frac{k+1}{2}\right) x^k \\
&= \left[\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} |d_j||d_k| 2^{(k+j)/2} \right. \\
&\quad \left. \times \Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{j+1}{2}\right) x^{j+k}\right]^{1/2} \\
&\leq \left[\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} |d_j||d_k| 2^{(k+j)/2} \right. \\
&\quad \left. \times \Gamma\left(\frac{k-1}{2} + \frac{j-1}{2} + 1\right)^{\frac{k-1}{k+j-2}}\right.
\end{aligned}$$

$$\times\Gamma\left(\frac{k-1}{2}+\frac{j-1}{2}+1\right)^{\frac{j-1}{k+j-2}} x^{j+k}\Bigg]^{1/2}$$

$$\leq \left[\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}|d_j||d_k|2^{(k+j)/2}\frac{\Gamma\left(\frac{k+j-1}{2}\right)(k+j)x^{k+j}}{\Gamma\left(\frac{3}{2}\right)}\right]^{1/2}$$

$$= O(1), \text{ as } n\to\infty$$

Now using Fubini's theorem, we get

$$\begin{aligned}
&\text{Bias}\left(\tilde{Z}(\mathbf{s}_0)\right)\\
&= E[\tilde{Z}(\mathbf{s}_0)-Z(\mathbf{s}_0)]\\
&= E\left[\phi\left(\mu(\mathbf{s}_0;\beta)+\tilde{\epsilon}_n(\mathbf{s}_0)\right)-\phi(\mu(\mathbf{s}_0;\beta)+\epsilon(\mathbf{s}_0))\right]\\
&= E\left(\sum_{k\geq 0}\frac{\phi^{(k)}(\mu(\mathbf{s}_0;\beta))}{k!}\tilde{\epsilon}_n(\mathbf{s}_0)^k\right)-E\left(\sum_{k\geq 0}\frac{\phi^{(k)}(\mu(\mathbf{s}_0;\beta))}{k!}\epsilon(\mathbf{s}_0)^k\right)\\
&= \sum_{k=1}^{\infty}\frac{\phi^{(2k)}(\mu(\mathbf{s}_0;\beta))}{(2k)!}\left\{E(\tilde{\epsilon}_n(\mathbf{s}_0)^{2k})-E(\epsilon(\mathbf{s}_0)^{2k})\right\}\\
&= \sum_{k=1}^{\infty}\frac{\phi^{(2k)}(\mu(\mathbf{s}_0;\beta))}{(2k)!}\left[\frac{(2k)!}{k!2^k}(\sigma_n^{2k}-\sigma_0^{2k})\right]\\
&= \sum_{k=1}^{\infty}\frac{\phi^{(2k)}(\mu(\mathbf{s}_0;\beta))}{k!2^k}(\sigma_n^{2k}-\sigma_0^{2k}),
\end{aligned}\tag{H.26}$$

where $\epsilon(\mathbf{s}_0)\sim N(0,\sigma_0^2)$ and $\tilde{\epsilon}_n(\mathbf{s}_0)\sim N(0,\sigma_n^2)$. By (H.26), we have

$$E\sum_{k=1}^{\infty}|d_k||w_k|<\infty$$

for $w_k\in\left\{\epsilon(\mathbf{s}_0)^k,\tilde{\epsilon}_n(\mathbf{s}_0)^k\right\}$, $k\geq 1$. Also, from (E.19), it follows that $\phi(\cdot)$ is infinitely differentiable and that $d_k=\phi^{(k)}(\mu(\mathbf{s}_0;\beta))/k!$, $k\geq 1$.

Next, fix $\delta\in(0,1/2)$. Let

$$A_n=\left\{\left|\mu(\mathbf{s}_0;\hat{\beta}_n)-\mu(\mathbf{s}_0;\beta)\right|\leq\delta\right\}\cap\left\{|\hat{\epsilon}_n(\mathbf{s}_0)-\tilde{\epsilon}_n(\mathbf{s}_0)|\leq\delta\right\},$$

for $n \geq 1$. Then,

$$\begin{aligned}
&\left|E\left[\hat{Z}(\mathbf{s}_0) - \tilde{Z}(\mathbf{s}_0)\right] \mathbb{1}(A_n)\right| \\
&= \left|E\left\{\phi\left(\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) + \hat{\epsilon}_n(\mathbf{s}_0)\right) - \phi\left(\mu(\mathbf{s}_0;\beta) + \tilde{\epsilon}_n(\mathbf{s}_0)\right)\right\} \mathbb{1}(A_n)\right| \\
&= \Bigg|E\Bigg\{\left[\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) - \mu(\mathbf{s}_0;\beta) + \hat{\epsilon}_n(\mathbf{s}_0) - \tilde{\epsilon}_n(\mathbf{s}_0)\right] \\
&\quad \times \int_0^1 (1-u)\phi''\left[u\left(\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) + \hat{\epsilon}_n(\mathbf{s}_0)\right) + (1-u)\left(\mu(\mathbf{s}_0;\beta)\right.\right. \\
&\quad \left.\left. + \tilde{\epsilon}_n(\mathbf{s}_0)\right) \right] du\Bigg\} \mathbb{1}(A_n)\Bigg| \\
&\leq \sqrt{2}\Bigg[E\left(\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) - \mu(\mathbf{s}_0,\beta)\right)^2 \mathbb{1}(A_n) \\
&\quad + E\left(\hat{\epsilon}_n(\mathbf{s}_0) - \tilde{\epsilon}_n(\mathbf{s}_0)\right)^2 \mathbb{1}(A_n) \Bigg]^{1/2} \\
&\quad \times \Bigg[\int_0^1 E\left\{\phi'\left(\mu(\mathbf{s}_0;\beta) + \tilde{\epsilon}_n(\mathbf{s}_0) + u\left[\{\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) - \mu(\mathbf{s}_0;\beta)\}\right.\right.\right. \\
&\quad \left.\left.\left. + \{\hat{\epsilon}_n(\mathbf{s}_0) - \tilde{\epsilon}_n(\mathbf{s}_0)\}\right]\right) \mathbb{1}(A_n)\right\}^2 du \Bigg]^{1/2}.
\end{aligned}$$

Note that by the Dominated Convergence Theorem (DCT),

$$E\left[\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) - \mu(\mathbf{s}_0;\beta)\right]^2 \mathbb{1}(A_n) \to 0 \text{ as } n \to \infty.$$

Similarly,

$$E(\hat{\epsilon}_n(\mathbf{s}_0) - \tilde{\epsilon}_n(\mathbf{s}_0))^2 \mathbb{1}(A_n) \to 0 \text{ as } n \to \infty.$$

Further, by (E.19), uniformly in $u \in (0,1)$,

$$\begin{aligned}
&E\left\{\phi'\left(\mu(\mathbf{s}_0;\beta) + \tilde{\epsilon}_n(\mathbf{s}_0) + u\left[\left\{\mu(\mathbf{s}_0;\beta) - \mu\left(\mathbf{s}_0;\hat{\beta}_n\right)\right\}\right.\right.\right. \\
&\qquad \left.\left.\left. + \{\hat{\epsilon}_n(\mathbf{s}_0) - \tilde{\epsilon}_n(\mathbf{s}_0)\}\right]\right) \mathbb{1}(A_n)\right\}^2 \\
&\leq E\left[\sum_{k=1}^{\infty} k|d_k| \left\{|\tilde{\epsilon}_n(\mathbf{s}_0)| + 2\delta\right\}^{k-1}\right]^2 \\
&\leq \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} kj|d_k d_j| E\left(|\tilde{\epsilon}_n(\mathbf{s}_0)| + 2\delta\right)^{k+j-2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} kj|d_k d_j| 2^{k+j-2}\left\{E|\tilde{\epsilon}_n(\mathbf{s}_0)|^{k+j-2} + (2\delta)^{k+j-2}\right\} \\
&\leq \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} kj|d_k d_j| 2^{(k+j-2)/2}\Gamma\left(\frac{k+j-1}{2}\right)\sigma_n^{j+k-2} + k(\delta) \\
&= O(1).
\end{aligned}
$$

Hence, it follows that

$$
\left|E\left[\hat{Z}_n(\mathbf{s}_0) - \tilde{Z}_n(\mathbf{s}_0)\right]\mathbb{1}(A_n)\right| = o(1). \tag{H.27}
$$

Also, by the Cauchy-Schwarz inequality and Condition E.1,

$$
\begin{aligned}
\left|E\left[\hat{Z}_n(\mathbf{s}_0) - \tilde{Z}_n(\mathbf{s}_0)\right]\mathbb{1}(A_n^C)\right| &\leq \left[E\left(\hat{Z}_n(\mathbf{s}_0)\right)^2\right]^{1/2}\left[P(A_n^C)\right]^{1/2} \\
&\quad + \left[E\left(\tilde{Z}_n(\mathbf{s}_0)\right)^2\right]^{1/2}\left[P(A_n^C)\right]^{1/2} \\
&= o(1).
\end{aligned} \tag{H.28}
$$

From (H.26)-(H.28), (E.21) follows. The proof of (E.22) is similar and hence, is omitted. □

Proof of Proposition E.2. First, consider the multiplicative bias-corrected predictor $\hat{Z}_{M,n}^{BC}(\mathbf{s}_0)$ and its ideal version $\tilde{Z}_{M,n}^{BC}(\mathbf{s}_0)$. Note that, with $c_n \equiv E\tilde{Z}_n(\mathbf{s}_0)/EZ(\mathbf{s}_0)$,

$$
\begin{aligned}
E\left(\tilde{Z}_{M,n}^{BC}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right) &= cE\tilde{Z}_n(\mathbf{s}_0) - EZ(\mathbf{s}_0) \\
&= 0,
\end{aligned}
$$

so that the ideal version $\tilde{Z}_{M,n}^{BC}(\mathbf{s}_0)$ is unbiased. It is easy to see that

$$
\begin{aligned}
&\text{Bias}\left(\hat{Z}_{M,n}^{BC}(\mathbf{s}_0)\right) \\
&= E\left(\hat{Z}_{M,n}^{BC}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right) \\
&= E\hat{c}_n\hat{Z}_n(\mathbf{s}_0) - EZ(\mathbf{s}_0)
\end{aligned}
$$

$$
\begin{aligned}
&= E\left\{\left(E_*Z^*(\mathbf{s}_0) - EZ(\mathbf{s}_0)\right)\hat{Z}_n(\mathbf{s}_0)/E_*\hat{Z}_n^*(\mathbf{s}_0)\right\} \\
&\quad -\frac{EZ(\mathbf{s}_0)}{E\hat{Z}_n(\mathbf{s}_0)}E\left[\left\{E_*\hat{Z}_n^*(\mathbf{s}_0) - E\hat{Z}_n(\mathbf{s}_0)\right\}\hat{Z}_n(\mathbf{s}_0)/E_*\hat{Z}_n^*(\mathbf{s}_0)\right] \\
&\equiv I_{1n} + I_{2n} \text{ (say).}
\end{aligned}
$$

Note that $\hat{\epsilon}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_{in}(\hat{\theta}_n)\epsilon(\mathbf{s}_i)$, where $\boldsymbol{\lambda}_n(\theta) = (\lambda_{1n}(\theta), \ldots, \lambda_{nn}(\theta))'$ is as given by Condition E.3. Note that

$$
\begin{aligned}
&E_*\hat{Z}_n^*(\mathbf{s}_0) \\
&= E_*\phi\left(\mu(\mathbf{s}_0;\beta_n^*) + \sum_{i=1}^n \lambda_{in}(\theta_n^*)\left[\epsilon^*(\mathbf{s}_i) - \left\{\mu(\mathbf{s}_i;\beta_n^*) - \mu\left(\mathbf{s}_i;\hat{\beta}_n\right)\right\}\right]\right) \\
&= E_*\phi\left(\mu\left(\mathbf{s}_0;\hat{\beta}_n\right) + \sum_{i=1}^n \lambda_{in}\left(\hat{\theta}_n\right)\epsilon^*(\mathbf{s}_i)\right. \\
&\quad + \sum_{i=1}^n \left(\lambda_{in}(\theta_n^*) - \lambda_{in}\left(\hat{\theta}_n\right)\right)\epsilon^*(\mathbf{s}_i) \\
&\quad - \sum_{i=1}^n \lambda_{in}(\theta_n^*)\left\{\mu(\mathbf{s}_i;\beta_n^*) - \mu\left(\mathbf{s}_i;\hat{\beta}_n\right)\right\} \\
&\quad \left. + \left[\mu(\mathbf{s}_0;\beta_n^*) - \mu\left(\mathbf{s}_0;\hat{\beta}_n\right)\right] \quad \right) \\
&= g\left(\hat{\beta}_n, \tau_n\left(\hat{\theta}_n\right)\right) + R_{1n},
\end{aligned}
$$

where $\tau_n^2(\theta) \equiv E_\theta\left(\sum_{i=1}^n \lambda_{in}(\theta)\epsilon(\mathbf{s}_i)\right)^2 = E_\theta\tilde{Z}(\mathbf{s}_0)^2$, and where, on the set

$$
\left\{\left\|\hat{\theta}_n - \theta\right\| < \delta_0/2\right\},
$$

for any $\delta \in (0, \delta_0/2)$,

$$
\begin{aligned}
&|R_{1n}| \\
&\leq E_*\left[\left(\sum_{i=1}^n k_0\alpha_i\left\{\left\|\theta_n^* - \hat{\theta}_n\right\||\epsilon^*(\mathbf{s}_i)| + k_1\left\|\beta_n^* - \hat{\beta}_n\right\|\right\}\right. \\
&\quad \left. + k_1\left\|\beta_n^* - \hat{\beta}_n\right\|\frac{1}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\times\Phi^{(1)}\left(\left|\mu\left(\mathbf{s}_0;\hat{\beta}_n\right)\right|+\left|\sum_{i=1}^{n}\lambda_{in}\left(\hat{\theta}_n\right)\epsilon^*(\mathbf{s}_i)\right|\right.\\
&\left.+\delta k_0\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|+k\left\|\beta_n^*-\hat{\beta}_n\right\|\right)\\
&\times\mathbb{1}\left(\left\|\theta_n^*-\hat{\theta}_n\right\|\leq\delta\right)\quad\Big]\\
&+E_*\left[\left(2k_0\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|+k\left\|\beta_n^*-\hat{\beta}_n\right\|\right)\right.\\
&\times\Phi^{(1)}\left(\left|\mu\left(\mathbf{s}_0;\hat{\beta}_n\right)\right|+3k_0\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|+k\left\|\beta_n^*-\hat{\beta}_n\right\|\right)\\
&\left.\times\mathbb{1}\left(\left\|\theta_n^*-\hat{\theta}_n\right\|>\delta\right)\quad\right]\\
\leq\;&\left[\delta k_0\left\{E_*\left(\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|\right)^2\right\}^{1/2}+k\left(E_*\left\|\beta_n^*-\hat{\beta}_n\right\|^2\right)^{1/2}\right]\\
&\times\left[E_*\left\{\Phi^{(1)}\left(\left|\mu\left(\mathbf{s}_0;\hat{\beta}_n\right)\right|+\left|\sum_{i=1}^{n}\lambda_{in}\left(\hat{\theta}_n\right)\epsilon^*(\mathbf{s}_i)\right|\right.\right.\right.\\
&\left.\left.\left.+\,\delta k_0\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|+k\left\|\beta_n^*-\hat{\beta}_n\right\|\right)\right\}^2\right]^{1/2}\\
&+k\left\{\left[E_*\left(\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|\right)^4\right]^{1/4}+\left(E\left\|\beta_n^*-\hat{\beta}_n\right\|^4\right)^{1/4}\right\}\\
&\times\left[P_*(\left\|\theta_n^*-\hat{\theta}_n\right\|)\right]^{1/4}\\
&\times\left[E_*\left\{\Phi^{(1)}\left(\left|\mu\left(\mathbf{s}_0;\hat{\beta}_n\right)\right|+3k_0\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|+k\left\|\beta_n^*-\hat{\beta}_n\right\|\right)\right\}^2\right]^{1/4}\\
\leq\;&\left[1+\left(E_*\left\{\Phi^{(1)}\left(\left|\mu\left(\mathbf{s}_0;\hat{\beta}_n\right)\right|+k\sum_{i=1}^{n}\alpha_i|\epsilon^*(\mathbf{s}_i)|\right.\right.\right.\right.\\
&\left.\left.\left.\left.+\;k\left\|\beta_n^*-\hat{\beta}_n\right\|\right)\right\}^2\right)^{1/2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&\times \left[\left(\sum_{i=1}^{n} \alpha_i E_* \epsilon^*(\mathbf{s}_i)^4 \right)^{1/4} \left(\sum_{i=1}^{n} \alpha_i \right)^{3/4} \right. \\
&\left. \times \left\{ \delta k_0 + k \left(P_* \left(\left\| \theta_n^* - \hat{\theta}_n \right\| > \delta \right) \right)^{1/4} \right\} + k \left(E_* \left\| \beta_n^* - \hat{\beta}_n \right\|^4 \right)^{1/4} \right] \\
\leq \; & \left[1 + k_2^{1/2} \right] \left[6 k_3^{1/2} \alpha_\infty \delta k_0 \right] \qquad \text{(H.29)}
\end{aligned}
$$

for all $n \geq n_0$ for some $n_0 = n_0(k_1, k_2, k_3)$. By a similar argument,

$$
E\hat{Z}_n(\mathbf{s}_0) = g(\beta, \tau_n(\theta)) + r_{1n}, \qquad \text{(H.30)}
$$

where $|r_{1n}| \leq \left[1 + k_2^{1/2} \right] \left[6 k_3^{1/2} \alpha_\infty \delta k_0 \right]$ for all $n \geq n_1$ for some $n_1 \geq 1$. Now, consider I_{1n}. Note that be Condition E.3,

$$
\begin{aligned}
|I_{1n}| &= \left| E \left[\left\{ g \left(\hat{\beta}_n, C \left(\mathbf{0}, \hat{\theta}_n \right) \right) - g \left(\beta, C \left(\mathbf{0}, \theta \right) \right) \right\} \hat{Z}_n(\mathbf{s}_0) / E_* \hat{Z}^*(\mathbf{s}_0) \right] \right| \\
&\leq \left\{ \left(E \left| g \left(\hat{\beta}_n, C \left(\mathbf{0}, \hat{\theta}_n \right) \right) - g \left(\beta, C \left(\mathbf{0}, \theta \right) \right) \right|^2 \right)^{1/2} \right. \\
&\quad \left. \left(E \left| \hat{Z}_n(\mathbf{s}_0) / E_* \hat{Z}^*(\mathbf{s}_0) \right|^2 \right)^{1/2} \right\} \\
&= o(1).
\end{aligned}
$$

Next, consider I_{2n}. By (H.29) and (H.30) and Condition E.3, we have

$$
\begin{aligned}
\left| \frac{E\hat{Z}_n(\mathbf{s}_0)}{EZ(\mathbf{s}_0)} I_{2n} \right| &= \left| E \left[\left\{ E_* \hat{Z}_n^*(\mathbf{s}_0) - E\hat{Z}_n(\mathbf{s}_0) \right\} \hat{Z}_n(\mathbf{s}_0) / E_* \hat{Z}_n^*(\mathbf{s}_0) \right] \right| \\
&\leq \left\{ \left(E \left[g \left(\hat{\beta}_n, \tau_n \left(\hat{\theta}_n \right) \right) - g \left(\beta, \tau_n(\theta) \right) \right]^2 \right)^{1/2} \right. \\
&\quad \left. + \left[12 \left(1 + k_2^{1/2} \right) k_3^{1/3} \alpha_\infty k_0 \right] \delta \right\} \\
&\quad \times \left[E \left(\hat{Z}_n(\mathbf{s}_0) / E_* \hat{Z}_n^*(\mathbf{s}_0) \right)^2 \right]^{1/2} \\
&\to 0,
\end{aligned}
$$

by letting $n \to \infty$ first and then $\delta \downarrow 0$. The proof for $\hat{Z}_{A,n}^{BC}$ is similar and is thus

omitted. This proves Proposition E.2. □

CHAPTER III

INFERENCE FOR COVARIANCE PARAMETERS OF REPLICATED TIME SERIES DATA

A. Introduction

In the analysis of time series, sometimes estimation of the mean parameter of a series is of little interest, while the estimation of the covariance parameters is much more important. Such a problem arises from an application in semiconductor manufacturing. Here, we propose resampling methods for the estimation of covariance parameters of replicated time series.

B. Problem Description

1. Stationary Time Series

Let $\mathbf{X} = \{X_t : X_t \in \mathbb{R}, t \in \mathbb{N}\}$ be a collection of observations from some random variable, with each X_t yielding (potential) observation taken at time t. For example, consider X_t to be the price of a given stock at the end of day t of trading, the air temperature at a given location at a time t, or the water level of a lake at time t. Clearly, in each of these examples, X_t will be correlated to X_{t-1}, and possibly earlier observations as well. When discussing time series, we consider the "lag" between two observations X_{t_1} and X_{t_2}, where the lag $h = |t_1 - t_2|$ is the difference in time between measurements. Time series where strong correlations exist for only small values of h are said to exhibit short range dependence, while those series where strong correlations exist for large values of h exhibit long range dependence.

Stationarity is a property of time series (and of other types of dependent processes such as spatial processes) under which the distributional properties of observations

do not change when the observations are moved in position, e.g. in time. A time series $\mathbf{X}$ is said to be intrisically stationary if

$$E(X_t - X_{t+h}) = 0, \tag{B.1}$$

and

$$\mathrm{Var}(X_t - X_{t+h}) = \mathrm{Var}(X_0 - X_h) = 2\gamma(h) \tag{B.2}$$

for all $(t, h) \in \mathbb{Z}^2$. (B.1) implies that the mean of X_t is the same for all times t, while (B.2) implies that the variance of the difference between two observations depends only on the lag between those two observations, and not on the location of t and $t+h$ within $\mathbb{Z}$. Further, the function $2\gamma(\cdot)$ is a special function known as the variogram, while $\gamma(\cdot)$ is known as the semivariogram.

A time series is said to be second order stationary if it meets a further requirement, that

$$\mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_0, X_h) = C(h), \tag{B.3}$$

for all $(t, h) \in \mathbb{Z}^2$, where $C(\cdot)$ is called the covariogram. (B.3) implies that the covariance between two observations depends only on the lag h. It can be shown that second order stationarity implies intrinsic stationarity, and also implies

$$\begin{aligned} 2\gamma(h) &= \mathrm{Var}(X_t - X_{t+h}) \\ &= \mathrm{Var}(X_t) + \mathrm{Var}(X_{t+h}) - 2\mathrm{Cov}(X_t, X_{t+h}) \\ &= 2(C(0) - C(h)). \end{aligned} \tag{B.4}$$

The variogram and covariogram generally depend on some (possibly vector) parameter θ. These definitions of stationarity can be extended into the spatial domain, where locations are considered rather than times, and the variogram and covariogram are functions of the lag vector $\mathbf{h}$, representing the vector difference between two locations.

2. Replicated Time Series

For $i = 1, \ldots, M$, let $\{X_{it}\}_{t \in \mathbb{N}}$ be independent second order stationary time series with unknown means μ_i but common autocovariance function $C(\cdot; \theta)$. Furthermore, suppose

$$C(h; \theta) = \theta_0 \rho\left(h/\xi; \alpha\right), h \in \mathbb{Z} \tag{B.5}$$

where $\theta_0 \equiv \sigma^2 \in (0, \infty)$, $\theta_1 \equiv \xi \in (0, \infty)$, and $\alpha \in A \subset \mathbb{R}^r$ and $\rho(\cdot; \alpha)$ is a known autocorrelation function on the continuum except for the parameter(s) α such that for all ξ, α,

1. $\rho(0; \alpha) = 1$
2. $\int |\rho(t; \alpha)| dt < \infty$,
3. $\int \rho(t; \alpha) dt \neq 0$.

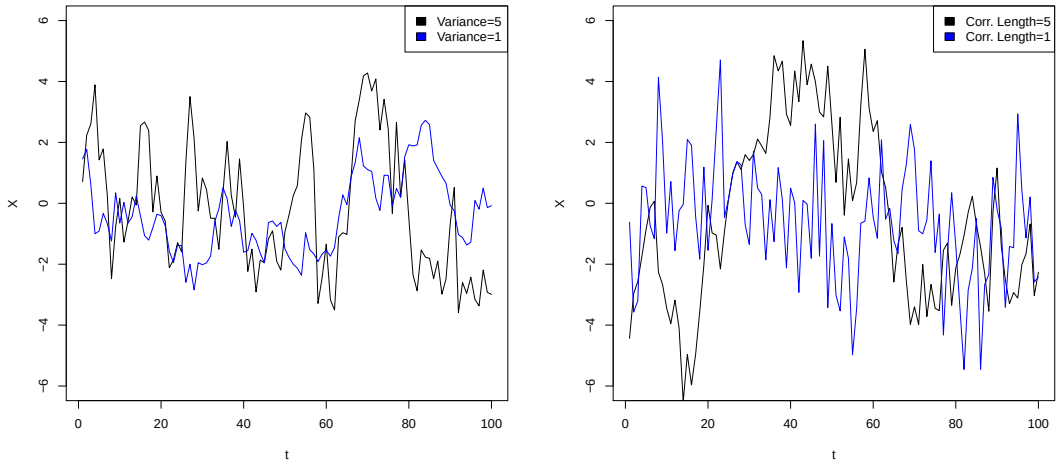
Further, notice that $C(0) = \sigma^2$ follows from the fact the correlation between X_t and itself is 1. Therefore, (B.5) implies that

$$\rho(h/\xi; \alpha) = \frac{C(h)}{C(0)},$$

and (B.4) implies that

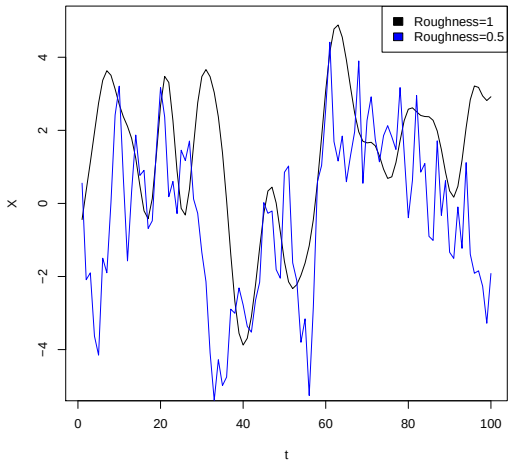
$$2\gamma(h) = 2\sigma^2(1 - \rho(h/\xi; \alpha)).$$

Our primary interest is the estimation of the vector covariance parameters $\theta = (\theta_0, \theta_1, \alpha')' \in \Theta$, where $\Theta = (0, \infty) \times (0, \infty) \times A$. The μ_i's are treated as nuisance parameters. The covariance parameters are defined as follows. σ^2 is the variance of the time series, i. e. $\sigma^2 = E[X_{it} - \mu_i]^2$. Larger values of σ^2 correspond to bigger variation from the mean μ_i as seen in Figure 8(a). The parameter ξ identifies the strength of correlation between two observations X_t and X_{t+h} at a given length. Small values of ξ correspond to a short range dependence structure, while higher values correspond to



(a)

(b)



(c)

Figure 8. Time series generated with differing values of (a) variance σ^2, (b) correlation length ξ, and (c) roughness exponent α.

long range dependence. As seen in Figure 8(b), larger values of ξ correspond to time series with stronger form of dependence at unit lag. Finally, the roughness exponent α also determines the roughness or smoothness of the line, but on a more local scale than ξ. Larger values of α correspond to smoother time series, as scene in Figure 8(c).

Suppose that we have observed $\mathbb{X}_n \equiv \{X_{it} : t = 1, \ldots, L, i = 1, \ldots, M\}$. Generally, the length of the series does not need to be the same for all i, but we do not pursue such generalization here. While we only have L observations for the estimation of each μ_i, we have ML observations for the estimation of the covariance parameters in θ. We propose methods for the estimation of θ at $O_p([ML]^{-1/2})$. Further, we propose block bootstrap methods for estimating the standard errors of these estimators.

C. Motivating Example

This problem is motivated by the estimation of line width roughness (LWR) parameters in the manufacturing process of semiconductors and transistors (Patel et al. 2010). Here, the process means are of little interest, while the covariance parameters are indicative of the overall quality of the final product.

The lines in question can be seen on on a scanning electron micrograph (SEM) image, containing 8-20 lines of 300-1500 nm in length. These lines represent patterned film, and the width of lines determine the resistance. In the manufacturing of transistors, smaller line width allows for more transistors to be included in a chip. This means chips will have smaller area and thus, lower cost. An example of an SEM image is given in Figure 9.

The lines lie adjacent to one another, with rough edges. If we limit ourselves to lines of the same length, then we can say we have M lines of length Ld, where d is

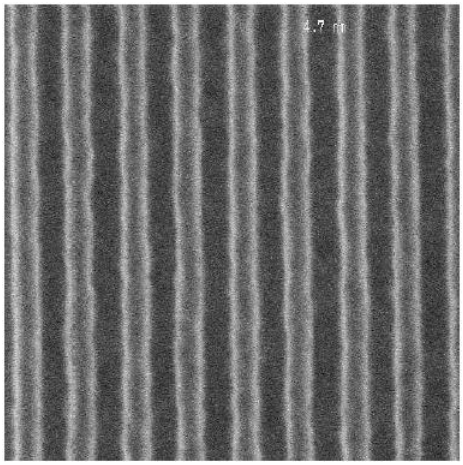


Figure 9. Example of an SEM image.

selected such that L is an integer. The variable of interest, X_{it}, represent the width of line i at location t, $t \in \{0, 1, \ldots, L\}$, as shown in Figure 10. Fluctuations in width are caused by LWR, which is caused by polymer aggregates being non-uniformly distributed along the edge of lines due to differing dissolution rates (Yoshimura et al. (1993) and Yamaguchi et al. (2003)). When line widths are small, the LWR can be very significant, affecting the transistor's performance. Therefore, investigation of LWR parameters is of vital interest.

Despite the spatial nature of the lines in an SEM image, we treat the observations as time series data because as we move along a line from $t = 0$ to $t = L$, we are only moving in one direction, leading to one-dimensional correlation similar to what is seen in time series data, rather than the higher-dimensional correlation seen in a spatial setting.

It is accepted that LWR can be described by three covariance parameters, variance of line width σ^2, correlation length ξ and roughness exponent α (Constantoudis et al. 2003). Leunissen et al. (2004) suggest a covariance model depending only on

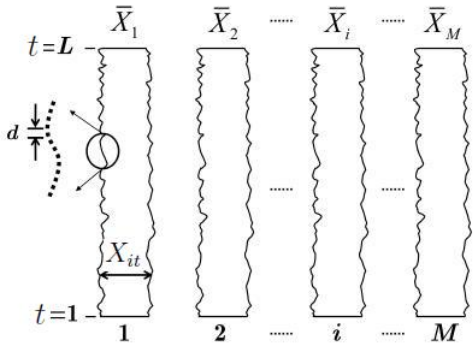


Figure 10. Framework of time time series representation of SEM image.

σ^2 and ξ, but Figure 8(c) demonstrates the importance of α's inclusion. The variance parameter describes fluctuations transversing the line, but does not incorporate any information regarding correlation between different locations along the same line. This type of correlation is described by the correlation length. The roughness exponent describes the short-range roughness of the line.

D. Inference Methodology Review

Here, we include a brief literature review, and describe the methods to be used in analyzing the repeated time series data. This involves a combination of block bootstrap and subsampling resampling algorithms, as well as parameter estimation using weighted least squares (WLS).

1. Weighted Least Squares

Weighted least squares (WLS) improve upon ordinary least squares (OLS) methods in that they are more suited for situations in which the observations do not have the same variance. For example, in simple linear regression, one key assumption is that

the residuals have constant variance. In this case, OLS is adequate for the estimation of the regression line's parameters. However, if the assumptions are relaxed such that the variance of the residuals can vary as the explanatory variable changes, WLS could be used, with the differing weights reflecting the change in various.

We will use WLS methods to estimate variogram parameters, as discussed by Cressie (1985). Lahiri et al. (2002) explored the asymptotic properties of variogram parameter estimates from WLS, as well as OLS and generalized least squares (GLS), and gave the regularity conditions under which the variogram estimates and parameter estimates from least squares methods were asymptotically normal.

The estimation of variogram parameters uses the method of moments variogram estimator $2\hat{\gamma}(h)$. If OLS were used, our estimator would be

$$\hat{\theta}_{OLS} = \text{argmin}_\theta \sum_h [2\hat{\gamma}(h) - 2\gamma(h;\theta)]^2.$$

Using WLS will allow for the fact that the variance of $2\hat{\gamma}(h)$ depends on the value of h. Using the reciprocal of the variance as the weight, we have

$$\hat{\theta}_{WLS} = \text{argmin}_\theta \sum_h \frac{1}{\text{Var}(2\hat{\gamma}(h))} [2\hat{\gamma}(h) - 2\gamma(h;\theta)]^2. \tag{D.6}$$

Also, note that the coefficient 2 in (D.6) has no effect on the minimization of the expression in (D.6), and may be removed.

2. Block Bootstrap

The block boostrap, developed independently by Künsch (1989) and Liu and Singh (1992), is a nonparametric resampling method for dependent observations, such as time series or spatial data. Suppose we have a series of n dependent observations $\mathbf{X} = X_1, X_2, \ldots, X_n$, and we wish to study the behavior of a statistic $R(\mathbf{X})$, such as the sample mean, sample variance, or a predictor based on the observed data.

We wish to create bootstrap replicates of length $l < n$. For simplicity, let $m = n/l$ be an integer. The i-th block would consist of observations $X_i, X_{i+1}, \ldots, X_{i+l-1}$ for $i = 1, \ldots, n - l + 1$.

The blocks are then sampled independently with replacement m times, and the m resulting blocks are concatenated to produce the bootstrap replicate

$$X_1^*, X_2^*, \ldots, X_n^*.$$

This is done a large number of times (at least in the hundreds), and we then calculate $R(\mathbf{X}^*)$ for each bootstrap replicate.

Clearly, the choice of block length l is of critical importance. Generally, the larger the block length, the smaller the bias of a bootstrap estimator, but the larger the variance. Therefore, the ideal block length should be one that minimizes the mean square error (MSE) of the bootstrap estimator. For example, for α, the optimal block length would be given by

$$l_{opt.} = \text{argmin}_l E\left[\alpha^* - \alpha\right]^2.$$

Early work indicated only that l should increase with n. Hall et al. (1995) and Bühlmann and Künsch (1999) explored methods for the selection of l. However, their methods were largely dependent on the data. Politis and White (2004) introduced a method that was less dependent on the data, based on a "flat-top" lag window introduced in Politis and Romano (1995). This method can be used to estimate the optimal block length for estimating the variance of a linear function of the data of the form (cf. Patel et al. (2010))

$$T_n = [M(L-h)]^{-1} \sum_{i=1}^{M} \sum_{t=1}^{L-h} g(X_{it}).$$

To estimate the variance of a LWR parameter estimator, we must identify a suitable

function $g(\cdot)$ by linearizing the estimator, as decribed in Serfling (1980). Once $g(\cdot)$ has been selected, the Politis and White (2004) method for selecting the optimal block length is given by the following steps:

1. Calculate the sample mean of $g(X_{it})$, given by
$$\bar{g} = [M(L-h)]^{-1} \sum_{i=1}^{M} \sum_{t=1}^{L-h} g(X_{it}).$$

2. Next, calculate the sample autocovariance of $\{g(X_{it})\}$ at lag k, given by
$$\hat{R}(k) = [M(L-h)]^{-1} \sum_{i=1}^{M} \sum_{t=1}^{L-h-|k|} [(g(X_{it}) - \bar{g})(g(X_{(i+|k|)t}) - \bar{g})].$$

3. Choose a value k_0 such that $\hat{R}(k) \approx 0$ for $k > k_0$.

4. Calculate the flat-top kernel of Politis and Romano (1995), given by
$$\lambda(t) = \begin{cases} 1 & \text{if } |t| \in [0, 1/2), \\ 2(1-|t|) & \text{if } |t| \in [1/2, 1], \\ 0 & \text{otherwise.} \end{cases}$$

5. Calculate
$$\hat{G} = \sum_{k=-2k_0}^{2k_0} \lambda(k/(2k_0))|k|\hat{R}(k).$$

6. Calculate
$$\hat{D} = \sum_{k=-2k_0}^{2k_0} \lambda(k/(2k_0))\hat{R}(k).$$

7. Then, the estimated optimal block size is given by
$$\hat{l}_{opt.} = \left(\frac{3[M(L-h)]}{2}\right) \left(\frac{\hat{G}}{\hat{D}}\right)^{2/3}.$$

Each covariance parameter will have its own unique influence function $g(\cdot)$.

3. Subsampling

Subsampling is a resampling method in which the sampling distribution of a statistic is studied using subsets of the data. As described in Politis et al. (1999), if we wish to investigate the properties of a statistic $R_n(\mathbf{X})$, where $\mathbf{X} = \{X_1, \ldots, X_n\}$, we calculate not only $R_n(\mathbf{X})$, but also the corresponding statistic $R_l(\mathbf{X}^{(i)})$, where $\mathbf{X}^{(i)}$ is the i-th subsample, or subset, of size l of $\mathbf{X}$, for some $l < n$.

If the data is independent, we look at all subsets of size l, so i ranges from 1 to $n!/(l!(n-l)!)$. in the case of dependent data, we care only about subsets where the l observations are highly correlated.

For time series, we choose l to be a lag such that if $i < j$, we consider X_i and X_j to be highly correlated if $j - i \leq l$, and not highly correlated if $j - i > l$. Then, the subsets of $\mathbf{X}$ we consider are $\mathbf{X}^{(i)} = \{X_i, \ldots, X_{i+l}\}$ for $i = 1, \ldots, n-l+1$. In the spatial setting, subsamples are often designed to be of the same shape as the original sampling area.

Notice the difference between the block bootstrap and subsampling. In the block bootstrap, we resample n observations from the empirical distribution a large number of times to get the bootstrap replicate $\mathbf{X}^*$, and then calculate $R(\mathbf{X}^*)$ for each replicate. Here, we are not generating observations from an empirical distribution, and we are dealing with subsamples of size $l < n$ rather than replicates of size n.

As in the case of the block boostrap, the choice of block length l is of vital importance, as it helps determine the behavior of the subsampled statistic $R_n(\mathbf{X}^{(i)})$.

E. Our Methodology

1. Estimation of Parameters

We naïvely estimate σ^2 by

$$s_n^2 = \frac{1}{ML}\sum_{i=1}^{M}\sum_{t=1}^{L}(X_{it} - \bar{X}_i)^2, \tag{E.7}$$

where $\bar{X}_i$ is the sample mean of the i-th series. Furthermore, we estimate the variogram $2\gamma(h) = 2[C(0;\theta) - C(h;\theta)]$ by

$$2\hat{\gamma}(h) = \frac{1}{M(L-h)}\sum_{i=1}^{M}\sum_{t=1}^{L-h}(X_{it} - X_{i(t+h)})^2. \tag{E.8}$$

Notice that s_n^2 is a poor estimator of σ^2, as

$$\begin{aligned} Es_n^2 &= \frac{1}{ML}\sum_{i=1}^{M}\sum_{t=1}^{L} E\left[(X_{it} - \mu_i) - (\bar{X}_i - \mu_i)\right]^2 \\ &= C(0;\theta) - L^{-2}\sum_{t_1=1}^{L}\sum_{t_2=1}^{L} C(|t_1 - t_2|;\theta) \\ &= C(0;\theta)\left[1 - L^{-2}\sum_{t_1=1}^{L}\sum_{t_2=1}^{L}\rho(|t_1 - t_2|;\xi,\alpha)\right] \\ &= \sigma^2\left[1 - f(L;\xi,\alpha)\right], \end{aligned} \tag{E.9}$$

where

$$f(L;\xi,\alpha) = L^{-1} + 2L^{-2}\sum_{k=1}^{L-1}(L-k)\rho(k/\xi;\alpha).$$

It can be shown that

$$f(L;\xi,\alpha) \sim L^{-1}\left[1 + 2\sum_{k=1}^{\infty}\rho(k/\xi;\alpha)\right] \text{ as } L \to \infty.$$

Therefore, the bias of s_n^2 is $O(L^{-1})$, but it can be significantly large in finite samples or if ξ is large. Figure 11 demonstrates that as ξ increases, the value of $f(L;\xi,\alpha)$

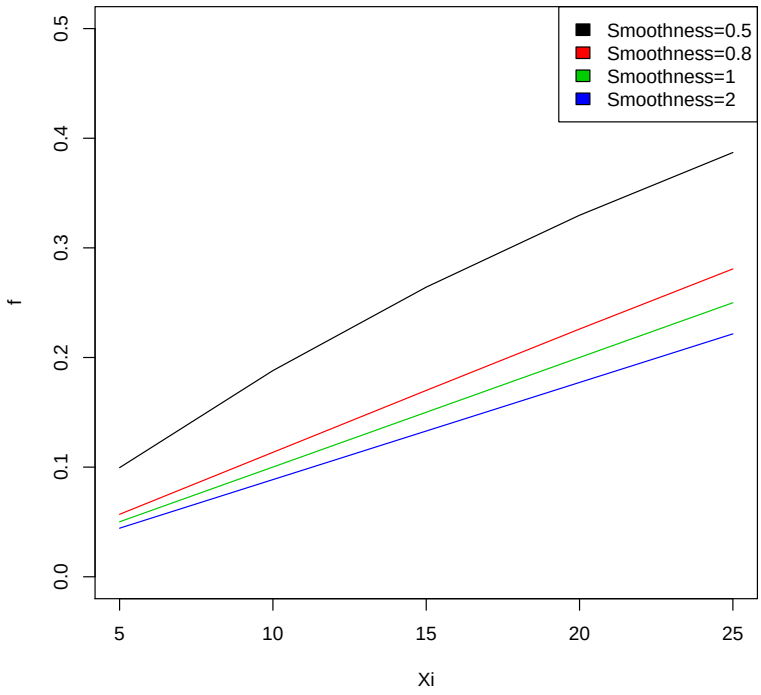


Figure 11. Approximate values of $f(L;\xi,\alpha)$ for $L = 200$ for varying values of α.

increases as well, leading to a value of (E.9) far from σ^2.

One way around the bias is to estimate the variance of the sample means $\bar{X}_i$, by using

$$\hat{\sigma}^2_{\bar{X}_i} = \frac{1}{M-1}\sum_{i=1}^{M}(\bar{X}_i - \bar{X})^2, \tag{E.10}$$

where $\bar{X}$ is average of the sample means $\bar{X}_i$. Leunissen et al. (2004) proposed

$$\hat{\sigma}^2 = s_n^2 + \hat{\sigma}^2_{\bar{X}_i}$$

as an unbiased predictor of σ^2. Furthermore, this sum is constant as L changes. However, the term in (E.10) is problematic because it also incorporates non-LWR

sources of variability. Therefore, a different bias correction is needed.

From (E.9), we see that $f(L;\xi,\alpha) = \sum_{t_1=1}^{L}\sum_{t_2=1}^{L}\rho(|t_1-t_2|/\xi;\alpha)$. A straightforward plug-in estimator for asymptotic unbiasedness will thus be given by

$$\hat{\sigma}^2_{BC} = \frac{s_n^2}{1 - f\left(L;\hat{\xi},\hat{\alpha}\right)}. \tag{E.11}$$

The strength of dependence and shape parameters will be estimated using weighted least squares methods, where the weights are based on block subsampling.

Here, the statistic $R(\cdot)$ of interest is the method of moments variogram estimate, because it is used to calculate the weights for WLS.

Let l_1, an integer in $(1, L)$, be the subsample size, selected via a method similar to that proposed in Politis and White (2004). Then

$$2\hat{\gamma}^{(i,j)}(h) = l_1^{-1}\sum_{t=j}^{j+l_1-1}(X_{it} - X_{i(t+h)})^2$$

for $j = 1,\ldots,(L-h-l_1+1)$, $i = i,\ldots,M$ and let $\bar{\gamma}$ be the average of $\hat{\gamma}^{(i,j)}(h)$ over all i, j. Then the weights for SWLS are

$$\hat{w}(h) = \frac{l_1}{M(L-h)}\frac{1}{M(L-h-l_1+1)}\sum_{i=1}^{M}\sum_{j=1}^{L-h-l_1+1}(\hat{\gamma}^{(i,j)}(h) - \bar{\gamma}(h))^2.$$

Using the weights as previously defined, the SWLS estimator of (σ^2,θ) is

$$(\hat{\sigma}^2,\hat{\theta}) = \text{argmin}_{\sigma,\theta}\sum_{h=1}^{h_{max}}\frac{[2\hat{\gamma}(h) - 2\gamma(h;\sigma,\theta)]^2}{\hat{w}(h)}. \tag{E.12}$$

This estimator has the advantages of avoiding the estimation of nuisance parameters and of having a mean squared error of $O([ML]^{-1})$. Furthermore, $\hat{\theta} = \left(\hat{\xi},\hat{\alpha}\right)'$ may be plugged into (E.11) for an unbiased estimator of σ^2.

2. Bootstrap Estimation of Parameter Estimate Variance

We propose using moving block bootstrap (MBB) methodologies, initially developed by Künsch (1989) and Liu and Singh (1992), to estimate the variance of the covariance parameter estimates.

Fix a block length $l_2 < L$. For simplicity, we suppose that L is a multiple of l_2, i.e. $b_2 = L/l_2$ is an integer, or work with the smallest integer greater than L/l_2 and retain only the first L resampled values for the reconstruction of each line. Since the means of the M lines can be different, we need to detrend the data in the formulation of the block bootstrap.

The steps in the block bootstrap algorithm are as follows:

1. Let $e_{it} = X_{it} - \bar{X}_i$, $t = 1, \ldots, L$, $i = 1, \ldots, M$ denote the residuals. Form the centered residuals

$$\tilde{e}_{it} = e_{it} - \bar{e}$$

where $\bar{e}$ is the grand mean of all e_{it}'s across all groups, given by

$$\bar{e} = \frac{1}{ML} \sum_{i=1}^{M} \sum_{t=1}^{L} e_{it}.$$

2. Form the overlapping blocks of centered residuals of length l_2 within each line, giving

$$\mathbb{B}(i, j) \equiv (\tilde{e}_{it} : t = j, \ldots, j + l_2 - 1)$$

for $j = 1, \ldots, L - l_2 + 1, i = 1, \ldots, M$.

3. Resample Mb_2 many blocks independently and reconstruct the MBB version of the original time series (of length ML), $\mathbb{X}^* = \{X^*_{it} : t = 1, \ldots, L, i = 1, \ldots, M\}$, where X_{it} is given by

$$X^*_{it} = \bar{X}_i + e^*_{it}$$

for all i, t, where $(e^*_{it}, t = 1, \ldots, L, i = 1, \ldots, M)$ is obtained by concatenating the values in Mb_2 resampled blocks.

4. Compute α^*, ξ^*, σ^{*2} based on this MBB data set. For example, to calculate σ^{*2} we plug in $\mathbb{X}$ into (E.7), giving

$$\sigma^{*2} = \frac{1}{ML}\sum_{i=1}^{M}\sum_{t=1}^{L}(X^*_{it} - \bar{X}^*_i)^2,$$

where $\bar{X}^*_i$ is the sample mean of the i-th time series for the bootstrap replicate, given by

$$\bar{X}^*_i = \sum_{t=1}^{L} X^*_{it}.$$

α^* and ξ^* are calculated making similar adjustments to (E.8) and (E.12), giving

$$2\gamma^*(h) = \frac{1}{M(L-h)}\sum_{i=1}^{M}\sum_{t=1}^{L-h}(X^*_{it} - X^*_{i(t+h)})^2$$

5. Repeat steps 3 and 4 B times where B is large.

The values of α^*, ξ^*, and σ^{*2}, the estimators of each parameter based on a bootstrap replicate, may be used to variance of the parameter estimates. For example, the variance of $\hat{\alpha}$ may be calculated as

$$\hat{\text{Var}}(\hat{\alpha}) = \frac{1}{B-1}\sum_{b=1}^{B}[\alpha^{*b} - \bar{\alpha}^*]^2, \tag{E.13}$$

where $\bar{\alpha}^*$ is the mean of $\alpha^{*1}, \ldots, \alpha^{*B}$. Similar adjustments will lead to variance estimates of $\hat{\sigma}^2$ and $\hat{\xi}$.

3. Bootstrap Confidence Intervals

Moreover, the distribution of bootstrap replicates may be used to produce bootstrap confidence intervals, giving a range of plausible values for a parameter. This is a

key use of bootstrap replicates, particularly when the distributive properties of $\mathbb{X}$ are unknown. As an example, we continue to consider the estimation of the roughness exponent α, and we look at the bootstrap estimators $\alpha^{*1}, \ldots, \alpha^{*B}$. Let $\alpha^*_{(n)}$ be the n-th order statistic of those estimators. Then a $100(1-\pi)\%$ nonparametric bootstrap confidence interval for α is given by

$$\left(\alpha^*_{(B(\pi/2))}, \alpha^*_{(B(1-\pi/2))}\right),$$

and similar procedures can be used for confidence intervals for σ^2 and ξ.

F. Theoretical Results

Theorem F.1. *Suppose that the following conditions hold:*

(F.1.1) $\{X_{it}\}$ *is a second order stationary sequence and for* $i \geq 1$, *the* $\{X_{it}\}$*'s are iid.*

(F.1.2) $E|X_{it}|^{4+2\delta} < k < \infty$ *for some* $\delta \in (0, \infty)$ *and* $k \in (0, \infty)$.

(F.1.3) $\int \alpha(t)^{\frac{\delta}{2+\delta}} dt < \infty$.

Then, as $L \to \infty$, $M \to \infty$,

$$s_n^2 - \sigma^2 = \frac{1}{\sqrt{ML}} W_{1n} + \frac{1}{L} W_{2n}$$

for some bivariate random vector (W_{1n}, W_{2n}) *such that*

$$\begin{pmatrix} W_{1n} \\ W_{2n} \end{pmatrix} \to^d \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}$$

and W_1, W_2 *are independent,* $W_1 \sim N(0, \tau_1^2)$ *with* $\tau_1^2 = \sum_{k \in \mathbb{Z}} Cov(X_{11}^2, X_{1k}^2)$ *and* $P(W_2 = -\tau^2) = 1$.

Theorem F.1 shows that the best possible rate of convergence for s_n^2 to σ^2 is $O_p(L^{1/2}[M \wedge L]^{-1/2})$, where $x \wedge y = \min\{x, y\}$, $x, y \in \mathbb{R}$. For $L \sim M$, the limit distribution is normal with a non-zero bias, which makes s_n^2 a worse estimator (in the sense of asymptotic MSE) compared to the estimator obtained by WLS variogram fitting. For $L \gg M$, this bias vanishes asympotically and both estimators have the same rate of convergence. For $M \gg L$, s_n^2 has a slower rate of convergence compared to $\hat{\sigma}^2_{WLS}$.

The next theorem investigates the asymptotic distribution of $\hat{\theta}$.

Theorem F.2. *Let* $\hat{\theta} = argmin_\theta \sum_{h=1}^{h_0} \hat{w}(h)[2\hat{\gamma}(h) - 2\gamma(h;\theta)]^2$. *Suppose that the conditions of Theorem F.1 hold. Further, suppose that*

(F.2.1) $\hat{w}(h) \to_p w(h)$ *for all* $h \in [1, h_0]$.

(F.2.2) $\gamma(\cdot, \theta)$ *is twice continuously differentiable in* θ *and*

$$\sup\left\{\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\gamma(h;\theta)\right| : 1 \le i, j \le p, 1 \le h \le h_0, \|\theta - \theta_0\| < \delta\right\} < \infty$$

for some $\delta > 0$.

(F.3.3) *The matrix*

$$\Gamma_0 \equiv \sum_{h=1}^{h_0} w(h)[2\gamma^{(1)}(h;\theta_0)][2\gamma^{(1)}(h;\theta_0)]'$$

is nonsingular.

Then,

(a) *If the parameter space is compact, then*

$$\hat{\theta}_{M,L} \to_p \theta$$

as $M, L \to \infty$.

(b) *For any p-consistent* $\left\{\hat{\theta}_{M,L}\right\}$,

$$\sqrt{ML}\left(\hat{\theta}_n - \theta\right) \to^d N(\mathbf{0}, \Sigma)$$

as $M, L \to \infty$, *where*

$$\Sigma \equiv \Gamma_0^{-1}\left(\sum_{k\in\mathbb{Z}} ET_1T_k'\right)\Gamma_0^{-1}$$

and

$$T_k = \sum_{h=1}^{h_0} w(h)\left\{(X_{1k} - X_{1(k+h)})^2 - 2\gamma(h;\theta_0)\right\} 2\gamma^{(1)}(h;\theta_0),$$

for $k \in \mathbb{Z}$.

G. Simulations

1. Framework

For simulations, we look to the setup in Patel et al. (2010), which looks at the motivating example regarding the estimation of LWR parameters in semiconductor manufacturing.

Suppose that $\rho(h/\xi;\alpha)$ is of the form

$$\rho(h/\xi;\alpha) = \exp\left(-\left[\frac{|h|}{\xi}\right]^{2\alpha}\right),$$

for $h \in \mathbb{Z}$. The lag between two observations h is divided by the scale parameter ξ, which controls the strength of the correlation, and α controls the smoothness of the correlation function. This makes the autocovariance function

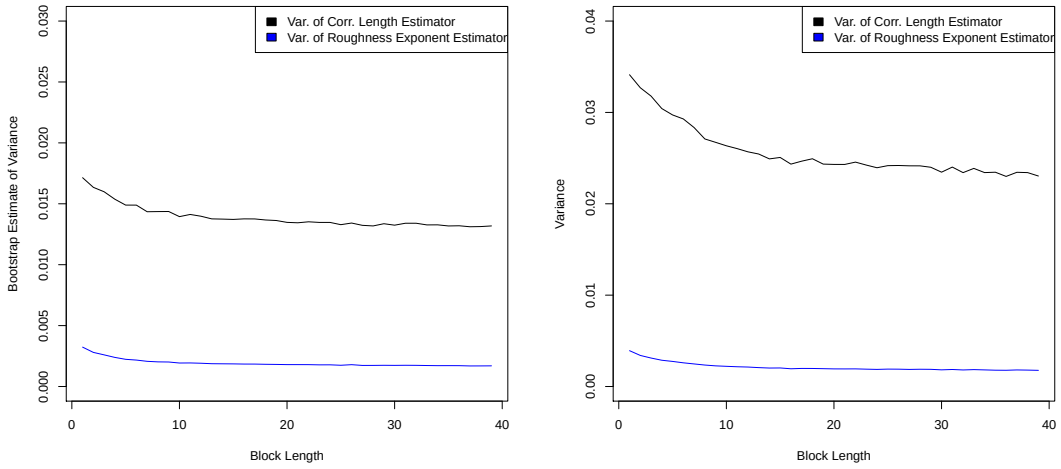
$$\begin{aligned} C(h;\theta) &= \sigma^2\rho(h/\xi;\alpha) \\ &= \sigma^2 \exp\left(-\left[\frac{|h|}{\xi}\right]^{2\alpha}\right), \end{aligned}$$

where $\theta = (\theta_0, \xi, \alpha)$. This particular form of $\rho(\cdot)$ was proposed by Sinha et al. (1988).

The variogram of $h \in \mathbb{Z}$ is then given by

$$2\gamma(h;\theta) = 2\sigma^2 \left[1 - \exp\left(-\left[\frac{|h|}{\xi}\right]^{2\alpha}\right)\right].$$

2. Results



(a) (b)

Figure 12. Bootstrap variance estimates for $\alpha = 0.05$ and (a) $\xi = 10$ and (b) $\xi = 15$.

For simulations, we set $M = 8$ and $L = 500$, given 8 independent time series each with 500 observations. Data was generated using $\sigma^2 = 1$, $\alpha = 0.5$ and $\xi \in \{10, 15\}$. One variable of interest was the bootstrap estimate of the variance of each predictor, given in (E.13).

Figure 12 shows the estimation of the variance of $\hat{\xi}$ and $\hat{\alpha}$ when $\xi = 10$ and 15 for various block lengths. Notice that the variance estimate stabilizes earlier for $\hat{\alpha}$, although a block length around 25 seems adequate for the variance of $\hat{\xi}$ to stabilize as well. We have observed similar results for other parameter values.

However, the larger ξ is, the longer the block length needs to be for the variance estimate to stabilize, which can be seen by comparing Figure 12(a) to Figure 12(b).

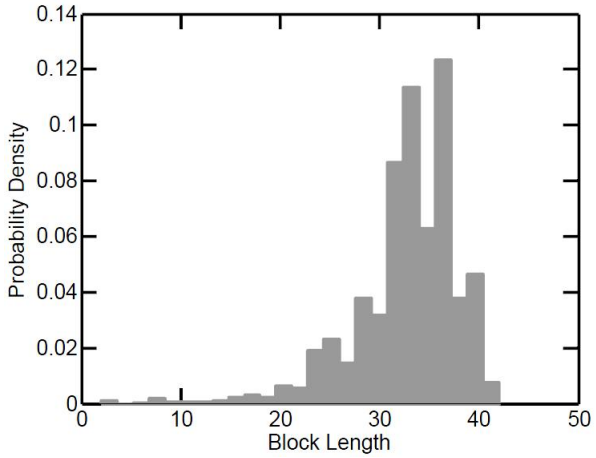


Figure 13. Optimal block length using the Politis and White (2004) method for $\xi = 15$.

Figure 13 shows the optimal block length calculated via the method proposed by Politis and White (2004) for $\xi = 15$. By comparing these results to Figure 12, it is easily seen that the Politis and White method tends to produce block lengths which allow for the stable bootstrap estimation of variance. Therefore, we recommend that this method be used for future applications.

Figure 14 demonstrates the bias correction capabilities of the predictor in (E.11), by comparing the bias of that predictor to the bias of the predictor in (E.7). Interestingly, the bias of the bootstrap predictor appears very stable for all block lengths.

H. Proofs

Proof of Theorem F.1. We begin by observing that

$$s_n^2 = \frac{1}{ML}\sum_{i=1}^{M}\sum_{t=1}^{L} X_{it}^2 - \frac{1}{M}\sum_{i=1}^{M} \bar{X}_{i\cdot}^2,$$

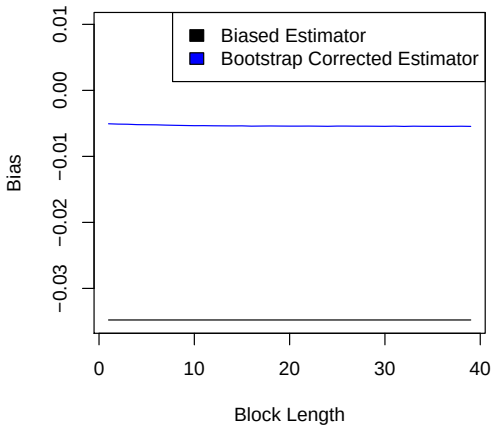


Figure 14. Bias of variance estimators in (E.7) and (E.11) for $\xi = 10$ and $\alpha = 0.5$.

which implies that

$$\begin{aligned} s_n^2 - \sigma^2 &= \frac{1}{ML}\sum_{i=1}^{M}\sum_{t=1}^{L}(X_{it}^2 - \sigma^2) - \frac{1}{L}\cdot\frac{1}{M}\sum_{i=1}^{M}\left(\sqrt{L}\bar{X}_{i\cdot}\right)^2 \\ &\equiv \frac{1}{\sqrt{ML}}W_{1n} + \frac{1}{L}W_{2n}, \text{ say.} \end{aligned}$$

Suppose we can show that

$$W_{1n} \to^d N(0, \tau_1^2),$$

with $\tau_1^2 = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(X_{11}^2, X_{1k}^2)$, and that

$$W_{2n} \to_p -\left[\sum_{k\in\mathbb{Z}} \mathrm{Cov}(X_{11}, X_{1k})\right].$$

Since W_{2n} has a degenerate limit, it follows that

$$\begin{pmatrix} W_{1n} \\ W_{2n} \end{pmatrix} \to^d \begin{pmatrix} W_1 \\ W_2 \end{pmatrix},$$

where W_1 and W_2 are independent, $W_1 \sim N(0, \tau_1^2)$ and $P(W_2 = -\tau) = 1$.

The convergence of W_{2n} to $-\tau^2$ in probability follows from Chebychev's inequality, Conditions (F.1.2) and (F.1.3), and the fact that

$$\begin{aligned} E\left(\sqrt{L}\bar{X}_{i\cdot}\right)^2 &= \frac{1}{L}\sum_{t=1}^{L}\sum_{s=1}^{L} C(t-s;\theta) \\ &= \frac{1}{L}\sum_{k=-(L-1)}^{L-1}(L-k)C(k;\theta) \\ &\to \sum_{k\in\mathbb{Z}} C(k;\theta) \\ &\equiv \tau^2. \end{aligned}$$

Condition (F.1.1) can be proven using arguments similar to those used for proving the Linedeberg-Feller Theorem. Here we provide the details for completeness. Let $\tau_1^2(L) = EZ_{1L}^2$, where

$$Z_{iL} = \frac{1}{\sqrt{L}}\sum_{t=1}^{L}(Y_{it}^2 - \sigma^2),\ 1 \le i \le M.$$

Note that by the independence of the $\{X_{it}\}$-series for $i \le 1$, $\{Z_{1n}, \ldots, Z_{Mn}\}$ are also independent. Fix $t \in \mathbb{R}$, $t \neq 0$. There, for any $\epsilon > 0$,

$$\begin{aligned} &\left| E\exp\left(\iota t\sum_{i=1}^{M} Z_{IL}/\sqrt{M}\right) - \exp\left(-t^2\tau_1(L)^2/2\right)\right| \\ &\le \sum_{i=1}^{M}\left|E\exp\left(\iota t Z_{iL}/\sqrt{M}\right) - \exp\left(-t^2\tau_1(L)^2/(2M)\right)\right| \\ &\le \sum_{i=1}^{M}\left|E\exp\left(\iota t Z_{iL}/\sqrt{M}\right) - \left[1 - \frac{t^2\tau_1(L)^2}{2M}\right]\right| \\ &\quad + M\left|\exp(-t^2\tau_1(L/(2M))^2) - \left[1 - \frac{t^2\tau_1(L)^2}{2M}\right]\right| \\ &\le \sum_{i=1}^{M} E\left\{\frac{|Z_{iL}|^3}{M^{3/2}} \wedge \frac{2Z_{iL}^2}{M}\right\} \end{aligned}$$

$$
\begin{aligned}
& + \frac{M}{2}\left[\frac{t^2\tau_1(L)^2}{2M}\right]^2 \exp\left(t^2\tau_1(L)^2/(2M)\right) \\
\leq\ & \epsilon \sum_{i=1}^{M} \frac{EZ_{iL}^2}{\left(\sqrt{M}\right)^L} + \frac{2}{M}\sum_{i=1}^{M} EZ_{iL}^2 \mathbb{1}\left(|Z_{iL}| > \epsilon\sqrt{M}\right) \\
& + \frac{t^4}{8M}\tau_1(L)^4 \exp\left(\frac{t^2\tau_1(L)^2}{2M}\right) \\
\equiv\ & I_1 + I_2 + I_3, \text{ say.}
\end{aligned}
$$

Under Conditions (F.1.2) and (F.1.3), $\tau_1(L)^2 \to \tau_1^2$ as $L \to \infty$. Hence, $I_3 = o(1)$ as $M, L \to \infty$. Also, by (F.1.2) and (F.1.3) and Denker (1986), $\{Z_{1L}^2\}_{L\geq 1}$ is uniformly integrable. Hence, for any $\epsilon > 0$,

$$
\begin{aligned}
I_2 &\leq \sup_{L\geq 1} EZ_{1L}^2 \mathbb{1}\left(|Z_{1L}| > \epsilon\sqrt{M}\right) \\
&\to 0 \text{ as } M \to \infty.
\end{aligned}
$$

Further, $\limsup_{M,L} I_1 \leq \epsilon\tau_1^2$ for every $\epsilon > 0$. Since $\epsilon > 0$ is arbitrary, and $\tau_1^2(L) \to \tau_1^2$, this implies $W_{1n} \to^d N(0, \tau_1^2)$. □

Proof of Theorem F.2. We first give a proof of (b). Note that by the smoothness of $\gamma(\cdot;\theta)$ in θ, the WLS estimator

$$
\hat{\theta} \equiv \text{argmin}_\theta \sum_{h=1}^{h_0} \hat{w}(h)[2\hat{\gamma}_n(h) - 2\gamma(h;\theta)]^2
$$

is also a solution to the estimating equation

$$
\sum_{h=1}^{h_0} \hat{w}(h)\left[2\hat{\gamma}_n(h) - 2\gamma\left(h;\hat{\theta}\right)\right] 2\gamma^{(1)}\left(h;\hat{\theta}\right) = 0,
$$

where $\gamma^{(1)}(\cdot;\theta)$ is the vector of partial derivatives of $\gamma(\cdot;\theta)$ with respect to θ. Using standard arguments from M-estimation theory, as in (Lahiri et al. (2002)), it follows that $\left(\hat{\theta} - \theta_0\right)$ has the same asymptotic distribution as its linear approximation, given

by

$$\left(\hat{\theta} - \theta_0\right) = \hat{\Gamma}^{-1} \sum_{h=1}^{h_0} \hat{w}(h)[2\hat{\gamma}_n(h) - 2\gamma(h;\theta_0)]2\gamma^{(1)}(h;\theta_0),$$

where

$$\begin{aligned} -\hat{\Gamma} &= \sum_{h=1}^{h_0} \hat{w}(h)[2\hat{\gamma}_n(h) - 2\gamma(h;\theta_0)]2\gamma^{(2)}(h;\theta_0) \\ &\quad - \sum_{h=1}^{h_0} \hat{w}(h)2\gamma^{(1)}(h;\theta_0)[2\gamma^{(1)}(h;\theta_0)]' \end{aligned} \tag{H.14}$$

and $2\gamma^{(2)}(\cdot;\theta)$ denotes the Hessian matrix of second order partial derivatives of $2\gamma(\cdot;\theta)$ with respect to θ. Next, note that for each h,

$$\begin{aligned} & 2\hat{\gamma}_n(h) - 2\gamma(h;\theta_0) \\ &= \frac{1}{M(L-h)} \sum_{i=1}^{M} \sum_{t=1}^{L-h} (X_{it} - X_{i(t+h)})^2 - 2\gamma(h;\theta_0) \\ &= \frac{1}{M(L-h)} \sum_{i=1}^{M} \sum_{t=1}^{L-h_0} U_{it}(h) + \frac{1}{M(L-h)} \sum_{i=1}^{M} \sum_{t=L-h_0+1}^{L-h} U_{it}(h) \end{aligned}$$

where $U_{it}(h) \equiv (X_{it} - X_{i(t+h)})^2 - 2\gamma(h;\theta_0)$, $i \geq 1, t \in \mathbb{Z}$, $1 \leq h \leq h_0$, and where, as a convention, we set $\sum_{t=a}^{b}(\cdot) = 0$ if $a > b$.

Let $\mathbf{U}_{it} = (U_{it}(1), \ldots, U_{it}(h_0))$. Then, by arguments similar to the proof of Theorem F.1, as $M, L \to \infty$,

$$\frac{1}{\sqrt{ML}} \sum_{i=1}^{M} \sum_{t=1}^{L-h} \mathbf{U}_{it} \to^d N\left(\mathbf{0}, \sum_{k\in\mathbb{Z}} E\mathbf{U}_{11}\mathbf{U}_{1k}'\right).$$

Hence, it follows that as $M, L \to \infty$,

$$\sqrt{ML}(2\hat{\gamma}_n(h) - 2\gamma(h;\theta_0) : h = 1, \ldots, h_0) \to^d N\left(\mathbf{0}, \sum_{k\in\mathbb{Z}} E\mathbf{U}_{11}\mathbf{U}_{1k}'\right). \tag{H.15}$$

In particular, this, in conjunction with Condition (F.2.1), implies that $\hat{\Gamma} \to \Gamma_0$ in probability as $M, L \to \infty$. Now, part (b) of Theorem F.2 follows from (H.14) and

(H.15). The proof of part (a) follows by (H.15) and the arguments in the proof of Theorem 2.1 of Lahiri et al. (2002). We omit the routine details. □

CHAPTER IV

CONCLUSION

In this dissertation, we have introduced new resampling methods for the analysis of dependent data. For time series data, we have introduced methods for the study of estimates of covariance parameters, motivated by the study of line width roughness in SEM images. In the area of spatial prediction, we have introduced bootstrap methods for bias correction.

A. Trans-Gaussian Kriging

For Trans-Gaussian data, we have introduced two new Kriging predictors, with bias correction factors calculated via parametric bootstrap methods. We have also discussed how bootstrap methods can be used to estimate the MSPE of these predictors, as well as two existing predictors, the naïve biased predictor and a predictor with an additive correction factor propsed by Cressie (1993).

We have shown that under certain regulatory conditions, the bias for the new predictors is of order $o(1)$, which is smaller than the bias for the existing predictors. Through a simulation study, we have shown that the naïve predictor as well as Cressie's predictor fail as the transformation function $\phi(\cdot)$ becomes more severe, and are greatly outperformed by the bootstrap predictors in this case.

Finally, we have analyzed a real world data set, and seen how Trans-Gaussian Kriging may be utilized to predict energy consumption, a growing concern in today's world.

B. Replicated Time Series Data

For replicated time series data, we have shown how covariance parameters may be estimated using weighted least squares subsampling. Further, we have discussed various bias correction methods for the variance estimator, and demonstrated how bootstrap methods may be used to estimate the variance of covariance estimators. The effectiveness of the Politis and White (2004) method for block length selection was demonstrated.

We discussed the asymptotic properties of the estimators, including discussion of the conditions that made them asymptotically normal, and finally saw their behavior through a simulation study.

REFERENCES

Bandyopadhyay, S. and Lahiri, S. N. (2010), "Resampling-based Bias-corrected Time Series Prediction," *Journal of Statistical Planning and Inference*, 140, 3775–3788.

Box, G. E. P. and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society B*, 26, 211–243.

Bühlmann, P. and Künsch, H. (1999), "Block Length Selection in the Bootstrap for Time Series," *Computational Statistics & Data Analysis*, 31, 295–310.

Constantoudis, V., Patsis, G. P., Tserepi, A., and Gogolides, E. (2003), "Quantification of Line-Edge Roughness of Photoresists. II. Scaling and Fractal Analysis and the Best Roughness Descriptors," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 21, 1019–1026.

Cressie, N. A. C. (1985), "Fitting Variogram Models by Weighted Least Squares," *Mathematical Geology*, 17, 563–586.

— (1990), "The Origins of Kriging," *Mathematical Geology*, 22, 239–252.

— (1993), *Statistics for Spatial Data*, New York: Wiley, 2nd ed.

De Oliveira, V., Kedem, B., and Short, D. A. (1997), "Bayesian Prediction of Transformed Gaussian Random Fields," *Journal of the American Statistical Association*, 92, 1422–1433.

Denker, M. (1986), "Uniform Integrability and the Central Limit Theorem for Strongly Mixing Processes," in *Dependence in Probability and Statistics*, ed. Nahapetian, B. S., Boston: Birkhäuser, pp. 269–289.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1–26.

Hall, P. (1985), "Resampling a Coverage Pattern," *Stochastic Processes and Their Applications*, 20, 231–246.

Hall, P., Horowitz, J., and Jing, B.-Y. (1995), "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561–574.

Krige, D. G. (1951), "A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand," *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52, 119–139.

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics*, 17, 1217–1241.

Lahiri, S. N. (1993), "On the Moving Block Bootstrap Under Long Range Dependence," *Statistics & Probability Letters*, 18, 405–413.

— (2003), *Resampling Methods for Dependent Data*, New York: Springer, 1st ed.

Lahiri, S. N., Lee, Y., and Cressie, N. (2002), "On Asymptotic Distributions and Asymptotic Efficiency of Least Squares Estimators of Spatial Variogram Parameters," *Journal of Statistical Planning and Inference*, 103, 65–85.

Leunissen, L. H. A., Lawrence, W. G., and Ercken, M. (2004), "Line Edge Roughness: Experimental Results Related to a Two-Parameter Model," *Microelectronic Engineering*, 73, 265–270.

Liu, R. Y. and Singh, K. (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of Bootstrap*, eds. LePage, R. and Billard, L., New York: Wiley, pp. 225–248.

Matérn, B. (1960), "Spatial Variation," *Meddelanden fran Statens Skogsforskningsinstitut*, 49.

Matheron, G. (1962), *Traité de Géostatistique Appliquée, Tome I: Mémoires du Bureau de Recherches Géologiques et Minières*, vol. 14, Paris: Editions Technip.

Oklahoma Climatological Survey (2010), *Mesonet*.
URL: *http://www.mesonet.org*

Patel, K., Lahiri, S. N., and Spanos, C. J. (2010), "Robust Estimation of Line Width Roughness (LWR) Parameters," Preprint.

Politis, D. N. and Romano, J. P. (1995), "Bias-Corrected Nonparametric Spectral Estimation," *Journal of Time Series Analysis*, 16, 67–103.

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*, New York: Springer.

Politis, D. N. and White, H. (2004), "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews*, 23, 53–70.

Ristinen, R. A. and Kraushaar, J. P. (2006), *Energy and the Environment*, Hoboken, NJ: Wiley, 2nd ed.

Schabenberger, O. and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Boca Raton: Chapman & Hall/CRC.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Shimizu, K. and Iwase, K. (1987), "Unbiased Estimation of the Autocovariance Function in a Stationary Generalized Lognormal Process," *Communications in Statistics, Theory and Methods*, 16, 2145–2154.

Singh, K. (1981), "On the Asymptotic Accuracy of Efron's Bootstrap," *The Annals of Statistics*, 9, 1187–1195.

Sinha, S. K., Sirota, E. B., Garoff, S., and Stanley, H. B. (1988), "X-ray and Neutron Scattering from Rough Surfaces," *Physical Review B*, 38, 2297–2311.

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

Yamaguchi, T., Yamazaki, K., and Nagase, Masao Namatsu, H. (2003), "Line-Edge Roughness: Characterization and Material Origin," *Japanese Journal of Applied Physics*, 42, 3755–3762.

Yeo, I.-K. and Johnson, R. A. (2000), "A New Family of Power Transformations to Improve Normality or Symmetry," *Biometrika*, 87, 954–959.

Yoshimura, T., Shiraishi, H., Yamamoto, J., and Okazaki, S. (1993), "Nano Edge Roughness in Polymer Resist Patterns," *Applied Physics Letters*, 63, 764–766.

VITA

Krista Dianne Rister graduated from Texas A&M university with a Bachelor of Science in Applied Mathematical Sciences in 2005. She received a Master of Science in Statistics in 2007, also from Texas A&M University. She received her Doctor of Philosophy in Statistics from Texas A&M University in December 2010. She may be reached at Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, or by email at kdrister@gmail.com.

The typist for this thesis was Krista Rister.