Consistency Properties for Growth Model Parameters Under an Infill Asymptotics Domain

DISSERTATION

David T. Mills, Major, USAF
AFIT/DAM/ENC/10-1

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/DAM/ENC/10-1

CONSISTENCY PROPERTIES FOR GROWTH MODEL PARAMETERS UNDER AN INFILL ASYMPTOTICS DOMAIN

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

David T. Mills, B.S., M.S.

Major, USAF

Sep 2010

AFIT/DAM/ENC/10-1

CONSISTENCY PROPERTIES FOR GROWTH MODEL PARAMETERS UNDER AN INFILL ASYMPTOTICS DOMAIN

David T. Mills, B.S., M.S.
Major, USAF

Approved:

Dr. Edward D. White Chairman	Date
Dr. Dursun A. Bulutoglu Member	Date
Dr. Kimberly K. Kinateder Member	Date
Dr. Mark E. Oxley Member	Date

Accepted:

M. U. Thomas Dean, Graduate School of Engineering and Management	Date

AFIT/DAM/ENC/10-1

Abstract

Growth curves are used to model various processes, and are often seen in biological and agricultural studies. Underlying assumptions of many studies are that the process may be sampled forever, and that samples are statistically independent. We instead consider the case where sampling occurs in a finite domain, so that increased sampling forces samples closer together, and also assume a distance-based covariance function. We first prove that, under certain conditions, the mean parameter of a fixed-mean model cannot be estimated within a finite domain. We then numerically consider more complex growth curves, examining sample sizes, sample spacing, and quality of parameter estimates, and close with recommendations to practitioners.

Table of Contents

Page

Acknowledgements

None of this would have been possible without the love and support of my wife and best friend, and I will forever be in her debt. Special thanks to my children, without whom the journey would be adventureless, and somewhat quieter. To our family at St. Andrew United Methodist, especially the blessed Boroff/Miller family, thanks.

My advisor, Dr. White, and my committee, Drs. Bulutoglu, Kinateder, and Oxley, invested much time and effort in this, for which I am very grateful. It is not often that I have been blessed to work so closely with such dedicated professionals, and I do recognize the sacrifices they have made for my benefit.

To the many other excellent educators I've been blessed to work with, especially Professors Kelly Mumbower and Gregory Passty, thanks.

List of Figures

List of Tables

CONSISTENCY PROPERTIES FOR GROWTH MODEL PARAMETERS UNDER AN INFILL ASYMPTOTICS DOMAIN

I. Introduction

1.1 Introduction

Much statistical theory and practice relies on two assumptions; the first that a stochastic process can be sampled infinitely often, with no dependence between samples, and the second that a process continues forever. Under these assumptions an asymptotic sampling result is applied to a finite data set for analysis, and inferences or predictions are made. The results, however, are only as valid as the assumptions.

When either assumption is violated the analysis must properly account for the situation, either through differing analytical techniques or through the interpretation of results. This work examines temporal growth curves under the conditions where both of these assumptions are violated; there is a distance-based dependence between samples, and we assume a process with a finite domain, sometimes called an *Infill Asymptotics* (IA) domain [7]. Within an IA domain, samples can be spread only so far apart, so increasing samples become increasingly closer together.

Without this last requirement, that is if the domain was unbounded, samples could simply be spaced far enough apart that the dependence is negligible, when in practice this might be impossible. Examples of this include a longitudinal study of a childhood illness, as the subjects will not be children for long, or a study of a seasonal growth process on a short-lived animal. If an infinite domain process is

assumed where a finite domain is actually correct, it is important that we consider the impact of the erroneous assumption. These may include overly tight confidence intervals, leading to optimistic beliefs about both parameter estimates and any resulting predictions, or false confidence in model identification.

1.2 Scope

This research is an examination of growth curve parameter estimation under an infill asymptotics domain. We restrict the scope of this work by assuming that we are not required to determine either the form of the growth model or the form of the error (covariance); the form of both is fully specified, although the required growth model parameters are estimated from observed data.

The process we examine consists of three distinct parts. The first two of these are the growth model itself and the error term, and we use these in the general additive form

$$Y_{\vec{\theta},\vec{\rho}}(t,d) = f(t;\vec{\theta}) + \epsilon(t,d;\vec{\rho}) \tag{1}$$

where $Y_{\vec{\theta},\vec{\rho}}(t)$ is the value at time t, and $f(t;\vec{\theta})$ is a deterministic function of t with parameters $\vec{\theta}$. The noise (or error) function $\epsilon(t,d;\vec{\rho})$ has parameters $\vec{\rho}$ and two arguments, t (time) and d. In certain cases, the error is dependent upon the errors elsewhere in the domain. In these cases we can denote the temporal difference $d = |t - t'|$, where t and t' are both within the domain of the model. The error term $\epsilon(t,d;\vec{\rho})$ may not actually be a function of t or d, and if either is omitted the assumption is that the error is independent of that variable. In all cases the parameters $\vec{\theta}$ are estimated; throughout this work we assume that $\vec{\rho}$ is fixed and known. The emphasis is on estimating $\vec{\theta}$ for different growth functions and different forms of $\epsilon(t,d;\vec{\rho})$.

Although these two components are given separately, they are inextricably

linked; the covariance form impacts estimation of the function parameters substantially. Throughout this work the term *model parameters* applies to the parameters of $f(t;\vec{\theta})$ and the term *covariance parameters* refers to the parameters of $\epsilon(t,d;\vec{\rho})$.

The third component of the model is the domain in which the model exists. We often pay little or no attention to the domain of our models, but this may have unintended consequences. Under a finite domain, sampling is restricted to a finite region, so increased sampling decreases the spread between samples. If we assume a distance-based covariance, this decreasing spread in turn affects the covariance and the estimates. Ignoring the possibility of a finite domain structure may lead to incorrect inferences such as choosing the wrong model or overconfidence in parameter estimates.

Under a finite domain much work has been done on determining the covariance structure ϵ and estimating covariance parameters $\vec{\rho}$, while very little has been devoted towards consistent estimators for the function parameters $\vec{\theta}$. Here we instead assume that the covariance is known and we devote our efforts to the estimation of model parameters, $\vec{\theta}$. First we prove that no consistent estimators exist for certain model parameters within a fixed domain, and then give a procedure for optimizing estimator performance based on a variance criterion.

1.3 Organization

Chapter 2 gives some historical context for this problem, and an overview of ongoing work in the field as found in a literature review. Chapter 3 is an examination of one particular error structure for a fixed-mean model within an infill asymptotics (IA) domain, with implications for every steady-state model in a fixed

domain. We begin with the stochastic process $Y_{\vec{\theta},\vec{\rho}}(t)$, defined as

$$Y_{\vec{\theta},\vec{\rho}}(t) = \alpha + \epsilon(t, d; \vec{\rho}) \tag{2}$$

where the only model parameter to be estimated is $\vec{\theta} = \{\alpha\}$. Using a simple linearly-decreasing covariance, we show that consistent estimation of α is not possible. Specifically, we derive a lower bound on the variance of any estimator of α within the IA domain, using the linearly-decreasing covariance. This lower bound is nonzero, and so increasing samples has a diminishing return, with an asymptotic bound which does not allow consistent estimation.

This bound is derived using two theorems we prove: The first gives a closed-form inverse of the covariance matrix, and the second makes use of this inverse to show that the summation of all elements of the inverse is finite. This summation represents the upper bound on the information available in a sample, under certain conditions. The lower bound on the variance is simply the reciprocal of the information available; as the information available is finite, the variance does not decrease beyond a certain amount. This proves that with this covariance α cannot be consistently estimated within a finite domain.

We then extend this result to show that α cannot be consistently estimated within a finite domain, by assuming that increasing the variance does not result in an improved ability to estimate the underlying the model parameters. With any reasonable covariance we may use the linearly-decreasing covariance to undercut the original covariance, resulting in less variance. If α cannot be estimated with the undercutting covariance, we can then assume α cannot be estimated with the original covariance. We then discuss the implications for more complex models.

In Chapter 4 we give a computational example, numerically considering a general model of the form of (1), where all applicable parameters $\vec{\theta}$ will be

estimated. We examine sample sizes, considering the information available with increasing samples. Next we consider spacing of samples to optimize total estimator variance, and give an empirically-optimal spacing of samples. Next, we examine parameter estimation in a curve-fitting context, considering Mean Squared Error and bias of the estimates, and finally we investigate the results a practitioner may encounter, regarding successful model identification. Chapter 5 is a summary, with discussions of implications and possible future research.

II. Background

Let (Ω, β, P) be a probability space, with Ω the set of all possible events, P the probability measure associated with each event, and β the σ-algebra of events. For t in an interval $I \subset \Re$, we consider a real-valued stochastic process given by (1):

$$Y_{\vec{\theta},\vec{\rho}}(t,d) = f(t;\vec{\theta}) + \epsilon(t,d;\vec{\rho})$$

Y is then a random variable. We consider the probability measure P given by the Multivariate Normal Distribution (MVN), where the mean is given by the deterministic function f, and the covariance of observations is given by ϵ. This provides the probability measure for the stochastic process Y, so for a given set of real numbers $\mathbf{c} = (c_1, c_2, ..., c_n)$, $c_i \in \Re$, and times $\{t_1, t_2, ..., t_n\} \in I$ such that $t_i \neq t_j$ for $i \neq j$, the MVN gives the associated cumulative probabilities
$Pr\left[Y(t_1) \leq c_1, Y(t_2) \leq c_2, ..., Y(t_n) \leq c_n\right]$.

To consider P then, we must consider both the mean and covariance; discussion of both the growth curves f for the mean and covariances ϵ follows, but we first define the domain in which the process Y resides. Our interest is focused on processes confined to a finite region, so $I = [a, b]$, where $a < b$ and $0 \leq a, b < \infty$. A process which is sampled from such a domain, even as the sample size increases, may be called an Infill Asymptotics (IA) domain process [7]. Within an IA domain the maximum distance between any two observations is bounded, leading to the definition:

Definition 1 (Infill Asymptotics)**.** *Suppose sampling within the domain of a process were to occur in a manner which spreads the samples as far apart as possible. If:*

$$\lim_{n\to\infty} \max_{i,j\in\{1,2,...,n\}} |t_i - t_j| = C \tag{3}$$

where $C \in \Re^+$ is a finite constant, $i, j \in Z^+ \cup \{0\}$ are indices which order the sample, and $t_i, t_j \in \Re^+ \cup \{0\}$ are time points corresponding to the i^{th} and j^{th} indices, respectively, then the domain of the process is an Infill Asymptotics (IA) domain.

An IA domain is sometimes referred to as a Fixed domain [32], and these terms are used interchangeably. Alternatively, a process which samples from an unbounded domain is referred to as an Increasing domain process:

Definition 2 (Increasing Domain)**.** *Suppose sampling within the domain of a process were to occur in a manner which spreads the samples as far apart as possible. If:*

$$\lim_{n \to \infty} \max_{i,j \in \{1,2,\ldots,n\}} |t_i - t_j| \to \infty \tag{4}$$

where $i, j \in Z^+ \cup \{0\}$ are indices which order the sample, and $t_i, t_j \in \Re^+ \cup \{0\}$ are time points corresponding to the i^{th} and j^{th} indices, respectively, then the domain of the process is an Increasing Domain.

The difference is not trivial. Within an IA domain, increasing samples sizes forces a smaller average distance between samples, and the possibility of a dependence among samples arises. Without a valid assumption of independence among samples, the analytical and inferential techniques must be able to properly account for the dependence structure. If an increasing domain is assumed where an IA domain is actually appropriate, this may cause significant errors in the resulting analysis. In a time-domain problem it is very important to consider the domain, as returning for more samples is not possible when the experiment has ended.

There is no shortage of discussion regarding the IA domain. As noted earlier, within a spatial context there is a large amount of work. However, this work is primarily regarding interpolation methods and estimation of variance components.

In many cases the underlying model is assumed to be constant and unknown. Little to no effort is given to model identification and model parameter estimation.

There is very little work simultaneously considering both growth curve parameter estimation and the IA domain. As we believe this represents a significant deficit, this is the focus of our work. The review of the literature presented here encompasses the portions of spatial statistics that may be relevant to this study of growth curves. In addition, several references are in the field of longitudinal models and data analysis; these are included as a natural application of the IA domain in temporal studies. What little work exists in our direct field of interest is of course given; the results are so sparse and so specialized that the focus of our work is shown to be highly relevant by the lack of previous research.

2.1 Model Identification and Selection

Growth curves are a broad collection of functions; in a statistical context many growth curves may be defined by a differential equation describing known or conjectured growth properties. The response may be discrete (i.e., a count) or continuous (i.e., a model for the weight of a fish). The function itself may be increasing, decreasing, flat, or any combination of these. The specifics of the problem may dictate a model form, or one may have to be hypothesized. The literature offers a great quantity of work regarding growth curves (including model selection, parameter estimation, and computational issues); discussion of several common growth curves follows. For more in-depth coverage the interested reader is referred to [35] or [27]. We restrict the discussion to functions of time as the independent variable, and restrict time to be nonnegative in all cases, so $t \in [0, \infty)$. We also restrict discussion to functions with an asymptotic upper limit. For several of these functions there are multiple parameterizations, and the original formulation

may not always be the clearest (see [15] for an example of this); the source of each parameterization listed is cited to avoid any confusion.

Several growth curves are found repeatedly in the literature, and are often derived from differential equations. The three-parameter logistic curve, as parameterized in [38]

$$f(t; K, a, b) = \frac{K}{1 + \exp(a - bt)} \tag{5}$$

is one of the more commonly found curves, where $K > 0$ is the asymptotic limit, $a \in \Re$ is a location parameter, and $b > 0$ is a rate parameter. Computation reveals that the inflection point (point of maximum growth) is rather inflexible, occurring after $K/2$ of the growth has occurred. In addition, the lower asymptote is always zero, meaning that modeling a process with a initial size (i.e. childhood growth starting at birth) requires that the curve be shifted to account for this, meaning some of the modeled growth has already occurred. This again affects the location of the inflection point. It is, however, not difficult to expand the model to account for an initial size while not shifting the inflection point.

The Gompertz curve, introduced by Benjamin Gompertz in 1825 [15], was initially used for actuarial projections. Winsor's 1932 reparameterization of the Gompertz curve in [38] is given by

$$f(t; K, a, b) = K \exp(-\exp(a - bt)) \tag{6}$$

where $K > 0$ is the asymptotic limit, $a \in \Re$ is a location parameter, and $b > 0$ is a rate parameter. Much like the logistic curve, however, the inflection point is still fixed (now when K/e of the growth has occurred, where e is the natural exponential base), and as with the logistic curve an additional parameter is required for a nonzero lower asymptotic value to not shift the inflection point.

Another growth function, often used for biological models, is the Bertalanffy function of [4], given by

$$f(t; L, l_o, k) = L - (L - l_0)\exp(-kt) \tag{7}$$

where $L > 0$ is the asymptotic limit, $l_0 \geq 0$ is the initial length at time zero, and $k > 0$ is a rate parameter. If $L > l_0$ then this is an increasing function, while if $L < l_0$ this is a decreasing function (and if $L = l_0$ this is the rather uninteresting function $f(t) = L$). As we concentrate on increasing curves, we will use $L > l_0$. Unlike the Logistic and Gompertz curves, the Bertalanffy curve has no inflection point, and does explicitly allow for an initial length. An alternative formulation of this curve, given in [21], uses a hypothetical (negative) time when the length is zero, rather than an initial length at time zero; this is simply a matter of preference.

In 1959 Richards formulated a growth curve which allows for the inflection point to be located anywhere between the asymptotes (or to be excluded), and the Logistic, Gompertz, and Bertalanffy curves are actually special cases of a Richards curve [29]. One parameterization of the Richards curve is given in [27] as

$$f(t; K, a, b) = \frac{K}{(1 + Qexp(a - bt))^{1/\nu}} \tag{8}$$

where the parameters $a \in \Re$, $b > 0$, and $K > 0$ are the same as in the 3-parameter logistics curve, the $Q > 0$ offers another rate parameter, specifically allowing for the rate at $t = 0$ to be set, and $v > 0$ allows some flexibility in the shape of the curve. This model also has its issues, including numerical difficulties in fitting the model to data and in interpreting the meaning of parameter estimates [5], and [27] goes so far as to recommend against using this curve based on these problems. For further description the reader may refer to [27]. Other parameterizations allow for nonzero

starting values as well.

Figure 1 shows examples of these four growth curves, all with an upper asymptote of one and a lower asymptote of zero, on a time scale from $t = 0$ to $t = 1$. The parameters are:

- Gompertz: $\{a, b, K\} = \{0, 5, 1\}$
- Logistic: $\{a, b, K\} = \{0, 5, 1\}$
- Richards: $\{a, b, K, Q, \nu\} = \{0, 5, 1, 2, 2\}$
- Bertalanffy: $\{l_0, k, L\} = \{0, 5, 1\}$

Note the differing function values of, and therefore the differing steepness before and after, the inflection point of the Gompertz, Richards, and Logistic curves. The Bertalanffy curve has the steepest growth at $t = 0$, and is concave down for all $t > 0$.

Research continues on postulating models which are more general still (see [5]), or on recognizing the linkages between the models (see [13]), or on forming a model with a finite time domain [39]. The [39] model brings up an interesting point of clarification regarding the infill asymptotics domain in a growth curve model. The process itself may be finite in time, in which case a finite-domain model is appropriate. This is the clearest corollary to the geospatial problems encountered in the literature. Alternatively, the sampling may be of fixed and finite length but the process itself continues indefinitely. The problem of samples becoming increasingly crowded will remain for both, but the choice of growth curve should reflect the nature of the problem, while the process of fitting the curve to the data will account for the infill asymptotics domain.

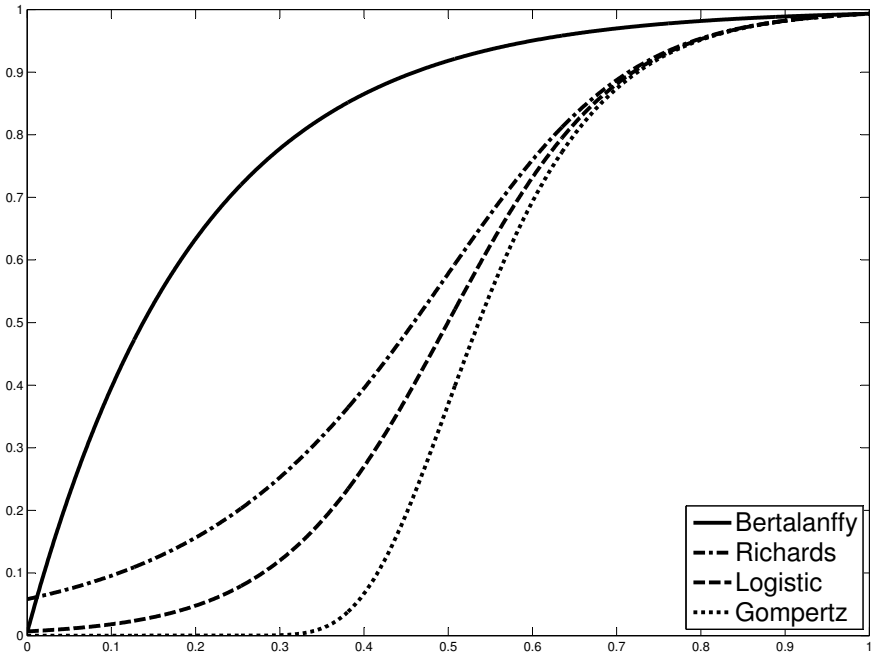

Figure 1. Four Common Growth Curves.

As we cannot consider every model, we will restrict this work to considering the 3-parameter Logistic model. While the others may be applicable we must scope the work to a reasonable quantity; this author believes the 3-parameter Logistic model offers a good array of growth curve shapes to consider.

2.2 Growth Curve Parameter Estimation

After the selection of an appropriate growth curve, either through observation or from foreknowledge of the problem, the applicable parameters must be estimated (that is, the curve fitted to the data). The domain type and covariance structure will often affect the estimation as well, and should be clearly stated as part of the model. Covariance functions and estimation will be discussed in greater depth in sections 2.3.1 and 2.3.2.

There are many methods of estimating parameters. One of the simplest to understand is the Method of Moments, where the theoretical moments of a process (mean, variance, skewness, etc.) are matched to the observed data and the resulting system of equations is solved to provide parameter estimates. Two other common choices for parameter estimation are the methods of Maximum Likelihood Estimation (MLE), which chooses the parameters most likely to have generated the observed data, and the method of Least Squares (LS), minimizing the distance between the observations and the fitted curve. Using these methods, estimation occurs in two distinct steps: First, a function to be optimized is created, linking the observed data to the postulated model. Next, the function is optimized (maximized for MLE, minimized for LS). If the function to be optimized is reasonably simple, a closed-form analytical solution may exist. However, if the function is complex a closed-form solution may be intractable and the optimization will require either an iterative or a heuristic approach.

Due to the nonlinearity of growth curves, parameter estimation for these is more likely to be complex. White's 1998 work [36] used several methods for the first step (including MLE), followed by various optimization algorithms to determine the parameters for several different curves. This was accomplished with an exponential error process (described in Section 2.3.1) within an IA domain. Later White established in [37] that the parameter estimates under these conditions were unbiased but inconsistent.

Growth curve parameter estimation cannot be discussed without discussing different covariance functions and their parameter estimation, as this will generally be accomplished simultaneously; thus, these sections follow immediately.

2.3 Covariance Functions and Parameter Estimation

The covariance structure is an important component of any stochastic process. Different covariance structures may affect selection of estimators and performance of the associated estimators, or a study may be undertaken without a known covariance structure. This section is separated into two subsections; the first examines several common covariance structures within an IA domain, and the second deals with current efforts to determine and estimate the components of an unknown covariance structure.

2.3.1 Covariance Functions.

There are many appropriate covariance functions within an infill asymptotics domain (actually, infinitely many). While the features required will differ from one model to the next, we will always assume that covariances will be nonnegative. In addition to this basic assumption, three attributes are of specific interest:

1. Isotropism
2. Dependence of samples
3. Stationarity

Isotropism refers to a covariance function which has no dependence based on direction. For the IA research in a spatial domain, this must be addressed. In a time-domain process, however, direction is not relevant, and so isotropism is not applicable.

With independent samples, the joint distribution of any two samples is the product of their marginal distributions. Independent samples require that knowledge of one does not offer any information of another. Dependence of samples refers to any violation of independence between samples. In the IA domain, a

common concern is that samples, specifically samples taken close together, may not be independent (although the definition of *close* depends on the specific features of the model, where distance may be measured in time, physical distance, travel costs, etc.). We assume that independence is violated by having a nontrivial covariance function, so that knowledge of one observation does provide some insight into other observations.

An example of this is a simple covariance function, called the *Triangular* or *Tent* covariance, mentioned repeatedly (see [3] or [7]) which addresses the covariance as a constant decreasing function of distance, within a given range, and then zero beyond that. For two samples Y_i and Y_j corresponding to times t_i and t_j respectively, and denoting $d = |t_i - t_j|$:

$$\epsilon(Y_i, Y_j; \sigma^2) = Cov(d; \sigma^2) = \begin{cases} \sigma^2(1 - d/h) & \text{if } d \leq h \\ 0 & \text{if } d > h \end{cases} \tag{9}$$

Figure 2 shows an example of this covariance for $\sigma^2 = 1$ and $h = 0.4$.

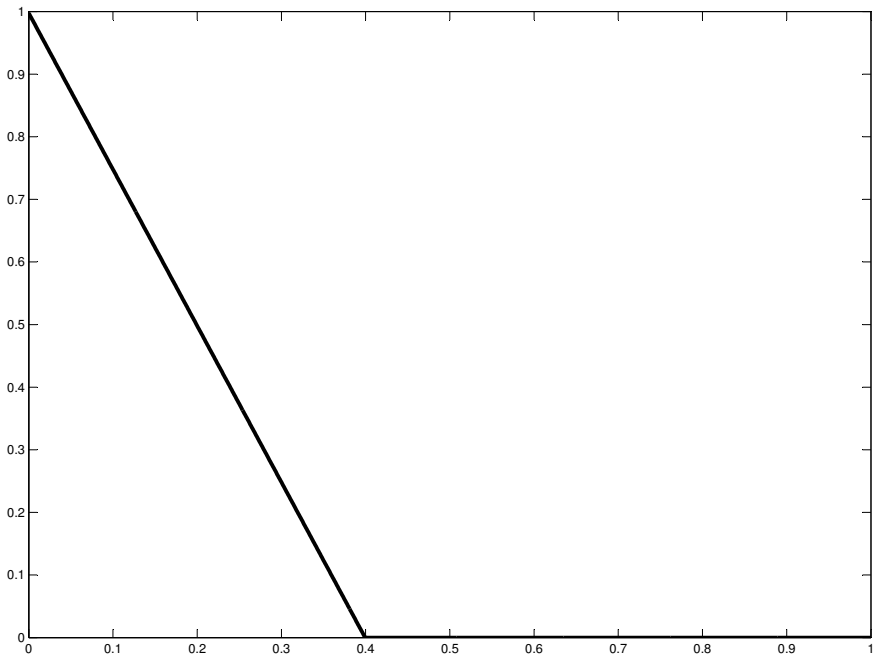

Figure 2. An example of a tent covariance function.

In addition, *stationarity* is often a desirable feature.

Definition 3. *A stationary covariance structure meets the requirement that, for all* $t > 0$*,*

$$Cov(Y_t, Y_{t+d}) = Cov(Y_0, Y_d) \tag{10}$$

where $d \geq 0$*,* $t \geq 0$*, and each* Y_i *is a random observations at time* t_i*.*

Stationarity means that the covariance is not dependent on the *location* on the time axis, and any dependence between two samples is strictly a function of the temporal *distance* between the two time instants. Unlike isotropism, which is a spatial statement, stationarity is a temporal statement and as such is specifically addressed for each covariance used this work. As the covariance is independent of location on the time scale a stationary covariance can be denoted simply Cov(d),

with d the temporal distance between two samples. A stationary covariance, when coupled with a nonstationary growth model, results in a nonstationary growth process.

A frequently-used class of functions to model these features is the Matérn class of covariance functions, a multi-parameter class of functions well-suited to modeling in an IA context [33]. An example of this is given by [12]

$$\epsilon(d;\phi,\nu,\alpha) = Cov(d) = \frac{\phi}{\Gamma(\nu)2^{\nu-1}}(\alpha d)^{\nu}K_{\nu}(\alpha d) \tag{11}$$

where $\phi > 0$ is a scale parameter, $\alpha > 0$ a rate parameter, $\nu > 0$ a smoothness parameter, K_ν is the modified Bessel function of the second kind of order ν, and Γ is the Gamma function. The Matérn class contains several common covariance structures. With $\nu = \frac{1}{2}$, and suitable reparameterizations, the exponential covariance model

$$\epsilon(d;\sigma^2,\lambda) = Cov(d) = \sigma^2\lambda^d,\ \lambda > 0,\ \sigma^2 > 0 \tag{12}$$

can be shown to be within the Matŕrn class (see [1] equation 9.7.2 for details). Figure 3 shows several Matérn covariances.

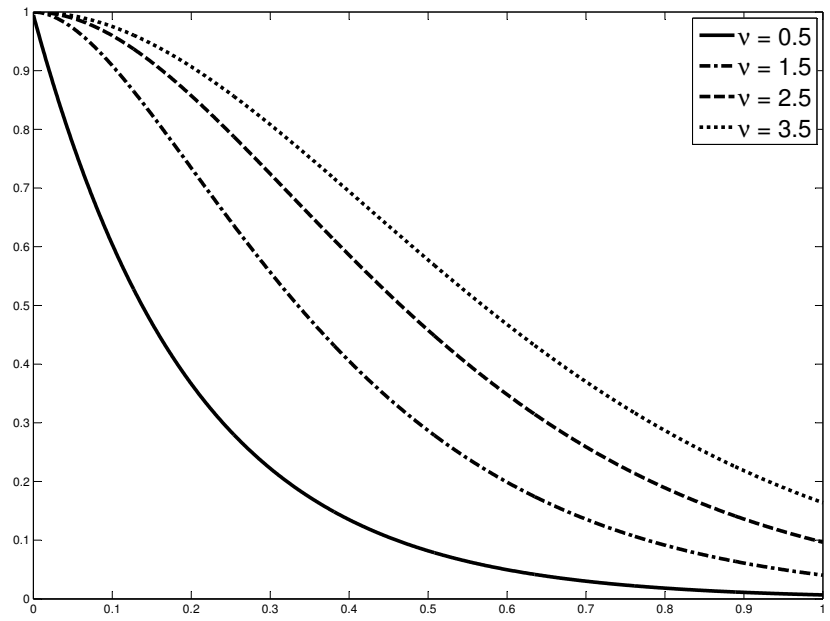

Figure 3. Matérn Covariances: $\phi = 1, \alpha = 5$

With some algebra (12) can be rewritten as

$$\epsilon(d; \sigma^2, \alpha) = Cov(d) = \frac{\sigma^2}{\alpha} e^{-d\alpha},\ \alpha > 0,\ \sigma^2 > 0 \tag{13}$$

which, in a continuous sample space, is also the result of choosing an Ornstein-Uhlenbeck (OU) error (see [11] for details). For a function $r(t)$ which varies about a fixed mean μ_r, the OU error can be expressed as the stochastic differential equation

$$dr(t) = -\alpha(\mu_r - r(t))dt + \sigma dB(t) \tag{14}$$

where $B(t)$ is the standard Brownian Motion [11]. Interesting components of (14) are $r(t)$ and μ_r; if these portions are replaced with a function (for example, a growth curve) and its associated time-dependent mean, the OU error can serve as the error

term whenever the exponential covariance structure is desired. The lowest curve in Figure 3 shows this covariance. Note that, unlike the Tent covariance, the OU covariance decays towards the asymptotic limit of zero, but under this structure there is always a nonzero covariance between two samples, no matter how far apart.

All of the covariance functions shown in Figure 3 decay rather quickly as a function of distance. This may not always be a reasonable assumption. When the dependence decays rather slowly this is often referred to as *long-memory dependence* or *persistence*, and research has addressed this type of problem (see [18] or [8]). We do not address this directly, however we do not specify a specific rate of decay in the theoretical portion of this work.

We close this section with an interesting note. The domain structure is clearly an important feature of any problem, and must be considered in the covariance structure. However, it may not always be clear which asymptotic domain, increasing or infill, is appropriate. Questions about which asymptotic domain to assume, either infill or increasing, is addressed in [41]. Specifically the discussion centers on the estimation of covariance components under each domain, and which estimators to use when the domain structure is in question. The conclusions are that if the domain is in question, use the IA domain as the inferences are more conservative, and will lead to fewer conclusive (and possibly wrong) statements.

2.3.2 Covariance Estimation.

Determination of the covariance form and estimation of the required parameters is one of the most studied fields in spatial statistics. While determining the form of the covariance function is not specifically of interest in this work, as we assume a known form of the covariance, this is a substantial portion of the work done in the field, and thus deserves at least some description.

Zhang showed that under the Matérn covariance structure of (11) certain covariance parameters cannot be estimated consistently, but other quantities based on these parameters can; specifically α and ϕ cannot be consistently estimated separately but the product $\alpha\phi$ can [40]. In the same article Zhang also discusses estimating functions (including a logistic function) which rely on local variation, but only when the growth function parameters are already known.

As with any statistical model, the quantities which can be estimated rely on the data collected. An example of sampling design for covariance parameter estimation under an IA domain can be found in [42], where the authors consider sampling patterns for different objectives. Although the discussion is in the context of a spatial problem, the results regarding minimizing the average kriging variance are echoed in the numeric example we give in Chapter 4.

When the goal is prediction of an unobserved value, a common method used in spatial statistics is kriging [7]. The method of ordinary kriging for interpolative prediction relies on the covariance matrix, and using the method of Restricted Maximum Likelihood the covariance parameters can be estimated without estimating the mean of the function; also, covariance estimation does not generally rely on estimation of a mean parameter, but estimation of a mean parameter does rely on estimation of the covariance [42]. This seems to be one reason that estimation of the mean is so infrequently of interest in spatial statistics.

Misspecification of the covariance function is also a topic of some research. As an example, Furrer et al. [12] misspecify the covariance to gain computational efficiency in a kriging problem, and demonstrates that under some regularity conditions the approach leads to an asymptotically optimal mean squared prediction error. Specifically, the covariance is tapered to only rely on local dependence, neglecting long-range dependence; beyond a certain distance the dependence is

assumed to be zero, much like the tent covariance discussed earlier. The resulting matrices are far more sparse, reducing the complexity involved in solving the linear system required for the kriging predictor. The essence of this approach is that the solutions are not to the same problem, but to problems that are related through the parameters of interest. Under certain conditions the tapered covariance solution does converge to the appropriate values, but only for covariance quantities already shown to have consistent estimators [10], but the estimation of model parameters is, once again, neglected. Other misspecifications may be unintentional, but when the true form of the covariance is unknown the possibility and impact of misspecification must be considered. Stein shows that, under certain conditions, not every misspecification ends with poor prediction performance [31].

Another computationally attractive family of procedures, bootstrap resampling (or simply bootstrapping), normally rely on independence of samples. The dependence induced in an IA domain make this an unreasonable assumption, but a valid bootstrapping approach developed by Loh and Stein accounts for certain types of dependence, and can be found in [23].

2.4 Estimator Consistency under Infill Asymptotics

Much research in the IA domain consists of estimating covariance parameters, and local variation. Difficulties in estimating mean and trend parameters under an infill asymptotics domain is well documented, but follows a slow progression.

In 1984 Morris and Ebey [24] demonstrated that under an IA domain with an AR(1) covariance structure the common estimator

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{15}$$

yields an estimate that is not only inconsistent, but whose variance actually *increases* as the sample size increases beyond a certain point, suggesting that there is an optimal and finite sample size for inferences on μ under this estimator (quite a strange statement).

White went on to demonstrate that the Maximum Likelihood Estimator of a mean-only model with the covariance of (12), and under an IA domain, is unbiased but inconsistent [37]. More specifically, when the data was evenly sampled the asymptotic value for the sample variance of the parameter was bounded above zero. Cressie notes in [7] that the estimator

$$\hat{\mu} = \frac{Y_1 + (1-\lambda)\sum_{i=2}^{n-1} Y_i + Y_n}{n - (n-2)\lambda} \tag{16}$$

which is the same estimator derived by White is indeed the minimum variance unbiased estimator for μ under the OU process covariance of (12). As this is the minimum variance unbiased estimator, it is clear that no consistent estimator exists for this model; either the variance will not vanish as the sample size tends to infinity, or the estimator will be biased.

S. N. Lahiri addressed both model and covariance parameter inconsistency in [22], focusing first on Least Squares estimators and then generalizing to a class of estimators based on smoothness and symmetry conditions of the estimators. The results remain the same as in [37] and [24], under the conditions stated consistent estimation cannot occur. Lahiri's work applies to a larger class of estimators than either [37] or [24], but still does not give a comprehensive conclusion.

Finally, based on the work of Grenander [17] Cressie suggests an inference method for statements on μ with the exponential covariance of (12), $Cov(h) = \sigma^2\lambda^h$, using a heuristic-based sample-size adjustment (which requires foreknowledge of λ),

but makes no claim on the consistency of the estimators themselves. Strangely, this method is based on the estimator of (15) for which the estimator variance is known to increase beyond some finite point, and the sample-size adjustment is an increasing function of the sample size. Cressie also uses similar methods, and assumptions on a known λ, for prediction rather than inference, but in all cases the consistency of the estimators is unaddressed. In a kriging context, Cressie also never shows consistency of trend or mean estimators, and many results directly acknowledge inconsistency, usually in the form of a bias.

2.5 Conclusion

Significant work has been done in parameter estimation, for both covariance and trend parameters, under increasing domain asymptotics. Also, significant work has been done in covariance estimation under an infill asymptotics domain. However, there is a lack of work in the estimation of *function* parameters under an IA domain, often making the assumption that these quantities are already known. While studies have been performed to address computational complexity of parameter estimation under an IA domain, once again these studies address covariance parameters, still leaving questions about function parameters unanswered. Work regarding the estimation of function parameters appears often haphazard, and does not address consistency of the estimators.

White's finding in [37] that the MLE of the mean parameter of model (2) under the OU error process in an IA domain was an interesting and compelling start, based in appropriate theory; indeed, coupling this result with Cressie's comment that this is the minimum variance unbiased estimator guarantees that no consistent estimator exists for this model. However, there was no description of the underlying reasons for the inconsistency (although obviously not from any bias, as that was

addressed). The current work lacks sufficient discussion of model parameter consistency. An in-depth examination of this, within an IA domain, is the next step.

III. Necessary and Sufficient Conditions

3.1 Introduction and Preliminaries

Before the main topic is addressed, discussion of three topics is needed:

- Toeplitz Matrices
- Reasonable covariances
- Cramér-Rao Lower Bound

We present two theorems regarding the Toeplitz matrix produced by the covariance function of (9). In the first theorem we show a closed-form inverse of this matrix under certain conditions, and in the second theorem we show the summation of the elements of the inverse is bounded. We then close with a discussion of the implications of these results in terms of the Cramér-Rao Lower Bound of any estimator of the only model parameter of (2) in an IA domain.

3.1.1 Toeplitz Matrices.

Definition 4 (Toeplitz Matrices)**.** *The matrix $A \in \Re^{n\times n}$ is said to be a Toeplitz Matrix if each entry $a_{i,j}$, $1 \leq i, j \leq n$, is defined solely by $i - j$ [16]. In the case of symmetry, each entry is defined solely by $|i - j|$.*

An example of this is:

$$\begin{pmatrix} a_0 & a_{-1} & a_{-2} & \dots & a_{-m+1} & a_{-m} \\ a_1 & a_0 & a_{-1} & \dots & a_{-m+2} & a_{-m+1} \\ a_2 & a_1 & a_0 & \dots & a_{-m+3} & a_{-m+2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m-1} & a_{m-2} & a_{m-3} & \dots & a_0 & a_{-1} \\ a_m & a_{m-1} & a_{m-2} & \dots & a_1 & a_0 \end{pmatrix} \tag{17}$$

For equally-spaced samples with an isotropic covariance function, the covariance matrix will be a Toeplitz matrix $[a_i]$ with the following properties:

1. $[a_i]$ is positive definite
2. $a_i \in \Re$ for all $1 \leq i \leq n$
3. $a_i \geq 0$ for all $1 \leq i \leq n$
4. $a_k = a_{-k}$ for all k

The first property is required of any covariance matrix, while the second and third properties reflect the real-valued nonnegativity of the covariances. The fourth property is symmetry, which follows from the fact that $Cov(a_i, a_j) = Cov(a_j, a_i)$. The third and fourth properties together make this a Hermitian matrix. The symmetry also allows an even simpler representation of the matrix – specifically a symmetric Toeplitz matrix can be completely reconstructed from only the first row or column.

Although the structure of Toeplitz matrices simplifies many operations, there is no closed-form formula for finding the inverse of a general Toeplitz matrix (although for several specific Toeplitz matrices closed-form solutions are known; the reader is

referred to [9]). W. F. Trench laid forth a recursive algorithm in 1965, which in simplified form is very well described in [43]. This method requires a condition referred to by Zohar as strong nonsingularity, where each of the principal minors is nonsingular. As we are dealing with covariance matrices, which are positive definite, this is a reasonable assumption. Another method of inverting Toeplitz matrices, introduced in [14] by Gohberg and Semencul (in Russian) and demonstrated in [19] by Iohvidov (in English), decomposes the inverse into the difference of the products of lower and upper triangular Toeplitz matrices. For the case of a symmetric, real-valued Toeplitz matrix $C = (c_i)_1^n$ the method requires solving the system of equations represented by

$$\begin{aligned} \sum_{j=1}^{n} x_j c_{i-j+1} &= \delta_{ij}, \quad i = 1, 2, \ldots, n \\ \sum_{j=1}^{n} y_{n-j-1} c_{i-j+1} &= \delta_{ij}, \quad i = 1, 2, \ldots, n \end{aligned} \tag{18}$$

where δ_{ij} is the Kronecker delta, and $x_j, y_n - j - 1$ are the coefficients to be found. A Toeplitz matrix is symmetric about both the upper-right to lower-left diagonal (standard matrix symmetry) and the upper-left to lower-right diagonal (persymmetry). Due to these, (18) reduces to the matrix/vector form as

$$C\vec{x} = e_0 \tag{19}$$

where $\vec{x}$ is an $n \times 1$ vector and e_0 is the first column of the $n \times n$ identity matrix [30]. If a solution for 19 exists, and $x_1 \neq 0$, then C is nonsingular and by the Gohberg-Semencul formula C^{-1} is:

$$C^{-1} = \frac{1}{x_1}\left(WW^T - QQ^T\right) \tag{20}$$

The solution vector $\vec{x} = (x_i)_1^n$ is used to construct the lower-triangular Toeplitz matrices W and Q as follows:

$$W = \begin{pmatrix} x_1 & 0 & 0 & 0 & ... & 0 \\ x_2 & x_1 & 0 & 0 & ... & 0 \\ x_3 & x_2 & x_1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ x_{n-1} & x_{n-2} & x_{n-3} & ... & x_1 & 0 \\ x_n & x_{n-1} & x_{n-2} & ... & x_2 & x_1 \end{pmatrix} \tag{21}$$

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 & ... & 0 \\ x_n & 0 & 0 & 0 & ... & 0 \\ x_{n-1} & x_n & 0 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ x_3 & x_4 & x_5 & ... & 0 & 0 \\ x_2 & x_3 & x_4 & ... & x_n & 0 \end{pmatrix} \tag{22}$$

Finding a solution to (19) is a sufficient but not necessary condition for finding the inverse (for details see [30]). If a solution can be found, the entire inverse is constructible from only the first column; finding this solution then becomes the problem at hand.

3.1.2 Reasonable Covariances.

There are many reasonable covariance structures to use within an IA domain; rather than study each individually, we instead choose to look at a floor on the total variance introduced into an estimator. We do this by choosing a covariance which, for each sample, introduces less variance than the actual covariance in the model. We assume that the covariances modeled are:

- Monotonically decreasing as a function of distance
- Discontinuous at a finite number of values in the IA domain (possibly zero)
- Nonzero for some distance beyond zero

We denote these as *reasonable* covariances. Without these assumptions it is possible to construct a pathological example which, while mathematically interesting, is of no practical use to a practitioner. The third item specifically excludes the assumption of independence, which leads to a trivial and noninteresting result (and which is often an unrealistic assumption anyway).

Given a reasonable covariance we can construct a linear covariance underneath, so that it is dominated by the actual model covariance; for examples see Figure 4 where, in each graph you have a reasonable model covariance, with a dashed line denoting a linear covariance beneath. The covariance curves are generated from Matérn curves, with all but the lower-right curve tapered as described in [12]. The specifics of the curves are rather unimportant, as we only attempt to demonstrate the ability to draw a line underneath.

As the linear (dominated) covariance is at most equal to the model covariance for each set of points sampled, a model with the dominated covariance has less overall variance, and in a signal-to-noise ratio sense we may consider this as a variance floor on the true model. We then examine the dominated covariance to make inferences on the model covariance.

3.1.3 The Cramér-Rao Lower Bound.

When considering the variance of an (unknown) estimator, a natural approach is to consider bounds. Given certain regularity conditions (easily met within a finite domain), the Cramer-Rao Lower Bound (CRLB) is a reasonable candidate. The

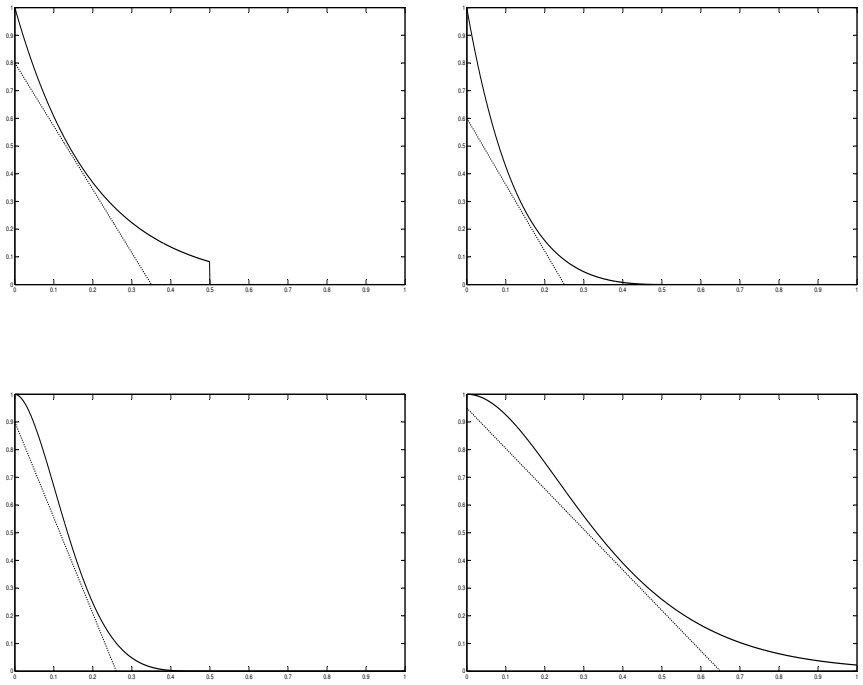

Figure 4. Matérn Covariances (over Linear Covariances)

CRLB gives the lowest variance an unbiased estimator *could* have, based on a given model, rather than considering the variance of a particular estimator. As we seek an unbiased estimator whose variance vanishes on increasing sample sizes, we require an asymptotic CRLB of zero. Note that there is no guarantee that an estimator achieving the CRLB even exists, so a bound of zero alone does not solve this problem. An asymptotically *nonzero* CRLB does, however, show nonexistence of a consistent estimator. For the mean-only model of (2),

$$Y_{\vec{\theta},\vec{\rho}}(t) = \alpha + \epsilon(t, h; \vec{\rho})$$

and assuming multivariate normality with covariance matrix C, the CRLB is [6]:

$$CRLB = \frac{1}{\vec{1_n}' C^{-1} \vec{1_n}} \tag{23}$$

Under these conditions the denominator is the sum of all elements of the inverse of the covariance matrix. An example of this is to revisit the problem of [37], a steady-state model under an IA domain sampled at equal intervals, with an Ornstein-Uhlenbeck covariance. This can be found in Appendix I.

3.2 An Inverse Formula

Theorem 1. *Consider a Toeplitz matrix of the type generated by (9), where* $0 < h < 1$ *is the proportion of nonzero elements in the first column,* n *is the dimension of the matrix, and* $m = nh$ *integer. Then* C *is nonsingular and a closed-form inverse exists.*

Proof. Without loss of generality let $\sigma^2 = 1$. To find the inverse, define

$$x_i = (y_a)_i + (y_b)_i + (r_a)_i + (r_b)_i \tag{24}$$

where

$$m = nh \tag{25}$$

$$k = \lceil 1/h \rceil \tag{26}$$

$$v = n - m(k-1) \tag{27}$$

$$p = \frac{1}{(k+1)(km+m-v+1)} \tag{28}$$

and, indexing from 1 to n, each portion of $\vec{x}$ is defined:

$$(y_a)_i = \begin{cases} \frac{km-i+1}{k+1} & \text{if } i \bmod m = 1 \\ 0 & \text{else} \end{cases} \tag{29}$$

$$(y_b)_i = \begin{cases} -\frac{km-i+2}{k+1} & \text{if } i \bmod m = 2 \\ 0 & \text{else} \end{cases} \tag{30}$$

$$(r_a)_i = \begin{cases} p(km-i+1) & \text{if } i \bmod m = 1 \\ 0 & \text{else} \end{cases} \tag{31}$$

$$(r_b)_i = \begin{cases} p(km-n+i) & \text{if } (n-i) \bmod m = 0 \\ 0 & \text{else} \end{cases} \tag{32}$$

As we use modulo m arithmetic, with required values of up to 2, we require $m > 2$ (and m is already required to be integer). This is reasonable, as in this structure $m = 1$ is the identity matrix and $m = 2$ is a symmetric Toeplitz tridiagonal matrix; neither is difficult to invert. The closed-form inverse is then given by (24), (20), (21), and (22).

As $x_1 \neq 0$, this reduces to showing that (29) – (24) satisfy (19). Matrix multiplication distributes over matrix addition, so:

$$C\vec{x} = C\vec{y}_a + C\vec{y}_b + C\vec{r}_a + C\vec{r}_b \tag{33}$$

The first portion to be computed is $C\vec{y}_a$; the nonzero part of each row of C is banded, and of total width of no more than $2m-1$ (actually, the first m and last v are shorter; all others are exactly $2m-1$ in length). The terms of $\vec{y}_a$ are only nonzero in cycles of length m, beginning with the first element. Each term in $C\vec{y}_a$ will then be the sum of the products of at most two nonzero elements from C with two nonzero elements from $\vec{y}_a$.

First, consider the first m elements in $C\vec{y}_a$. For the i^{th} entry in $C\vec{y}_a$, $1 \leq i \leq m$,

the product is:

$$\begin{aligned}(C\vec{y}_a)_i &= \frac{m-(i-1)}{m}\left(\frac{km}{k+1}\right)+\frac{i-1}{m}\left(\frac{(k-1)m}{k+1}\right)\\ &= \frac{mk-k(i-1)+(i-1)(k-1)}{k+1}\\ &= \frac{mk-(i-1)}{k+1}\end{aligned} \tag{34}$$

Second, consider all but the final v elements. If $h \geq 1/2$, there is no complete cycle, and the first m elements and the final v elements comprise all n elements. If $h < 1/2$, each element down in the product represents a corresponding shift in the row of C used for the computation, and there are $k-2$ complete cycles of length m (which also explains why $h \geq 1/2$ gives no complete cycles). As C is banded, every m elements shifted in C results in a different nonzero pair from $\vec{y}_a$. To capture this pattern, suppose $i = (r-1)m + j$ for $2 \leq r \leq k-2$ and $1 \leq j \leq m$; for each $i \in m+1, n-j-1$ there is exactly one corresponding r, j pair. This gives all but the first m and final v elements, so for $m+1 \leq i \leq n-v$:

$$\begin{aligned}(C\vec{y}_a)_i &= \frac{m-(j-1)}{m}\left(\frac{(k-r+1)m}{k+1}\right)+\frac{j-1}{m}\left(\frac{(k-r)m}{k+1}\right)\\ &= \frac{(k-r+1)(m-(j-1))+(j-1)(k-r)}{k+1}\\ &= \frac{m(k-r)+m-(i-(r-1)m-1)}{k+1}\\ &= \frac{mk-(i-1)}{k+1}\end{aligned} \tag{35}$$

Finally, consider the last v elements. Following the pattern given in the preceding section, for j in the last v elements of $C\vec{y}_a$, the shift is $n-v+j$ rows down in C, but the only nonzero element from $\vec{y}_a$ is the last nonzero element,

$m/(k+1)$. Let $i = n - v + j$ for $1 \leq j \leq v$:

$$\begin{aligned}(C\vec{y}_a)_i &= \frac{m-(j-1)}{m}\left(\frac{m}{k+1}\right) \\ &= \frac{m-(j-1)}{k+1} \\ &= \frac{m-(i-n+v-1)}{k+1}\end{aligned}$$

Now, recalling the definition $v = n - m(k-1)$ from (27), we may substitute this in and continue the simplification:

$$\begin{aligned}(C\vec{y}_a)_i &= \frac{m-(i-n+v-1)}{k+1} \\ &= \frac{m-(i-n+n-m(k-1)-1)}{k+1} \\ &= \frac{mk-(i-1)}{k+1}\end{aligned} \tag{36}$$

In all cases, (34), (35), and (36), the product is identical:

$$(C\vec{y}_a)_i = \frac{mk-(i-1)}{k+1} \tag{37}$$

Next, $(y_b)_i = -(y_a)_{i-1}$ for $i = 2, 3, \ldots, n$; as C is Toeplitz we only have to find the first element of $\vec{y}_b$ separately, and the formula is then:

$$(C\vec{y}_b)_i = \begin{cases} -\frac{mk-k}{k+1} & \text{if } i = 1 \\ -\frac{mk-(i-2)}{k+1} & \text{if } i = 2, 3, \ldots, n \end{cases} \tag{38}$$

The sum of (37) and (38) computes to:

$$(C\vec{y}_a)_i + (C\vec{y}_b)_i = \begin{cases} \frac{k}{k+1} & \text{if } i = 1 \\ \frac{-1}{k+1} & \text{if } i = 2, 3, \ldots, n \end{cases} \tag{39}$$

Note that $\vec{r}_a = p(k+1)\vec{y}_a = \vec{y}_a/(km+m-v+1)$, so:

$$(C\vec{r}_a)_i = \frac{mk-(i-1)}{(k+1)(km+m-v+1)} \tag{40}$$

Finally, $\vec{r}_b$ is the same as $\vec{r}_a$ indexed backward, and C is Toeplitz, so:

$$\begin{aligned}(C\vec{r}_b)_i &= (C\vec{r}_a)_{n-i+1} \\ &= \frac{mk-(n-i)}{(k+1)(km+m-v+1)}\end{aligned} \tag{41}$$

Adding (40) and (41):

$$\begin{aligned}(C\vec{r}_a)_i + (C\vec{r}_b)_i &= \frac{mk-(i-1)+mk-(n-i)}{(k+1)(km+m-v+1)} \\ &= \frac{2mk-n+1}{(k+1)(km+m-v+1)}\end{aligned}$$

Again recalling the definition of v from (27), this sum computes to:

$$\begin{aligned}(C\vec{r}_a)_i + (C\vec{r}_b)_i &= \frac{2mk-n+1}{(k+1)(km+m-v+1)} \\ &= \frac{2mk-n+1}{(k+1)(km+m-(n-m(k-1))+1)} \\ &= \frac{2mk-n+1}{(k+1)(mk+m-n+mk-m+1)} \\ &= \frac{2mk-n+1}{(k+1)(2mk-n+1)} \\ &= \frac{1}{k+1}\end{aligned} \tag{42}$$

Adding (39) to (42) provides the final answer:

$$(C\vec{x})_i = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{if } i = 2,3,\ldots,n \end{cases} \tag{43}$$

This satisfies the requirements of (19), so in conjunction with (20) this is a closed-form inverse. □

3.3 Summation of Inverse Elements

Theorem 2. *Consider a Toeplitz matrix of the type generated by (9), where* $0 < h < 1$ *is the proportion of nonzero elements in the first column,* n *is the dimension of the matrix, and* $m = nh$ *integer. Then the summation of all elements of* C^{-1} *is bounded above by a function of* h*, not dependent upon* n*.*

Proof. For this proof, define $\vec{r}_a$ and $\vec{r}_b$ as before, and:

$$\vec{y} = \vec{y}_a + \vec{y}_b \tag{44}$$

To find the summation, first note that $y_i \neq 0$ requires $i \bmod m \in \{1, 2\}$, and for each i such that $i \bmod m = 1$, the pair y_i, y_{i+1} sums to zero. This results in the following:

$$\sum_{i=t}^{n} y_i = 0 \; \forall \; \{t : t \bmod m \neq 2\} \tag{45}$$

$$\sum_{i=1}^{t} y_i = 0 \; \forall \; \{t : t \bmod m \neq 1\} \tag{46}$$

Based in part on (45) and (46), $\sum_1^t y_i \neq 0$ requires $t \bmod m = 1$, and $\sum_t^n y_i \neq 0$ requires $t \bmod m = 2$. More specifically these values are:

$$\sum_{i=t}^{n} y_i = y_t \; \forall \; \{t : t \bmod m = 2\} \tag{47}$$

$$\sum_{i=1}^{t} y_i = y_t \; \forall \; \{t : t \bmod m = 1\} \tag{48}$$

Finally, note that r^a and r^b are the same vector indexed from opposite ends, so:

$$\sum_{i=1}^{t} r_i^a = \sum_{i=n-t+1}^{n} r_i^b \; \forall \; 1 \leq t \leq n \tag{49}$$

$$\sum_{i=n-t+1}^{n} r_i^a = \sum_{i=1}^{t} r_i^b \; \forall \; 1 \leq t \leq n \tag{50}$$

Recalling the definitions of W and Q from (21) and (22) we form the sums separately:

$$\begin{aligned}
\sum_{i,j} WW^T = {} & x_1 \sum_{i=1}^{n} x_i \\
& + x_2 \sum_{i=1}^{n} x_i + x_1 \sum_{i=1}^{n-1} x_i \\
& + x_3 \sum_{i=1}^{n} x_i + x_2 \sum_{i=1}^{n-1} x_i + x_1 \sum_{i=1}^{n-2} x_i \\
& \vdots \\
& + x_n \sum_{i=1}^{n} x_i + x_{n-1} \sum_{i=1}^{n-1} x_i + ... + x_2 \sum_{i=1}^{2} x_i + x_1 \sum_{i=1}^{1} x_i \\
& = \left(\sum_{i=1}^{n} x_i \right)^2 + \left(\sum_{i=1}^{n-1} x_i \right)^2 + \ldots + \left(\sum_{i=1}^{2} x_i \right)^2 + \left(\sum_{i=1}^{1} x_i \right)^2
\end{aligned} \tag{51}$$

$$
\begin{aligned}
\sum_{i,j} QQ^T &= x_n \sum_{i=2}^{n} x_i \\
&\quad + x_{n-1} \sum_{i=2}^{n} x_i + x_n \sum_{i=3}^{n} x_i \\
&\quad + x_{n-2} \sum_{i=2}^{n} x_i + x_{n-1} \sum_{i=3}^{n} x_i + x_n \sum_{i=4}^{n} x_i \\
&\quad \vdots \\
&\quad + x_2 \sum_{i=2}^{n} x_i + x_3 \sum_{i=3}^{n} x_i + ... + x_{n-1} \sum_{i=n-1}^{n} x_i + x_n \sum_{i=n}^{n} x_i \\
&\quad = \left(\sum_{i=2}^{n} x_i \right)^2 + \left(\sum_{i=3}^{n} x_i \right)^2 + \ldots + \left(\sum_{i=n-1}^{n} x_i \right)^2 + \left(\sum_{i=n}^{n} x_i \right)^2
\end{aligned} \tag{52}
$$

Reversing the sum of QQ^T, arrange the difference as:

$$
\begin{aligned}
\sum_{i,j} WW^T - \sum_{i,j} QQ^T &= \left(\sum_{i=1}^{n} x_i \right)^2 \\
&\quad + \left(\sum_{i=1}^{n-1} x_i \right)^2 - \left(\sum_{i=n}^{n} x_i \right)^2 \\
&\quad + \left(\sum_{i=1}^{n-2} x_i \right)^2 - \left(\sum_{i=n-1}^{n} x_i \right)^2 \\
&\quad \ldots \\
&\quad + \left(\sum_{i=1}^{2} x_i \right)^2 - \left(\sum_{i=3}^{n} x_i \right)^2 \\
&\quad + \left(\sum_{i=1}^{1} x_i \right)^2 - \left(\sum_{i=2}^{n} x_i \right)^2
\end{aligned}
$$

Using the identity $a^2 - b^2 = (a-b)(a+b)$ rewrite this:

$$\begin{aligned}\sum_{i,j} WW^T - \sum_{i,j} QQ^T = &\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right) \\ &+ \left(\sum_{i=1}^{n-1} x_i - \sum_{i=n}^{n} x_i\right)\left(\sum_{i=1}^{n-1} x_i + \sum_{i=n}^{n} x_i\right) \\ &+ \left(\sum_{i=1}^{n-2} x_i - \sum_{i=n-1}^{n} x_i\right)\left(\sum_{i=1}^{n-2} x_i + \sum_{i=n-1}^{n} x_i\right) \\ &\cdots \\ &+ \left(\sum_{i=1}^{2} x_i - \sum_{i=3}^{n} x_i\right)\left(\sum_{i=1}^{2} x_i + \sum_{i=3}^{n} x_i\right) \\ &+ \left(\sum_{i=1}^{1} x_i - \sum_{i=2}^{n} x_i\right)\left(\sum_{i=1}^{1} x_i + \sum_{i=2}^{n} x_i\right)\end{aligned}$$

Note that one factor in each term simplifies to $\sum_{i=1}^{n} x_i$; we can factor this out and rewrite as

$$\sum_{i,j} WW^T - \sum_{i,j} QQ^T = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i + A\right) \tag{53}$$

where A is defined as

$$\begin{aligned}A = &\left(\sum_{i=1}^{n-1} x_i - \sum_{i=n}^{n} x_i\right) + \left(\sum_{i=1}^{n-2} x_i - \sum_{i=n-1}^{n} x_i\right) \\ &+ \left(\sum_{i=1}^{n-3} x_i - \sum_{i=n-2}^{n} x_i\right) + \ldots + \left(\sum_{i=1}^{3} x_i - \sum_{i=4}^{n} x_i\right) \\ &+ \left(\sum_{i=1}^{2} x_i - \sum_{i=3}^{n} x_i\right) + \left(\sum_{i=1}^{1} x_i - \sum_{i=2}^{n} x_i\right)\end{aligned} \tag{54}$$

Consider A in two pieces, $A = A_y + A_r$. For A_y:

$$A_y = \sum_{i=1}^{n-1} y_i - \sum_{i=n}^{n} y_i + \sum_{i=1}^{n-2} y_i - \sum_{i=n-1}^{n} y_i + \sum_{i=1}^{n-3} y_i - \sum_{i=n-2}^{n} y_i + \ldots$$
$$+ \sum_{i=1}^{3} y_i - \sum_{i=4}^{n} y_i + \sum_{i=1}^{2} y_i - \sum_{i=3}^{n} y_i + \sum_{i=1}^{1} y_i - \sum_{i=2}^{n} y_i$$

Rearranging the terms of A_y gives:

$$A_y = \sum_{i=1}^{n-1} y_i + \sum_{i=1}^{n-2} y_i + \sum_{i=1}^{n-3} y_i + \ldots + \sum_{i=1}^{3} y_i + \sum_{i=1}^{2} y_i + \sum_{i=1}^{1} y_i$$
$$- \sum_{i=n}^{n} y_i - \sum_{i=n-1}^{n} y_i - \sum_{i=n-2}^{n} y_i - \ldots - \sum_{i=4}^{n} y_i - \sum_{i=3}^{n} y_i - \sum_{i=2}^{n} y_i$$

By (46) and (48), the nonzero positive-signed sums are those for which $(n-l) \bmod m = 1$, and the value of these is specifically y_{n-l}. There are k values of $l, 0 \leq l \leq n-1$, for which $(n-l) \bmod m = 1$. Likewise, by (45) and (47), the nonzero negative-signed sums are those for which $(n-l) \bmod m = 2$, and the value of these is specifically y_{n-l}. Considering only these nonzero values, the equal

magnitudes of the y_i, y_{i+1} pairs, and the definition of y_i from (44):

$$\begin{aligned}
A_y &= \sum_{i=0}^{k-1} y_{1+im} - \sum_{i=0}^{k-1} y_{2+im} = 2\sum_{i=0}^{k-1} y_{1+im} \\
&= 2\sum_{0}^{k-1} \frac{mk-1-im+1}{k+1} = 2\sum_{0}^{k-1} \frac{mk-im}{k+1} \\
&= 2\frac{m}{k+1}\left(\sum_{0}^{k-1} k - \sum_{0}^{k-1} i\right) = 2\frac{m}{k+1}\left(k^2 - \frac{k(k-1)}{2}\right) \\
&= 2\frac{m}{k+1}\left(k^2 - \frac{k(k-1)}{2}\right) = \frac{m}{k+1}\left(k^2+k\right) \\
&= \frac{m}{k+1}\left(k(k+1)\right) \\
&= mk
\end{aligned} \tag{55}$$

Next, consider A_r:

$$\begin{aligned}
A_r = &\sum_{i=1}^{n-1} r_i^a - \sum_{i=n}^{n} r_i^a + \sum_{i=1}^{n-2} r_i^a - \sum_{i=n-1}^{n} r_i^a + \sum_{i=1}^{n-3} r_i^a - \sum_{i=n-2}^{n} r_i^a + \dots \\
&+ \sum_{i=1}^{3} r_i^a - \sum_{i=4}^{n} r_i^a + \sum_{i=1}^{2} r_i^a - \sum_{i=3}^{n} r_i^a + \sum_{i=1}^{1} r_i^a - \sum_{i=2}^{n} r_i^a \\
&+ \sum_{i=1}^{n-1} r_i^b - \sum_{i=n}^{n} r_i^b + \sum_{i=1}^{n-2} r_i^b - \sum_{i=n-1}^{n} r_i^b + \sum_{i=1}^{n-3} r_i^b - \sum_{i=n-2}^{n} r_i^b + \dots \\
&+ \sum_{i=1}^{3} r_i^b - \sum_{i=4}^{n} r_i^b + \sum_{i=1}^{2} r_i^b - \sum_{i=3}^{n} r_i^b + \sum_{i=1}^{1} r_i^b - \sum_{i=2}^{n} r_i^b
\end{aligned}$$

Rearranging the elements of A_r:

$$\begin{aligned}
A_r = &\sum_{i=1}^{n-1} r_i^a + \sum_{i=1}^{n-2} r_i^a + \sum_{i=1}^{n-3} r_i^a + \ldots + \sum_{i=1}^{3} r_i^a + \sum_{i=1}^{2} r_i^a + \sum_{i=1}^{1} r_i^a \\
&- \sum_{i=2}^{n} r_i^b - \sum_{i=3}^{n} r_i^b - \sum_{i=4}^{n} r_i^b - \ldots - \sum_{i=n-2}^{n} r_i^b - \sum_{i=n-1}^{n} r_i^b - \sum_{i=n}^{n} r_i^b \\
&+ \sum_{i=1}^{n-1} r_i^b + \sum_{i=1}^{n-2} r_i^b + \sum_{i=1}^{n-3} r_i^b + \ldots + \sum_{i=1}^{3} r_i^b + \sum_{i=1}^{2} r_i^b + \sum_{i=1}^{1} r_i^b \\
&- \sum_{i=2}^{n} r_i^a - \sum_{i=3}^{n} r_i^a - \sum_{i=4}^{n} r_i^a - \ldots - \sum_{i=n-2}^{n} r_i^a - \sum_{i=n-1}^{n} r_i^a - \sum_{i=n}^{n} r_i^a
\end{aligned}$$

In this order, (49) shows that the first two rows sum to zero, and (50) shows that the last two rows sum to zero, so:

$$A_r = 0 \tag{56}$$

Combining (55) and (56) then gives that $A = mk + 0 = mk$. The other sum necessary to evaluate (53) is:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} r_i^a + \sum_{i=1}^{n} r_i^b$$

As $\sum_{i=1}^{n} y_i = 0$ trivially, and $\sum_{i=1}^{n} r_i^a = \sum_{i=1}^{n} r_i^b$, we only need consider:

$$\sum_{i=1}^{n} x_i = 2 \sum_{i=1}^{n} r_i^a$$

Now, for $i \bmod m \neq 1$ $(r_a)_i = 0$, and there are k nonzero $(r_a)_i$'s. With this, the

definition of p from (28), and (31):

$$\begin{aligned}
\sum_{i=1}^{n} x_i = 2\sum_{i=1}^{n} r_i^a &= 2\sum_{j=0}^{k-1} r_{jm+1} \\
&= 2\sum_{j=0}^{k-1} p(km - jm) = 2\sum_{j=0}^{k-1} pm(k-j) \\
&= 2pm\left(\sum_{j=0}^{k-1} k - \sum_{j=0}^{k-1} j\right) = 2pm\left(\sum_{j=0}^{k-1} k - \sum_{j=0}^{k-1} j\right) \\
&= 2pm\left(k^2 - \frac{k(k-1)}{2}\right) = pm\left(k^2 + k\right) \\
&= \frac{m}{(k+1)(km+m-v+1)}\left(k\left(k+1\right)\right) \\
&= \frac{km}{(km+m-v+1)}
\end{aligned} \tag{57}$$

Combining (53) with (57), (28), (44), and (31) we have the formula for the sum:

$$\begin{aligned}
\sum_{i,j}\left(WW^T - QQ^T\right) &= \frac{km}{(km+m-v+1)}\left(\frac{km}{(km+m-v+1)} + km\right) \\
&= \frac{km}{(km+m-v+1)}\left(\frac{km + km(km+m-v+1)}{(km+m-v+1)}\right) \\
&= \frac{km}{(km+m-v+1)}\left(\frac{km(km+m-v+2)}{(km+m-v+1)}\right)
\end{aligned} \tag{58}$$

Next, the formula for the first term required in (20):

$$\begin{aligned}
x_1 &= \frac{km}{k+1} + \frac{km}{(k+1)(km+m-v+1)} \\
&= \frac{km(km+m-v+1) + km}{(k+1)(km+m-v+1)} \\
&= \frac{km(km+m-v+2)}{(k+1)(km+m-v+1)}
\end{aligned} \tag{59}$$

Combining (58) and (59):

$$\frac{1}{x_1}\sum_{i,j}\left(WW^T - QQ^T\right) = \frac{km(k+1)}{km+m-v+1} \tag{60}$$

Recalling the definitions $v = n - m(k-1) = n - km + m$ and $m = nh$:

$$\begin{aligned}\frac{1}{x_1}\sum_{i,j}\left(WW^T - QQ^T\right) &= \frac{km(k+1)}{km+m-n+km-m+1} \\ &= \frac{km(k+1)}{2km-n+1} \\ &= \frac{knh(k+1)}{2knh-n+1} \\ &= \frac{kh(k+1)}{2kh-1+1/n}\end{aligned} \tag{61}$$

From this we compute the asymptotic limit

$$lim_{n\to\infty}\left(\vec{1}^T C^{-1}\vec{1}\right) = \frac{kh(k+1)}{2kh-1} \tag{62}$$

and as $1/n > 0$ the bound

$$\left(\vec{1}^T C^{-1}\vec{1}\right) \leq \frac{kh(k+1)}{2kh-1} \tag{63}$$

is not only an asymptotic limit but an upper bound for all finite n. □

3.4 Discussion And Conclusion

The summation of all elements of the inverse of C represents the Fisher Information, and therefore the reciprocal of the Cramér-Rao Lower Bound (CRLB), on the variance of any estimator of the mean parameter of model (2). We did not show that the CRLB is $(2kh-1)/(kh(k+1))$; instead we have shown that the

CRLB certainly cannot be less than $(2kh-1)/(kh(k+1))$, as we have identified a nontrivial, countably infinite sequence for which the CRLB is asymptotically bounded below by $(2kh-1)/(kh(k+1)) > 0$, which is not a function of the sample size. If $h = a/b$ this sequence is $\{b, 2b, 3b, ...\}$, and so what we have shown is that for this covariance structure the CRLB is not asymptotically zero.

More importantly, if we scale this covariance appropriately we can undercut any other (reasonable) covariance. If the lower bound on the variance of any estimator does not vanish, it is reasonable to assume that another variance structure, which *always* introduces more variance, will also not allow an estimator with a vanishing variance.

We know that for a given nonsingular square matrix D and a scalar $\alpha \neq 0$, $(\alpha D)^{-1} = (1/\alpha)D^{-1}$; for $0 < \alpha < 1$, $0 < h < 1$, and $0 < t < 1$ we can construct a covariance of the form:

$$y(t) = \begin{cases} \alpha - \frac{\alpha}{h}t & \text{if } 0 \leq t \leq h \\ 0 & \text{if } t > h \end{cases} \tag{64}$$

For a mean-only model with even sampling, an estimator for the mean of this model has a CRLB bounded below by:

$$\frac{2kh-1}{\alpha(kh(k+1))} > 0 \tag{65}$$

The implications of this are that if we can construct a line below a covariance function, there is no consistent estimator for the mean value of a constant-only model for a model with this covariance within an infill asymptotics domain. Within the constraints of a reasonable covariance, constructing a line underneath is trivial.

While conclusive, this does leave more questions than it answers; what about nontrivial models? What about unevenly spaced samples? While the problems

become theoretically more complex, an examination of the unanswered questions remains interesting. If not addressed theoretically, a numerical inspection is certainly possible. We cannot exhaustively consider all possibilities, but we can consider a reasonable subset of the possibilities. This becomes the next step.

IV. Numeric Exploration

4.1 Introduction

For a mean-only model in an IA domain, we have shown that, with any reasonable nonindependent covariance and evenly spaced samples, no consistent estimator for the mean exists. This does not address the many cases that differ from these requirements:

- Unevenly spaced samples
- Nontrivial models

These changes may seem small, but the complexities build substantially when either case is considered. For unevenly spaced samples the computational advantages of Toeplitz matrices are no longer applicable, and without a mean-only model the Fisher Information is no longer the sum of the elements of the inverse covariance. In the case of the multi-parameter growth models presented earlier (or any multi-parameter model), the Fisher Information is a matrix rather than a scalar. With these complexities in mind we now present a numeric examination of issues arising within an IA domain.

While much of the work presented here was based on the triangular covariance this was only as an undercutting covariance, and it appears only infrequently in modeling examples in literature. We instead base our exploration on the exponential covariance, as it is commonly used within an IA domain. Additionally, if the samples are spaced evenly the resulting covariance has a well-known and very simple closed-form inverse. Recall the definition of the exponential covariance from (12):

$$\epsilon(d; \sigma^2, \rho) = Cov(d) = \sigma^2 \rho^d$$

The growth curves we consider are bounded between zero and one, so for all cases we will use $\sigma^2 = 0.1$ and $\rho = 0.1$, as these represent a balance between so much noise that the curve is buried in the noise, and so little noise that the curve fitting becomes trivial.

We use the A-Optimality criterion based on the Fisher Information Matrix (FIM), which is the trace of the inverse of the FIM [26], as a metric for the information available in a sample. This represents a measure of the sum of the variances of all parameters to be estimated. Other optimality criterion could be used; this criterion is chosen here for its easy interpretation. This criterion requires that the function be a linear transform, and the growth curves presented are generally not linear transforms; however, in the limited domain of an IA model, a polynomial can be constructed to provide a linear transform which is arbitrarily close to the growth curve, and so we assume applicability of the criterion.

For this chapter we consider the Logistic Growth Curve, with several parameter combinations, totaling 9 different curves. From (5), the equation for the 3-parameter logistic curve is:

$$f(t; K, a, b) = \frac{K}{1 + exp(a - bt)}$$

Graph of these are given in Figures 5, 6, and 7.

Within this construct we examine several different items:

- Sample Sizes. Using the A-criterion, we examine sample sizes for several different model parameter combinations. We determine that beyond about 50 samples the returns diminish substantially in all cases.

- Sample Spacing. Using the A-criterion, we examine sample spacing for the same set of model parameter combinations. We determine that spacing

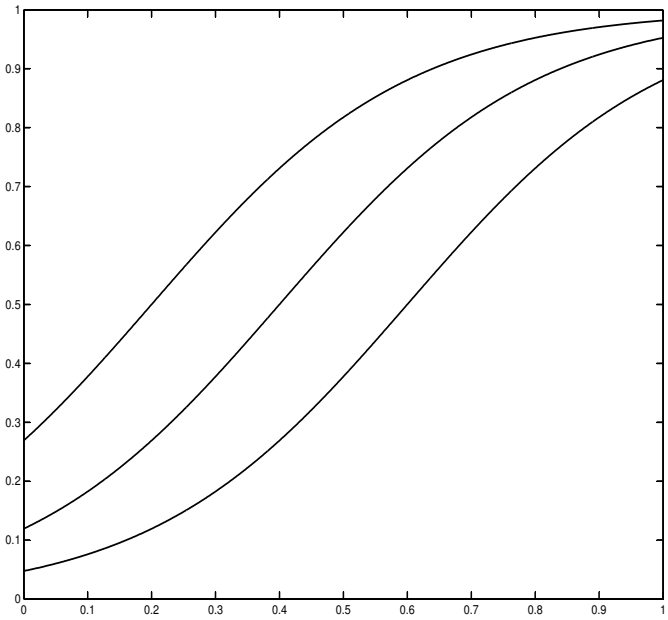

Figure 5. Logistic Curves: $a \in \{1, 2, 3\}, b = 5, K = 1$

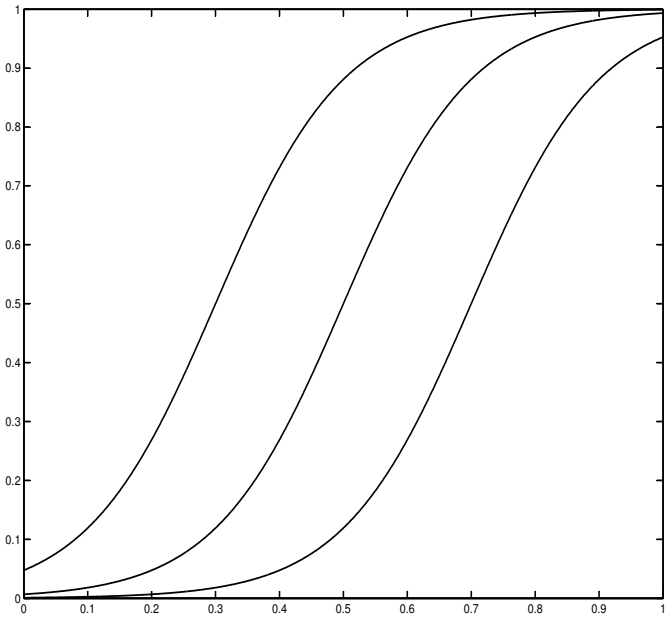

Figure 6. Logistic Curves: $a \in \{3, 5, 7\}, b = 10, K = 1$

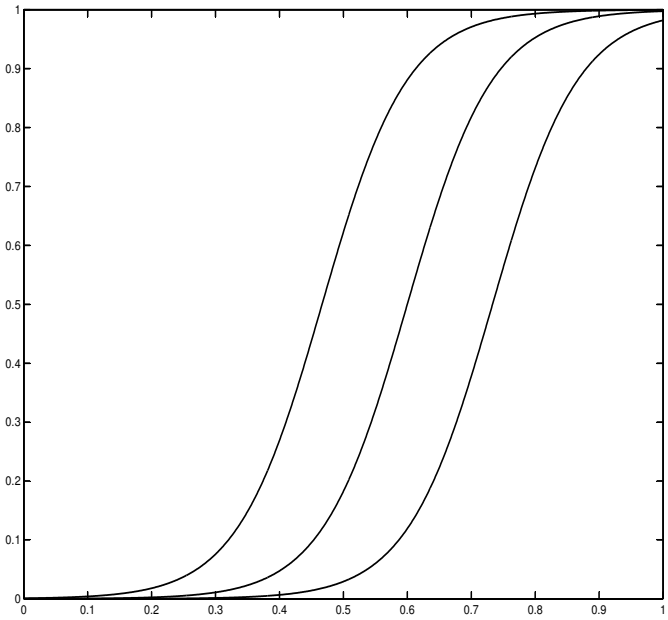

Figure 7. Logistic Curves: $a \in \{7, 9, 11\}, b = 15, K = 1$

samples evenly along the time-axis is best.

- Parameter Estimation. We consider the estimation of model parameters within the IA domain, both in terms of accuracy (MSE) and in terms of how often we may identify the model without reparamaterizing the model. We show that in this example MSE is poor for all parameters, and the ability to fit the model within a reasonable range of the known parameters is also poor.

4.2 Sample Sizes

A crucial decision in any statistical modeling application is to consider the sample size (*a priori* if at all possible). When the samples are independent, it is generally not difficult to set the sample size based on an acceptable level of error, or if cost is an issue there is at least a rather good understanding of the tradeoffs in

increased predictor variance from smaller, and thus lower-cost, sampling. In an IA domain these common approaches are not necessarily valid. We then begin our exploration by considering sample size.

With a mean-only model, we have shown that no consistent estimator exists for the unknown parameter under an IA domain with even sampling; however, if the covariance structure and parameters are known, we may still be able to make valid inferences, acknowledging the fact that the variance will never tend to zero. One approach to this is, instead of considering the variance of the estimator, to consider the marginal return of additional sampling. In the case of covariance matrices with known inverses, such as the exponential covariance and now the triangular covariance, this approach is not particularly difficult. With more complex growth functions, or with covariance matrices without closed-form inverses, the difficulties expand significantly.

The procedure is to first compute the underlying growth curve, using the Logistic curve, then using the exponential covariance compute the appropriate noise function for an evenly-spaced sample of size n; these are then used to find the A-Criterion based on the Fisher Information Matrix. This is a deterministic procedure – no randomness is included. For $n = 5$ samples to $n = 100$ we consider the marginal improvement M for an additional sample:

$$M_i = \frac{A_i - A_{i-1}}{A_{i-1}},\ i = 2, 3, ..., n \tag{66}$$

Note that this is defined as an *a priori* additional sample; for each sample set of size n, the spacing is anchored at $t = 0$ and the samples are then spaced at distances of $1/n$ with the last point at $(n-1)/n$, so an additional sample is not placed within the previous set, but instead defines an entirely new set of data. Introducing an additional sample always induces a new covariance matrix, and the location of the

additional sample defines the new covariance; rather than deal with the continuum of possibilities here we instead define a new sample set constructed to take advantage of the closed-form inverse available with the evenly-spaced sampling. We compute these marginal improvements for a variety of parameter settings, across the different models, with the intention of gaining insight into the effect sample sizes have on the expected quality of estimates. Figures 8, 9, and 10 plot the M_i for $i = 5, 6, 7, ...100$ with several parameter combinations.

It is apparent that, in the settings examined, the marginal return tapers off rather quickly; after a rather small sample size the inferences drawn will not differ much. Note also that the marginal return, while strictly positive, is not necessarily monotonic, as seen when $b = 15$; it appears that when very few samples are spaced far apart, and the growth portion is steep and brief, the growth portion is easy to miss. Adding just one or two more samples then improves the A-criterion substantially. However, even in these cases after large initial improvement, later gains taper off quickly. In all cases the gains past about 50 samples are negligible. With this in mind we may limit the sample sizes considered in subsequent sections to a manageable size.

4.3 Spacing of Samples

After considering sample size, we next consider the spacing of the samples within the context of the model in question; we assume, of course, an IA domain. Without loss of generality we scale the domain from $t = 0$ to $t = 1$, and assume a distance-dependent exponential error between samples. Within an IA domain, and with distance-dependent errors, the spacing of the samples drives the covariance matrix. Two experiments with the same sample size, spaced differently within the domain, will have differing covariances, hence different estimators. The question is

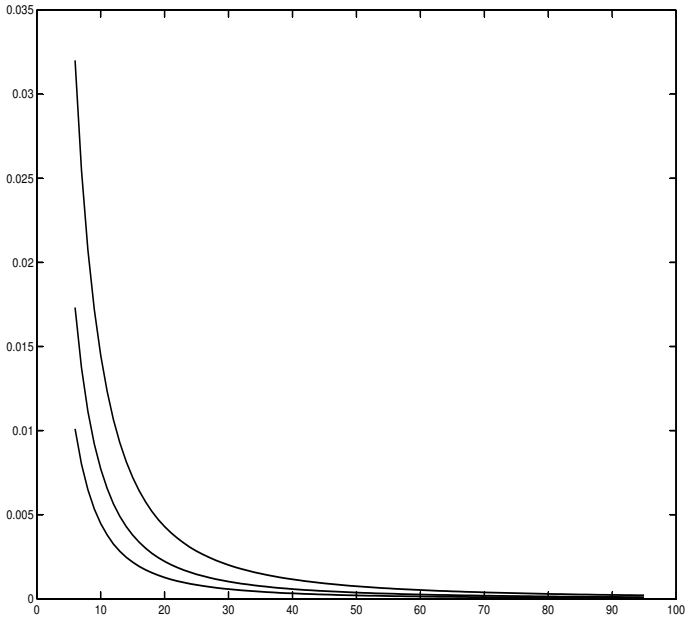

Figure 8. Marginal Return: $a \in \{1, 2, 3\}, b = 5, K = 1$

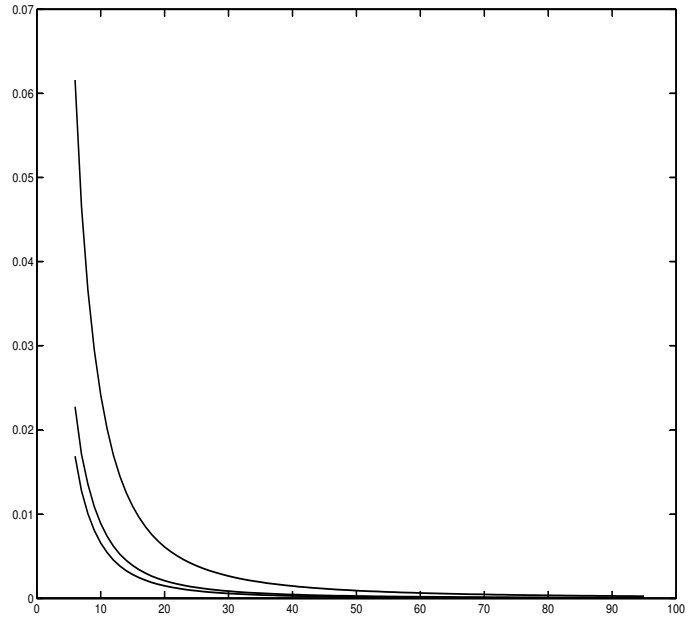

Figure 9. Marginal Return: $a \in \{3, 5, 7\}, b = 10, K = 1$

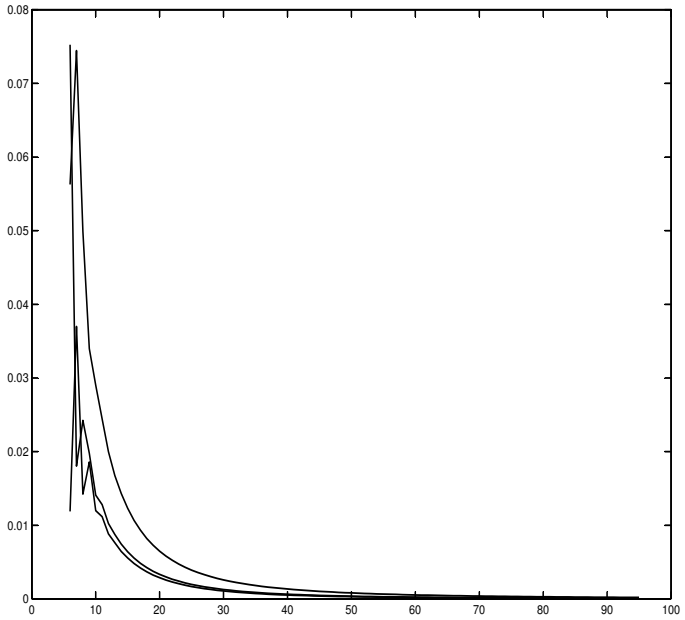

Figure 10. Marginal Return: $a \in \{7, 9, 11\}, b = 15, K = 1$

which sample spacing offers the most information about the parameters.

In an infinite domain, the optimal answer to spacing samples in a distance-dependent covariance is simple; space the samples farther apart to minimize the error. In an IA domain, however, the answer is more nuanced; for a given number of samples, increasing the distance between any two points inherently decreases the distance between another two. Without prior knowledge, we do not know the optimal spacing of samples within the finite domain.

There are four sampling options that can be called even for a sample of size n:

1. First sample at $t = 0$, distance between of $1/n$ (Anchored left)

2. First sample at $t = 1/n$, distance between of $1/n$ (Anchored right)

3. First sample at $t = 1/(n+1)$, distance between of $1/(n+1)$ (Anchored center)

4. First sample at $t = 0$, distance between of $1/(n-1)$ (Anchored on both ends)

We denote these as Even1, Even2, Even3, and Even4, respectively. Asymptotically, these are the same; even for small samples sizes the differences are small. The largest distance between any two corresponding points within the sample is for the first and second options, with a distance of exactly $1/n$ for corresponding points; for obvious reasons this is achieved by these methods actually sharing all but the endpoints.

In addition to these, we also consider a case where the points are evenly spaced and anchored at both ends, but instead of spacing equally on the time axis, we space the points approximately equally along the arc-length of the growth curve itself. Placing the points along the arc-length is computationally prohibitive; instead, we approximate the spacing. To approximate the arc-length, we subdivide the interval into subsections, then take the first derivative in the center of each subsection. The derivative determines the proportion of samples within the subsection. Unlike the other spacing schemes, this requires that the parameters to be estimated are known *a priori*; in application this is of course useless. However, in the context of research we wish to know if any knowledge can be used to better estimate the parameters. Spacing the samples in this manner specifically puts more samples in regions where the growth curve is steepest. In application we will not know the specific parameters, but we may have a general idea of the overall shape. As we again use the A-Criterion with no data, this is a (predictive) deterministic approach; there is no randomness, and each spacing scheme will return the exact same answer each time.

The arc-length spacing places samples more densely on the time-axis where the curve is steepest; we also briefly examined the case where the samples were spaced more densely where the curve was flattest, by using the reciprocal of the derivative but keeping the remainder of the process the same as the arc-length spacing.

Results ranged from merely poor to many orders worse than the remaining schemes. These results are not included, and this scheme pursued no further. Finally, for each parameter setting we also generated 1000 pseudo-random $U(0,1)$ sampling vectors, found the corresponding A-criterion for each, and kept the best; in no case did any of these ever best the last even sampling scheme, and so these results are also not listed. Table 1 gives the results of this experiment for several combinations of a and b, with $K = 1, \sigma^2 = 0.1$, and $\rho = 0.1$. Surprisingly, the evenly-spaced schemes were the best, and in all cases Even2 or Even3 gave the best results. Indeed, each of the even spacing schemes were rather close overall (not surprising, given that they are all very similar). The strategy for practitioners appears to be to space samples as uniformly as possible across the entire time window. However, note that the A-criterion should be indicative of the sums of the variances, and these appear so large in comparison to the parameter values that estimation may be impossible.

4.4 Parameter Estimation: Estimate Quality

Given that the spacing of samples appears to be best left at an even spacing, we next consider the quality of parameter estimates in a simulation setting. To consider

Table 1. A-criterion: Selected Parameter Combinations

a	b	Even1	Even2	Even3	Even4	Line	Best
1.0	5.0	24.9	25.5	24.6	25.8	32.7	Even3
2.0	5.0	27.9	27.4	27.3	27.9	32.1	Even3
3.0	5.0	42.8	41.1	41.1	42.8	43.3	Even3
3.0	10.0	51.3	51.4	51.1	51.5	70.5	Even3
5.0	10.0	60.1	59.7	59.6	60.2	76.9	Even3
7.0	10.0	89.7	85.8	85.8	89.6	90.6	Even2
7.0	15.0	89.4	89.1	89.2	89.4	112.0	Even2
9.0	15.0	103.0	102.0	102.0	103.0	120.0	Even2
11.0	15.0	130.0	126.0	126.0	130.0	149.0	Even2

the quality of parameter estimates, the approach is to simulate pseudo-random data with the appropriate covariance structure, add this to the growth curve, and fit this "observed" data to the model. By doing this repeatedly we can record the parameter estimates and then examine the distribution of these to gain insight. Unlike the previous section, this is pseudo-random. To generate the observations we use the built-in Matlab Normal data generator *normrnd*, and correlate using the method of [20].

Curve fitting may be accomplished using many methods; here we choose the method of Quasi-likelihood Estimation as outlined in [25], a method based on weighted least squares. Myers notes that this iterative method is well-suited to models with correlated responses, appropriate in the processes we examine.

Suppose we have a time-based sample of n data points, and a growth curve with p parameters. Beginning with an initial estimate vector Θ_0, and a known covariance matrix V the estimates are updated by iterating

$$\Theta_{i+1} = \Theta_i + \left(D_i^T V^{-1} D_i\right)^{-1} D_i^T V^{-1} \left(y - \mu_i\right) \tag{67}$$

where y is the observed data, μ_i is the growth function value using the Θ_i values in the growth curve, and D_i is the $n \times p$ matrix of partial derivatives of the curve evaluated at Θ_i. Note that $\left(D_i^T V^{-1} D_i\right)$ is an estimate of the Fisher Information matrix at iteration i. There is a potential complication: If at any step the estimated Fisher Information Matrix is singular (or at least ill-conditioned) the method will fail; in practice this would indicate that one of the parameters has little to no effect on the outcome, and the model would usually be reformulated to not use this parameter, possibly by setting the value to some nominal level, or a different model chosen. In this artificial setting we know the form of the growth curve and that each parameter is significant, so in this case the noise has overcome the model itself.

Reformulating the model to exclude one parameter, however, alters the estimates of the remaining parameters. As we are attempting to consider the distribution of parameter estimates, we are left with no choice but to discard a data set where this occurs. This will be explored in greater depth in the next section.

For sample spacing, we assume, as indicated by the A-criterion experiment, that even spacing is optimal; we then consider spacing evenly, but in addition to sampling across the entire domain we consider the case where the experiment is restricted to the left 3/4, or 75%, of the domain (L75) and the right 3/4 of the domain (R75), while the full domain is denoted F. The optimization terminates when some predetermined criteria is met:

- Ill-conditioned estimated FIM (condition number smaller than 100 times Matlab's default ϵ, $100 * \epsilon \approx 2.2 * 10^{-14}$); discard the data set
- Excess iterations (more than 200 iterations); discard the data set
- Successful convergence (step size $\leq 10^{-9}$)

Figure 11 depicts one step of this algorithm, and Tables 2, 3, and 4 give the results of this for a 3-parameter logistic growth curve with several parameter combinations, chosen for the purpose of showing an array of sigmoidal shapes. Based on the sample sizes discussed in section 4.2 we choose $n = 50$, replicate 500 times, and then compute the summary statistics. Note that, in this context, the MSE for each parameter is computed using the known value of the parameter. Normally the MSE is computing from each sample to the end-result estimate, but in this case we know the actual value. We choose to report MSE as it takes into account both bias and variance of the estimate.

Several things are immediately apparent:

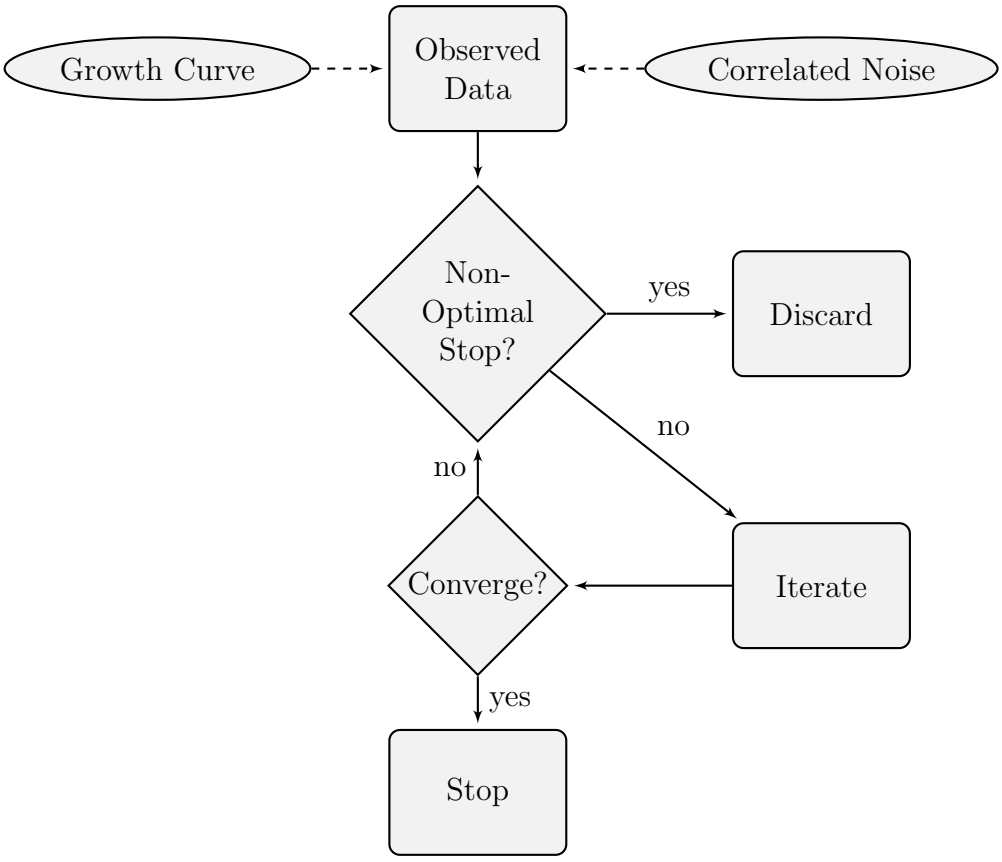

Figure 11. Optimization Algorithm

Table 2. Summary Statistics: $a \in \{1, 2, 3\}$**,** $b = 5$**,** $K = 1$

	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (1, 5, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	16.48	126.3	1072
$MSE_{\hat{b}}$	443.7	2200	$1.5 * 10^4$
$MSE_{\hat{K}}$	0.0622	0.05	0.0682
Bias($\hat{a}$)	0.146	0.783	-20.7
Bias($\hat{b}$)	-3.09	-5.89	-77.7
Bias($\hat{K}$)	-0.0346	-0.0097	0.074
A-Crit.	24.92	31.56	73.88
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (2, 5, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	0.592	5.891	305
$MSE_{\hat{b}}$	43.85	140.8	4386
$MSE_{\hat{K}}$	0.053	0.184	0.0476
Bias($\hat{a}$)	-0.211	-0.289	-6.07
Bias($\hat{b}$)	-0.734	-0.872	-22.73
Bias($\hat{K}$)	-0.0594	-0.056	0.1252
A-Crit.	27.89	43.01	36.30
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (3, 5, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	4.871	27.35	8.78
$MSE_{\hat{b}}$	216.7	1562	119.5
$MSE_{\hat{K}}$	3.862	0.1352	2.196
Bias($\hat{a}$)	-0.52	-0.806	-0.737
Bias($\hat{b}$)	-1.50	-4.368	-1.69
Bias($\hat{K}$)	-0.1818	0.1265	-0.150
A-Crit.	42.80	97.10	41.94

Table 3. Summary Statistics: $a \in \{3, 5, 7\}$**,** $b = 10$**,** $K = 1$

	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (3, 10, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	0.627	0.8147	634.7
$MSE_{\hat{b}}$	5.888	7.953	8209
$MSE_{\hat{K}}$	0.003	0.0061	0.0229
Bias($\hat{a}$)	-0.197	-0.2664	-10.94
Bias($\hat{b}$)	-0.71	-0.850	-43.17
Bias($\hat{K}$)	-0.0040	0.0023	0.0583
A-Crit.	51.28	55.68	71.59
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (5, 10, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	1.473	1.885	2.783
$MSE_{\hat{b}}$	5.620	7.867	10.06
$MSE_{\hat{K}}$	0.029	0.012	0.00767
Bias($\hat{a}$)	-0.275	-0.425	-0.580
Bias($\hat{b}$)	-0.534	-0.881	-1.02
Bias($\hat{K}$)	-0.0043	-0.009	-0.0022
A-Crit.	61.13	86.55	62.39
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (7, 10, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	2.988	212.4	4.037
$MSE_{\hat{b}}$	84.37	2033	8.199
$MSE_{\hat{K}}$	0.048	0.242	0.00520
Bias($\hat{a}$)	-0.29	-0.96	-0.569
Bias($\hat{b}$)	-0.84	-4.36	-0.803
Bias($\hat{K}$)	-0.0206	0.075	-0.0044
A-Crit.	89.66	408.37	86.32

Table 4. Summary Statistics: $a \in \{7, 9, 11\}$**,** $b = 15$**,** $K = 1$

	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (7, 15, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	1.85	2.075	3.123
$MSE_{\hat{b}}$	8.092	13.69	12.74
$MSE_{\hat{K}}$	0.0026	0.0037	0.0042
Bias($\hat{a}$)	-0.27	-0.228	-0.622
Bias($\hat{b}$)	-0.593	-0.561	-1.260
Bias($\hat{K}$)	0.00193	-0.0066	0.00859
A-Crit.	89.42	99.67	92.99
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (9, 15, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	3.059	4.562	2.989
$MSE_{\hat{b}}$	8.5	13.67	7.992
$MSE_{\hat{K}}$	0.0021	0.021	0.00247
Bias($\hat{a}$)	-0.234	-0.366	-0.252
Bias($\hat{b}$)	-0.437	-0.565	-0.446
Bias($\hat{K}$)	-0.0061	-0.027	-0.0045
A-Crit.	103.03	175.29	102.33
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (11, 15, 1)$		
	F	L75	R75
$MSE_{\hat{a}}$	4.873	457.1	4.752
$MSE_{\hat{b}}$	9.028	2621	8.697
$MSE_{\hat{K}}$	0.003	0.273	0.0022
Bias($\hat{a}$)	-0.421	0.321	-0.435
Bias($\hat{b}$)	-0.576	0.162	-0.597
Bias($\hat{K}$)	-0.0047	0.196	-0.00574
A-Crit.	129.92	1096	125.71

- The MSE's computed for a and b are generally rather large as compared to the values of the parameters themselves.

- Estimates for K are generally better (less bias, lower MSE) than for a and b, by orders of magnitude. This is very interesting, as a good argument can be made that the theoretical work regarding the mean-only model applies to the estimators for K, but no such argument holds for the other (intrinsically nonlinear) parameters.

- The A-criterion was only a rough indicator of the final result; noting that the A-criterion should indicate the sum of the variances, and MSE is normally a good estimator for MSE. However, as we knew the actual value beforehand, our MSE is not the standard definition of MSE. In general when one A-criterion value was substantially worse than the others, the estimates were also worse.

- Using the full domain for sampling generally resulted in better estimates (no surprise). Under certain combinations of parameters the difference is drastic.

The inaccurate parameter estimates encountered leads to the question of how often we may be reasonably satisfied with our estimates. We next consider this question.

4.5 Parameter Estimation: Estimators with Restricted Range

In the previous model-fitting many of the estimates were wildly inaccurate; suppose however that some reasonable range of the parameters is already known, maybe by previous experience or some sort of physical constraints. One option is to restart the fitting procedure with a different initial estimate; this may result in an improved estimate. In this section we try exactly that, by specifying a constraint

box, and we also track the number of times the procedure is restarted. After some maximum number of restarts we are, again, forced to discard the data set. We consider a constraint box bounding a and b between zero and twice their actual known values, and K between 0.25 and 2; K near zero is a flat line, and as we maintain a fixed value of $K = 1$ this is a reasonable range of acceptable values.

The process for one set of data is shown in Figure 12, and is largely unchanged from the previous process. Noise data (pseudorandom) and model data (deterministic) combine to simulate observations as before. Next the decision block for *Non-Optimal Stop*; this is a catch-all for situations which cannot be overcome:

- Ill-conditioned estimated FIM (condition number smaller than 100 times the Matlab default ϵ, $100 * \epsilon \approx 2.2 * 10^{-14}$)
- Excess iterations (more than 200 iterations)
- Excess restarts (more than 20 restarts)
- One or more parameters beyond acceptable range
- Excessive step size (more than 3 times the sum of all parameter ranges)

The biggest difference is the restart step. If the Non-Optimal Stop criteria are met, rather than discard the data set we attempt a restart with different initial values. After this the optimization iteration continues, using (67). If after a maximum number of restarts no successful optimization occurs we then discard the data set. A successful optimization terminates at iteration j when $|\Theta_j - \Theta_{j-1}| < 10^{-9}$, and all parameter estimates are within the acceptable range.

Computing the MSE and bias of the parameters in this process is useless; we have discarded results outside of a predetermined box, artificially reducing the MSE and possibly affecting the bias as well. The major insight to be gained here is the proportion of data which cannot be fitted to the model which actually generated it.

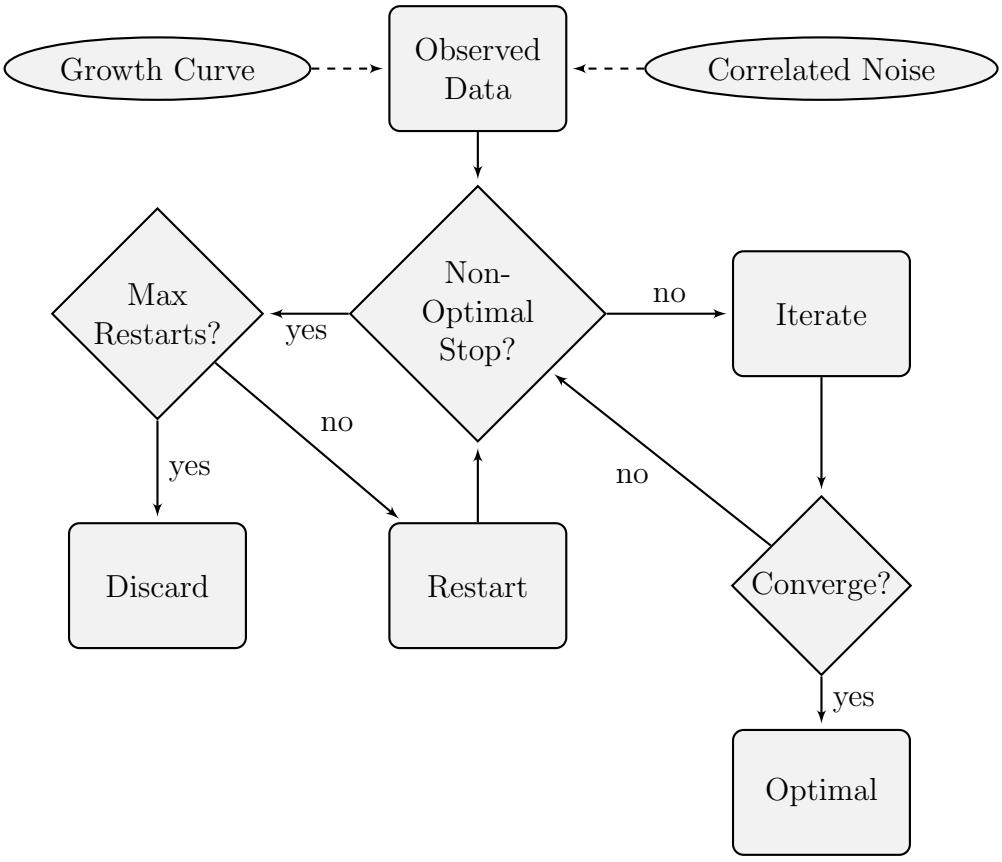

Figure 12. Restricted Range Optimization Algorithm

This algorithm was run 1000 times at each parameter combination; the results are given in Tables 5, 6, and 7. The results are, unfortunately, not significantly better than before. Data requiring a restart did not often gain from it in our example; in this context excess noise apparently has more impact than can be overcome by choosing a differing initial point.

The problems seen here in the data fitting, both in the unrestricted and the restricted parameter ranges, are described in [27] as indicative of an overparameterized model; as noted, in practice a reduced model would probably be the choice to address this. These problems do preclude a direct comparison of the MSE of the simulation to the A-criterion calculated beforehand. This does not detract from the A-criterion as an *a priori* scheme to determine the information in a sample, and the author believes that in practice the sampling schemes indicated by the A-criterion will prove to be superior to other designs.

4.6 Discussion and Conclusions

The numeric exploration of this problem is rather disheartening. We initially assumed that $\sigma^2 = 0.1$ and $\rho = 0.1$ would provide a manageable amount of variance, while still allowing an exploration of curve fitting within an IA domain. With the covariance assumptions given, estimation of the model parameters is at best troublesome:

- Even with the form of the model known in advance, a large portion of the generated data sets could not be fit to the model.
- More samples are unlikely to help.
- Spacing the samples differently is also unlikely to help.
- While it can be argued that our variance was large, it is not difficult to find

Table 5. Fitted and Discarded: $a \in \{1,2,3\}$**,** $b = 5$**,** $K = 1$

	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (1, 5, 1)$		
	F	L75	R75
Discarded	411	543	845
Fitted; no restart	556	435	150
Successful restart	33	22	5
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (2, 5, 1)$		
	F	L75	R75
Discarded	311	499	659
Fitted; no restart	665	488	330
Successful restart	24	13	11
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (3, 5, 1)$		
	F	L75	R75
Discarded	434	696	599
Fitted; no restart	544	298	384
Successful restart	22	6	17

Table 6. Fitted and Discarded: $a \in \{3,5,7\}$**,** $b = 10$**,** $K = 1$

	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (3, 10, 1)$		
	F	L75	R75
Discarded	147	201	624
Fitted; no restart	826	760	354
Successful restart	27	39	22
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (5, 10, 1)$		
	F	L75	R75
Discarded	360	477	379
Fitted; no restart	596	488	588
Successful restart	44	35	33
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (7, 10, 1)$		
	F	L75	R75
Discarded	516	787	542
Fitted; no restart	426	195	419
Successful restart	58	18	39

Table 7. Fitted and Discarded: $a \in \{7, 9, 11\}$**,** $b = 15$**,** $K = 1$

	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (7, 15, 1)$		
	F	L75	R75
Discarded	398	451	433
Fitted; no restart	537	497	512
Successful restart	65	52	55
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (9, 15, 1)$		
	F	L75	R75
Discarded	525	666	531
Fitted; no restart	450	308	431
Successful restart	50	26	38
	$(\sigma^2, \rho) = (0.1, 0.1)$ $(a, b, K) = (11, 15, 1)$		
	F	L75	R75
Discarded	637	873	629
Fitted; no restart	334	117	326
Successful restart	29	10	45

examples in the literature with large variances (i.e. [18] for a long-range dependence), and in actual applications variance is not under the control of the experimenter.

What then is a practitioner to do? It is not enough to ignore the problems inherent and revert to assumptions which oversimplify the problem at hand. When such a model is to be examined, there are a few steps that can be taken:

- First, and most importantly, consider the domain of the model. If an IA domain is or even may be appropriate, do not use the assumptions of an increasing domain, as this will invalidate most inferences. Recognize rather than ignore the inherent difficulties of the IA domain.

- Nonintuitive results seem to be the norm rather than the exception; when faced with this situation a practitioner may want to simulate several scenarios

before the actual data collection activities are begun, so as to anticipate these situations.

- Space samples evenly along the time-axis to minimize covariance, if estimating model parameters is the main goal. If covariance parameters are also to be estimated simultaneously, *a priori* simulations with guessed values, followed by an optimization scheme to determine a good spacing scheme, may be an acceptable approach.

- Recognize the diminishing returns from additional sampling; if possible use any *a priori* knowledge to determine a reasonable sample size.

- Unless the model is known, do not assume the power to determine the form of the model from the data.

- Approach inferences with caution; do not assume that your model form is correct, and do not assume an asymptotically vanishing variance for any parameter.

This work has revealed many difficulties unexpected by this author; the IA domain brings with it many unforeseen results, but recognizing the strange behaviors present allow for a much more informed analysis. It is a mistake to assume that the behaviors encountered in the increasing domain apply to models in an IA domain, and it is our opinion that statisticians need to carefully analyze all assumptions carefully when such a problem is encountered. To ignore this is to risk invalid analysis.

V. Summary and Discussion

5.1 Unique Contributions

There are four unique contributions of this work. They are:

- Closure on the question of a consistent estimator for a fixed mean in an IA domain, specifically that even with this simplest model there is no MSE-consistent estimator for the one model parameter
- Exploration of spacing schemes for growth curve parameter estimation within an IA domain; equally spacing samples along the time-axis appears to be the best option for model parameter estimation
- Demonstrating the use of the A-Criterion as a proxy for parameter estimator quality
- An inverse for a specific Toeplitz Matrix

5.1.1 Estimator consistency in an IA domain.

Difficulties in estimating a fixed mean within an IA domain have been addressed before; [24] demonstrated strange behavior for a common estimator, specifically variance increasing as samples increased beyond a certain finite point, and [37] showed that the MLE was an unbiased, but not consistent, estimator when the covariance structure was an exponential covariance. Attempting to find the necessary and sufficient conditions for a consistent estimator to exist was the original intent of this work. However through simulation and experimentation it became readily apparent that these conditions, except the trivial case of independence, may not exist.

In research we must go where the facts lead us, even when those facts lead us to a different conclusion than first expected. We were first attempting to find the necessary and sufficient conditions for consistent estimators to exist under an IA domain. Instead, we have extended previous work by proving that a fixed-mean model in an IA domain, with any reasonable covariance structure other than uncorrelated, has no consistent estimator for the unknown mean. This extends the previous work by considering all estimators rather than just one, and considering all covariances, not just one. The necessary and sufficient condition for a consistent estimator to exist within an IA domain is that the covariance be trivial – each point uncorrelated with all others.

In addition, this has implications for models other than fixed-mean. As an example, consider again the three-parameter logistic model. In an epsilon-delta fashion, for any $\epsilon > 0$ there exists a point in time δ, beyond which the remaining growth in a finite domain is less than ϵ (so the model is effectively in steady-state). When this ϵ is less than the CRLB associated with the model covariance, attempts to estimate the asymptotic upper limit will be overwhelmed by the estimator variance. Indeed, any possibility of consistent estimation of the ratio of parameters K/a within this model must be conditioned on $b \neq 0$, as this would become a mean-only model if $b = 0$. Similar implications for other models are also apparent.

The grander implications are also troubling. If a practitioner mistakenly models an IA domain as increasing domain, all inferences are likely to be wrong, unless there is true independence of samples. If cost of sampling is an issue, time or money may be wasted on increasing samples with little or no additional information, or predictions of reduced variance and improved point estimates may be false. This should drive home the need to consider the domain the model resides in, not just the exact model and covariance form.

Finally, consider that the fixed-mean model is by far the simplest parametric model available. Adding complexity does not often make estimation simpler; estimating parameters in a more complex model is unlikely to be possible if even the single parameter of the fixed-mean is unestimable.

5.1.2 Exploring spacing schemes.

After proving that under equally spaced sampling consistent estimation of a fixed mean is not possible, we must consider the limitations of this: Specifically the questions of unequal spacing and nontrivial models were unaddressed in theory. With unequal spacing, we lose the advantages of the Toeplitz matrix as a covariance, and with nontrivial models we then must perform a weighted sum of the elements of the covariance to find the Cramér-Rao bound.

We chose to consider this computationally. We chose an exponential covariance, and a nontrivial growth curve, and numerically explored the estimation of parameters based on differing spacing of the samples. As we considered multiparameter models, the A-Optimality criterion was chosen due to easy interpretation, and because it can be considered to be a multiparameter version of the CRLB. While this approach does not constitute a proof, the results suggest that adding model complexity does not make parameter estimation easier.

The results of this experiment were rather surprising; while in some cases a pattern emerged of what a *bad* spacing scheme might look like, none of the *good* schemes were better than equally-spaced sampling. This has been suggested in spatial statistics where the average kriging prediction variance was optimized for a 2-D space (see [42]), but for growth curve parameter estimation this appears to be a new result.

5.1.3 Estimate accuracy.

While the A-Optimality criterion suggested some spacing schemes to be better than others, the true measure of performance here is to examine the results of parameter estimates and correct model identification, using the spacing schemes suggested by the A-criterion. The A-criterion is based on linear transforms, which we readily admit is violated in our analysis. However, the empirical results obtained uphold the even spacing scheme universally suggested.

5.1.4 An inverse for a specific Toeplitz matrix.

The author never intended to find a new inversion formula for any matrix. However, as it became clear that such an inverse would be useful for the computation of a Cramér-Rao Lower Bound it became a priority.

Many inversion formulas exist for specific Toeplitz matrices (see [34] and [9] for examples), and this specific matrix was claimed to be inverted by [28] and [2]; careful reading of these works revealed that no solution was actually given for a general case, instead alluding to the existence of a solution which could, with proper effort, be found. That effort has been provided here. Although finding this inverse was not the primary goal, the inverse itself has value. Specifically, it is this closed-form inverse which allows for the undercutting variance approach to the mean-only estimator problem. In addition, the value of this inverse is emphasized by two previously published, but unsuccessful, inversion attempts.

5.2 Future Research

The limitations of the preceding work lay the foundation for future suggested research. Specifically, in theoretical work we have addressed:

- A mean-only model

- Equally-spaced sampling
- A variance floor

We did address departures from the first two aspects numerically, in specific examples; however, there is no guarantee that these examples actually cover the space, that the sample sizes, growth curves and parameters demonstrated, variance parameters, etc. actually describes the problem. The empirical suggestions are tantalizing, and the results do seem similar to the theoretical results, but a true theoretical conclusion cannot be drawn based on the work here.

The third aspect above is also of interest. In the extension of White's work in Appendix I we calculated an exact bound on the variance of a mean-only estimator with an exponential covariance. In Chapter 3 we calculated an exact bound on the variance for the same model with a linear covariance. Other models, and other covariances, remain unaddressed. While the linear covariance allows for a floor on all other covariances, this may not be a very accurate bound. Computing this exactly for a robust family of covariances, say the Matérn family, may have value, especially if the results can be extended to nontrivial models with these covariances. Until this is done, it is difficult for a practitioner to know the results of increased sampling.

Finally, there is a good knowledge base of growth curve parameter estimation issues in an increasing domain (see [27] for many examples), including which parameters have better estimation behavior and alternate parameterizations for some models for improved estimation. This knowledge base does not exist for an IA domain. We believe that the approach used in this work could be applied across multiple models, with multiple covariance structures, with many differing parameter settings for both the model and the error terms, and the examples would collectively constitute at least a rough start to improving this knowledge base.

Appendix I: White's Inconsistent Estimator Proof Revisited

In [37], White considered the mean-only model with covariance $Cov(t_i, t_j) = \sigma^2\beta^h$, where $h = |t_i - t_j|$ and $\sigma^2 > 0$, $0 < \beta < 1$. Without loss of generality we may set $\sigma^2 = 1$, and using equally spaced samples the distance between consecutive points is uniformly $1/(n-1)$. We may then make the substition $\rho = exp(-\beta/(n-1))$, yielding the covariance matrix:

$$C = \begin{pmatrix} 1 & \rho & \rho^2 & ... & \rho^{n-1} & \rho^n \\ \rho & 1 & \rho & ... & \rho^{n-2} & \rho^{n-1} \\ \rho^2 & \rho & 1 & ... & \rho^{n-3} & \rho^{n-2} \\ ... & ... & ... & ... & ... & ... \\ \rho_{n-1} & \rho^{n-2} & \rho^{n-3} & ... & 1 & \rho \\ \rho^n & \rho^{n-1} & \rho^{n-2} & ... & \rho & 1 \\ \end{pmatrix}$$

The inverse of this is well-known:

$$C^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & ... & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & ... & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & ... & -\rho & 1 \\ \end{pmatrix}$$

As the model in question is a mean-only model (or equivalently a model in steady state), the Fisher Information is the sum of the elements of the inverse

covariance. The sum of the elements of C^{-1} is:

$$\begin{aligned}
\vec{1_n}^T C^{-1} \vec{1_n} &= \frac{1}{1-\rho^2}(2 + (n-2)(1+\rho^2) - 2(n-2)\rho) \\
&= \frac{1}{1-\rho^2}(2 + (n-2)(1+\rho^2 - 2\rho)) \\
&= \frac{1}{1-\rho^2}(2 + (n-2)(1-\rho)^2) \\
&= \frac{2 + (n-2)(1-\rho)}{1+\rho} \\
&= \frac{n(1-\rho) + 2\rho}{1+\rho}
\end{aligned}$$

So the lower bound on the variance of the unknown mean parameter is then:

$CRLB = \frac{1+\rho}{n(1-\rho)+2\rho}$

From here forward the remainder of the calculations closely mirror those in [37]. Using the substitution $\rho = exp(-\beta/(n-1))$, we can write

$CRLB = \frac{1+exp(-\beta/(n-1))}{n(1-exp(-\beta/(n-1)))+2exp(-\beta/(n-1))}$

As n increases, the numerator clearly approaches a value of $1 + exp(0) = 2$, and the second term in the denominator approaches $2exp(0) = 2$. The first term in the denominator requires a Taylor series expansion of the exponential to see the asymptotic behavior:

$$\begin{aligned}
n\left(1 - exp(-\beta/\left(n-1\right))\right) &= n\left(1 - \left(1 - \frac{\beta}{n-1} + \frac{1}{2!}\left(\frac{\beta}{n-1}\right)^2 - \frac{1}{3!}\left(\frac{\beta}{n-1}\right)^3 + ...\right)\right) \\
&= \frac{n\beta}{n-1} - \frac{n}{2!}\left(\frac{\beta}{n-1}\right)^2 + \frac{n}{3!}\left(\frac{\beta}{n-1}\right)^3 - ...
\end{aligned}$$

Every term of this after the first asymptotically vanishes; the first term approaches β as n increases. Putting these all together yields:

$lim_{n\to\infty} CRLB = \frac{2}{\beta+2}$

This bound was shown by White to be the variance of the MLE, so in this model the MLE is an efficient estimator. However, as this bound does not asymptotically tend to zero, there is no unbiased estimator with an asymptotically vanishing variance; then there is no consistent estimator for the mean parameter of this model.

Bibliography

[1] Abramowitz, Milton and Irene A. Stegun. *Handbook of Mathematical Functions*. Washington D.C.: National Bureau of Standards, 1972.

[2] Allgower, E. L. "Exact Inverses of Certain Band Matrices," *Numer. Math.*, *21*:279–284 (1973).

[3] Banerjee, Sudipto, et al. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, 2004.

[4] Bertalanffy, Ludwig von. "A Quantitative Theory of Organic Growth," *Human Biology*, *10*:181–213 (1938).

[5] Birch, Colin P. D. "A New Generalized Logistic Sigmoid Growth Equation Compared with the Richards Growth Equation," *Annals of Botany*, *83*:713–723 (1999).

[6] Cramér, Harald. *Methods of Mathematical Statitics*. Princeton, NJ: Princeton University Press, 1946.

[7] Cressie, Noel A. *Statistics for Spatial Data*. New York: Wiley, 1991.

[8] Di Battista, T., et al. "The Environmental Long-Memory Space-Time Series Prediction," *Environmental Informatics Archives*, *2*:289–296 (2004).

[9] Dow, Murray. "Explicit Inverses of Toeplitz and Associated Matrices," *ANZIAM Journal*, *44*:185–215 (2002).

[10] Du, Juan, et al. "Fixed Domain Asymptotic Properties of Maximum Likelihood Estimators," *Annals of Statistics (in press)* (2008).

[11] Duan, Jason A., et al. "Modeling Space-Time Data with Stochastic Differential Equations," *Working Paper* (2008).

[12] Furrer, R., et al. "Covariance Tapering for Interpolation of Large Spatial Datasets," *J. Comput. Graph. Stat.*, *15*(3):502–523 (2006).

[13] Garcia, Oscar. "Unifying Univariate Sigmoid Growth Functions," *Forest Biometry, Modelling and Information Sciences*, *1*:63–68 (2005).

[14] Gohberg, I. and A. A. Semencul. "On Inversion of Finite-Section Toeplitz Matrices and their Continuous Analogs," *Matem. Issled., Kishinev (Russian)*, *7*:201–224 (1972).

[15] Gompertz, Benjamin. "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies," *Philosophical Transactions of the Royal Society of London*, *115*:513–585 (1825).

[16] Gray, Robert M. *Toeplitz and Circulant Matrices: A Review*. Boston: Now Publishers Inc., 2005.

[17] Grenander, U. "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance," *Annals of Mathematical Statistics*, *25*:252–272 (1954).

[18] Haslett, John and Adrian E. Raftery. "Space-Time Modeling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource," *Applied Statistics*, *38*:1–50 (1989).

[19] Iohvidov, I. S. *Hankel and Toeplitz Matrices and Forms*. Boston, MA: Birkhäuser, 1982.

[20] Kaiser, Henry F. and Kern Dickman. "Sample and Population Score Matrices and Sample Correlation Matrices from an Arbitrary Population Correlation Matrix," *Psychometrika*, *27* (1962).

[21] Kimura, Daniel K. "Likelihood Methods for the von Bertalanffy Growth Curve," *Fishery Bulletin*, *77*:765–776 (1980).

[22] Lahiri, Soumendra Nath. "On Inconsistencies of Estimators Based on Spatial Data Under Infill Asymptotics," *Sankhya: The Indian Journal of Statistics*, *58*:403–417 (1996).

[23] Loh, J. M. and Michael L. Stein. "Spatial Bootstrap with Increasing Observations in a Fixed Domain," *Statistica Sinica*, *18*:667–688 (2008).

[24] Morris, M. D. and S. F. Ebey. "An Interesting Property of the Sample Mean Under a First-Order Auto-Regressive Model," *American Statistician*, *38*:127–129 (1984).

[25] Myers, Raymond H., et al. *Generalized Linear Models with Applications in Engineering and the Sciences*. Wiley, 2002.

[26] Pukelsheim, Friedrich. *Optimal Design of Experiments*. New York: Wiley, 1993.

[27] Ratkowsky, D. A. *Handbook of Nonlinear Regression Models*. New York: Marcel Dekker, 1990.

[28] Rehnqvist, Lars. "Inversion of Certain Symmetric Band Matrices," *Nordisk. Tidskr. Informationbehandslung (BIT)*, *12*:90–98 (1972).

[29] Richards, F. J. "Fixed Domain Asymptotic Properties of Maximum Likelihood Estimators," *Journals of Experimental Botany*, *10*:290–300 (1959).

[30] Rodman, Leiba and Tamir Shalom. "On Inversion of Symmetric Toeplitz Matrices," *SIAM Journal on Matrix Analysis and Applications*, *13*:530–549 (1992).

[31] Stein, Michael L. "Asymptotically Efficient Prediction of a Random Field with a Misspecified Covariance Function," *Annals of Statistics*, *16*:55–63 (1988).

[32] Stein, Michael L. "Fixed-Domain Asymptotics for Spatial Periodograms," *Journal of the American Statistical Association*, *90*:1277–1278 (1995).

[33] Stein, Michael L. "Space-Time Covariance Functions," *Journal of the American Statistical Association*, *100*:310–321 (2005).

[34] Trench, William F. "Explicit Inversion Formulas for Toeplitz Band Matrices," *SIAM J. Alg. Disc. Meth.*, *6*:546–554 (1985).

[35] Vonesh, Edward F. and Vernon M. Chinchilli. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: CRC Press, 1996.

[36] White, Edward Dalton III. *A PC Program for the Parameter Estimation of Nonlinear Growth Models with Unequally Spaced Correlated Observations*. PhD dissertation, Texas A&M University, College Station, TX, 1998.

[37] White, Edward Dalton III. "Investigating Estimator Consistency of Nonlinear Growth Curve Parameters," *Communications In Statistics - Simulation and Computation*, *30*:1–10 (March 2001).

[38] Winsor, Charles P. "The Gompertz Curve as a Growth Curve," *Proceedings of the National Academy of Sciences*, *1*:1–8 (1932).

[39] Yin, Xinyou, et al. "A Flexible Sigmoid Function of Determinate Growth," *Annals of Botany*, *91*:361–371 (2003).

[40] Zhang, Hao. "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," *Journal of the American Statistical Association*, *99*:250–261 (March 2004).

[41] Zhang, Hao and Dale L. Zimmerman. "Towards Reconciling Two Asymptotic Domains in Spatial Statistics," *Biometrika*, *92*:921–936 (January 2005).

[42] Zhu, Zhungyuan and Hao Zhang. "Spatial Sampling Design Under the Infill Asymptotic Framework," *Environmetrics*, *17*:323–337 (2006).

[43] Zohar, Shalhav. "Toeplitz Matrix Inversion: The Algorithm of W. F. Trench," *Journal of the ACM*, *16*:592–601 (October 1969).

Vita

Major David T. Mills was born in San Antonio, Texas. A homeschooled student, he earned a GED in 1992, attended Palo Alto Community College in San Antonio, Texas, and graduated from Southwest Texas State University in 1997 (BS, Mathematics, Magna Cum Laude), and the Air Force Institute of Technology (MS, Operations Research 1999). He was commissioned into the US Air Force in 1997, and has been stationed in Ohio, Maryland, New Mexico, and Colorado. He married in 2000, and is the father of three children. He has served on the faculty of the US Air Force Academy.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
17-09-2010	Doctoral Dissertation	Sep 2007 -- Sep 2010

4. TITLE AND SUBTITLE

Consistency Properties for Growth Model Parameters Under an Infill Asymptotics Domain

5a. CONTRACT NUMBER

5b. GRANT NUMBER
None

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Mills, David T, Maj USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/DAM/ENC/10-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Intentionally left blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Growth curves are used to model various processes, and are often seen in biological and agricultural studies. Underlying assumptions of many studies are that the process may be sampled forever, and that samples are statistically independent. We instead consider the case where sampling occurs in a finite domain, so that increased sampling forces samples closer together, and also assume a distance-based covariance function. We first prove that, under certain conditions, the mean parameter of a fixed-mean model cannot be estimated within a finite domain. We then numerically consider more complex growth curves, examining sample sizes, sample spacing, and quality of parameter estimates, and close with recommendations to practitioners.

15. SUBJECT TERMS
Infill Asymptotics; Toeplitz Matrices

16. SECURITY CLASSIFICATION OF: U			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
REPORT	**ABSTRACT**	**c. THIS PAGE**			Edward D. White, AFIT/ENC
U	U	U	UU	92	**19b. TELEPHONE NUMBER** *(Include area code)* 937-255-6565 x 4540

Standard Form 298 (Rev: 8-98)
Prescribed by ANSI Std. Z39-18