ABSTRACT

DAVIS, JIMENA LAMANDA. Uncertainty Quantification in the Estimation of Probability Distributions on Parameters in Size-Structured Population Models. (Under the direction of Professor H.T. Banks.)

We consider ordinary least squares (OLS) parameter estimation problems in which the underlying dynamics are described by partial differential equations and the unknown parameter of interest is a probability distribution describing the variability of growth rates across a size-structured population. The focus of our work is the development of an inverse problem computational methodology for the estimation of functional parameters in the presence of (model and data) uncertainty. Since this optimization problem involves both an infinite-dimensional state space and an infinite-dimensional parameter space, computationally efficient approximation methods for both parametric and non-parametric versions of the OLS inverse problem are developed and discussed. The approximation methods that we present are applicable to a variety of inverse problems, including *Type II* problems in which only aggregate or population level longitudinal data is available.

We compare computational and statistical results of a delta function approximation method, a spline based approximation method, and a standard parametric OLS formulation. The latter uses an a priori probability distribution in the inverse problem for estimation of distributions of growth rates in size-structured marine populations. After summarizing the underlying theoretical framework, we present several numerical examples as validation of the theory. Convergence as well as sensitivity of the estimates with respect to noise in the data is discussed for both approximation methods.

A computational framework for quantification of uncertainty associated with the estimated parameters is given and sample numerical findings are presented. We demonstrate how to construct "functional" confidence bands that will aid in quantifying the uncertainty in estimated probability distributions by extending the standard asymptotic theory for finite-dimensional OLS estimators. Using our inverse problem methodology, we present results for the estimation of growth rate distributions in size-structured marine populations illustrating the strengths and weaknesses associated with the three different computational schemes.

Uncertainty Quantification in the Estimation of Probability Distributions on Parameters in Size-Structured Population Models

by
Jimena Lamanda Davis

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fullfillment of the
requirements for the Degree of
Doctor of Philosophy

Computational Mathematics

Raleigh, North Carolina

July 2008

APPROVED BY:

Dr. Alun L. Lloyd	Dr. Hien T. Tran
Dr. H.T. Banks Chair of Advisory Committee	Dr. Marie Davidian

DEDICATION

For my three "Angels" whom I love and miss so dearly,
I truly wish you were all here to share in this with me.

My Grandmother Lula Mae Owens
You will always be my heart and my motivation. Everything I have accomplished and will ever accomplish is a tribute to you. The promises I made to you kept me going when I wanted to give up. I pray that I have made you proud of me!

My Aunt Lois Ann Owens
You always saw the best in me. I am so grateful for the special way you had of making me smile in some of my most difficult times. Even though you are not here physically I know you are watching over me and still cheering for me!

My Best Friend LaTina Renee Wilson
You are truly my inspiration, and I am a stronger individual because of you. I hope you know how valuable and irreplaceable your friendship was and still is to me. You will always hold a special place in my heart!

BIOGRAPHY

Jimena Lamanda Davis was born on November 3, 1980 to Sandra and Clifton Davis, Jr. She has one younger sister Sherena. Jimena grew up in the small town of Mullins, South Carolina. She graduated from Mullins High School as the Valedictorian of the class of 1999.

Jimena then attended Clemson University as a Palmetto Fellows Scholar. Her major was initially computer engineering, but she soon realized that math was her true passion. Jimena graduated Summa Cum Laude in May 2003 with a Bachelor of Science in Mathematical Sciences and a minor in Computer Science.

After completing her undergraduate studies, Jimena moved to Raleigh, North Carolina to continue her education. She was awarded a SAMSI Graduate Fellowship for her first year of graduate school at North Carolina State University. Jimena was then chosen as a recipient of a Department of Energy Computational Science Graduate Fellowship in 2004. She received her Master of Science in Applied Mathematics in December 2005. During the summer of 2006, Jimena traveled to Albuquerque, New Mexico for a summer internship at Sandia National Laboratories. She will have earned a Doctor of Philosophy in Computational and Applied Mathematics with the completion of this dissertation.

Jimena has accepted a post-doctoral fellowship position in the National Center for Computational Toxicology in the United States Environmental Protection Agency in Research Triangle Park, North Carolina.

ACKNOWLEDGMENTS

Trust in the Lord with all thine heart; and lean not unto thine own understanding. In all thy ways acknowledge him, and he shall direct thy paths. - Proverbs 3:5-6

First, I give all honor and praise to God for without Him I would not have completed this task. I am most grateful for His grace, His mercy, and His favor, and God continues to get all glory for the great and marvelous things He has done for me!

This has been a long journey filled with ups and downs, and there are so many people who have had an impact on me during this process. I would like to thank my advisor Dr. H.T. Banks for being a wonderful teacher and mentor to me during my time at North Carolina State University. I have truly learned so much both inside and outside of the classroom from you, and I have grown as an individual and a mathematician because of your guidance and support. I would also like to thank Dr. Banks and Mrs. Susie Banks for welcoming me into their home on so many occasions.

There are many individuals at NC State who have been very supportive of me. I want to express my gratitude to my committee members, Dr. Marie Davidian, Dr. Alun Lloyd, and Dr. Hien Tran. Thank you for being so willingly to serve on my committee. I appreciate your time and contributions to my work and professional development. I would also like to thank Dr. Stephen Campbell and Dr. Ernie Stitzinger for recruiting me to the mathematics graduate program and taking an interest in my growth here at NC State. I am also grateful to Dr. Robert Martin for being very supportive of me. I am especially appreciative of Denise Seabrooks for her encouragement and optimism throughout this process. You were an immense help and a bright spot every time I was in your presence.

Before studying at NC State, I was given a solid foundation as a result of the wonderful mathematical sciences department at Clemson University. I must first thank Dr. Terri Johnson for opening my eyes to the field of mathematics. As a result of many discussions with you, I chose to change my major to mathematical sciences, and I am appreciative of your continued encouragement and support. I would also like to express my gratitude to Dr. Joel Brawley, Dr. Robert Fennell, and Dr. Kevin James for all of your support. Each of you played a significant role in my development as a mathematician at Clemson, and I am truly thankful that I had the opportunity to work with you.

During my time at NC State, I was blessed with the opportunity to travel to Albuquerque, New Mexico for a summer internship at Sandia National Laboratories. I was

even more blessed by the mentoring of Dr. Elebeoba "Chi-Chi" May. Thank you so much for expressing an interest in working with me. This experience not only helped me to grow professionally but also personally, and I am so very grateful that you have continued to serve as a mentor to me. I would also like to give a special thanks to Dr. Elena Artimovich and Dr. Arun Dhar of Advanced Bionutrition Corporation and Dr. Craig Browdy of Marine Resources Research Institute for their collaborative efforts and for providing experimental data.

I have had the support of a wonderful family and great friends throughout my life. I am so grateful for the love and support of my immediate family, my parents Clifton and Sandra and my sister Sherena. I really hope that I have made you proud of me. Thank you Mom for your never ending prayers and for always encouraging me when I was low in spirits. I am so very appreciative of you Daddy for always believing that I could do anything even when I wasn't sure and for being willing to come to my rescue. I would like to give a special thanks to Sherena for providing me with many laughs and for encouraging me to stay strong when others tried to bring me down.

My grandparents, Robert and Lula Mae Owens and Clifton and Rosephine Davis, were also very supportive of me, and although they did not get to see this in person, I pray you are celebrating with me in spirit. I love each of you and hope that I have become a woman that you would all be proud of today.

Although my immediate family is small, I have been blessed with numerous "extended" family members who are also very dear to me. I would like to acknowledge the support and advice of my spiritual "mother" Carolyn Wilson. You have truly been a blessing to my entire family, and I am so thankful for everything you have done for us. I would also like to recognize my two little "sisters" Ashley and Chelsea Godbold. You both mean so much to me, and I have enjoyed watching you both grow into beautiful young ladies. I am so very proud of you!

I would like to thank my second "mom" and kindergarten teacher Sharon Hollen. You have been so supportive of me throughout my life, and I will always be grateful to you for your love and encouragement.

I want to express my sincere gratitude to my friend and "sister" of many years Lanita Coleman for being there for me through so much. Thank you for answering the early morning and late night calls and text messages about everything. I want to thank my "brother" Johnnie Smith for not putting me completely in the "red" when I failed to keep

in contact. I truly appreciate my friend Shinique McNeil– along with Nita, you have helped me to relax and enjoy the moment whenever I managed to take a break from work. I would also like to thank my friend Edwin Smith for the many conversations about everything and nothing.

While at Clemson University, I met two wonderful young ladies, Christie Dyson and Jametha Washington, who have been so supportive of me and my goals. Thank you for reminding (and sometimes "forcing") me to take time out to take care of me. You have both been so understanding when I couldn't be there physically because of school. I truly appreciate the friendships that I have developed with you.

I have also been blessed to meet some great individuals at NC State who have not only helped in my studies but who have also proven to be great friends. I would like to extend my thanks to those I have had the pleasure of sharing an office with– Shuhua Hu and Johnny Samuels. I have learned a great deal from both of you and have really enjoyed working with you and getting to know you. Thanks to April Alston for your friendship and for the countless conversations both in person and over IM. When we met at the recruitment weekend, I had no idea we would become such good friends, but I am so glad that our paths crossed again. I would like to give a sincere thanks to Karen Dillard– we both endured a long summer while preparing for the PDEs qualifier but there is no one else I would have preferred to go through it with than you. I also want to thank you for being such a thoughtful and encouraging friend through everything. Thanks so much to Kelly Dickson for being a supportive friend not only while we were in Raleigh but also in Albuquerque. I would like to acknowledge Martene Fair, Gloria Porter and Monique Taylor for celebrating my triumphs and sharing in this experience with me. I am grateful for the support that each of you have given me. I would also like to thank my friend Johnsie Gunter for listening to all of my stories and for offering me your advice and support when I needed it most.

There are so many individuals that have offered me encouraging words, prayers, and support throughout this process, and while I cannot list them all individually here, I would like to thank each of you for your well wishes. There are also those who tried to distract me and discourage me from achieving this goal. I thank you as well because your negativity served as motivation and I persevered in spite of your ill intentions.

My graduate studies, including this research, have been supported in part by the NC State University Mathematics Department and by the National Science Foundation under grant DMS-0112069 to the Statistical and Applied Mathematical Sciences Institute

(SAMSI). This dissertation work was also supported in part by the US Department of Energy Computational Science Graduate Fellowship under grant DE-FG02-97ER25308. Many thanks to the Krell Institute and the Department of Energy for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The importance of estimating time and spatially dependent function parameters as coefficients in distributed parameter models has been recognized for some time [22]. This is especially true when one is trying to determine mechanistic-based terms in a model. General theoretical and computational ideas (called *function space estimation convergence* or *FSPEC* in [22]) for approximation schemes for such problems were developed some years ago and now are used somewhat routinely by practitioners. A diverse range of examples involving systems of the form

$$\frac{\partial u}{\partial t} + V \cdot \nabla u = \nabla \cdot (D \nabla u) - \mu u$$

for the state variables $u = u(t, x)$ is discussed in Chapter 7 of [22], where parameters to be estimated are generally vector *functions* of the form $q = (D, V, \mu)$ and are to be chosen from some set Q of admissible parameter functions. As summarized in [22], spatially dependent coefficients $D = D(x)$ are used in [26] to study the effects of bioturbation on volcanic ash records in core samples from deep sea sediments. Functional coefficients are also needed in the insect dispersal studies of [19, 21] where vegetation effects on dispersal lead to spatially dependent advection $V = V(x)$ and time-dependent emigration/immigration $\mu = \mu(t)$ terms are important in capture-mark-release flea beetle experiments (these are used to characterize "initial disturbance" effects due to the trauma from capture, handling, etc.). Similar studies involving time-dependent anemotaxis ($V = V(t)$) and emigation/immigration ($\mu = \mu(t)$) in cabbage root fly dispersal [44] are described in [20].

In these problems one uses data $\{y_k\}$ for the parameter dependent model values $u(\tau_k; q)$ (where typically $\tau_k = (t_i, x_j)$ are time and/or spatial covariates) to estimate

functions $q \in Q$. The data $\{y_k\}$ can be regarded as a realization of the observation process

$$Y_k = u(\tau_k; q_0) + \epsilon_k, \quad k = 1, \ldots, n,$$

where the ϵ_k are measurement or observation errors and q_0 are underlying "true" parameters (assumed to exist in theoretical formulations). This leads to estimates $\hat{q}$ defined by

$$\hat{q} = \arg\min_{q \in Q} \sum_{k=1}^{n} [u(\tau_k; q) - y_k]^2$$

and corresponding ordinary least squares (OLS) estimator

$$q_{OLS}(Y) = \arg\min_{q \in Q} \sum_{k=1}^{n} [u(\tau_k; q) - Y_k]^2,$$

which is a Q-space valued random variable. The distribution of this infinite-dimensional random variable is called the *sampling distribution* and is a probability distribution on Q. It is of great interest since knowledge of this will lead to information about the uncertainty associated with the estimates $\hat{q}$. In finite-dimensional problems, there is a rather complete asymptotic theory to provide such results (see Chapter 12 of [60]). The major focus of this dissertation is the development of an infinite-dimensional analogue.

Another class of problems to which such an infinite-dimensional theory would be immediately applicable is that involving estimation of parameters in the Fokker-Planck or forward Kolmogorov equation [1, 41] for transition probabilities $p(s, y; t, x)$ for the stochastic diffusion process $X(t)$ for a growth process

$$\frac{\partial p}{\partial t} + \frac{\partial [a(t,x)p]}{\partial x} = \frac{1}{2} \frac{\partial^2 [b(t,x)p]}{\partial x^2}.$$

Here $a(t, x)$, the "drift" or mean growth rate, and $b(t, x)$, the "diffusion" or second moment of the rate of increase, are the functional parameters $q = (a, b)$ to be estimated. Because the population density $u(t, x)$, where growth is assumed to be a stochastic diffusion process, also satisfies such an equation (see [53]), this model can be used as a stochastic alternative (e.g., see [27]) to the Sinko-Streifer deterministic growth model [5, 46]

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(g(t,x)v) = -\mu(t,x)v.$$

The estimation of time-dependent mortalities in these equations is important in recent problems for sublethal effects of pesticides [2, 3] in insect populations where constant parameters μ prove inadequate in describing population life data.

In this dissertation we consider another class of estimation problems where the functions to be estimated are actually probability distributions or densities. This class of problems arises when one assumes that a probability distribution describes the distribution of growth rates g in the Sinko-Streifer deterministic growth model. Such formulations are called *Growth Rate Distribution (GRD)* models [7, 8, 13, 14]. We consider approximation methods for this general class of estimation or inverse problems wherein the quantity of interest is a probability distribution. In particular we assume that we have a parameter ($\omega \in \Omega$) dependent system with model responses $x(\omega)$ describing in some manner a population of interest. For data or observations, we are given a set of values $\{z_l\}$ for the expected values

$$\mathcal{E}[x_l(\omega)|P] = \int_\Omega x_l(\omega) dP(\omega)$$

with respect to an unknown probability distribution P describing the distribution of the parameters ω over the population under investigation. We wish to use this data to choose from a given family $\mathcal{P}(\Omega)$ the distribution P^* that gives a best fit of the underlying model to the data. Here we formulate an ordinary least squares (OLS) version of the problem, but this is not essential to our results and one could equally well use a weighted least squares, a maximum likelihood estimator, etc., approach. Thus we seek to minimize

$$J(P) = \sum_l |\mathcal{E}[x_l(\omega)|P] - z_l|^2$$

over $P \in \mathcal{P}(\Omega)$.

Even for simple dynamics for x_l, this is in general an infinite-dimensional optimization problem and approximations that lead to computationally tractable schemes are desirable. That is, it is useful to formulate methods to yield finite-dimensional sets $\mathcal{P}^M(\Omega)$ over which to minimize $J(P)$. Of course, we wish to choose these methods so that "$\mathcal{P}^M(\Omega) \to \mathcal{P}(\Omega)$" in some sense. In this case we shall use the Prohorov metric [6, 29] of weak star convergence of measures to assure the desired approximation results. A general theoretical framework is given in [6] with specific results on the approximations we use here given in [4, 24]. Briefly, ideas for the underlying theory are as follows. One argues continuity of $P \to J(P)$ on $\mathcal{P}(\Omega)$ with the Prohorov metric (this is trivial in the cases considered here). If Ω is compact (again, easily established in our case) then it is known that $\mathcal{P}(\Omega)$ is a complete metric space and is also compact when taken with the Prohorov metric. Moreover, if the approximation families $\mathcal{P}^M(\Omega)$ are chosen so that elements $P^M \in \mathcal{P}^M(\Omega)$ can

be found to approximate elements $P \in \mathcal{P}(\Omega)$ in the Prohorov metric, then well-posedness (existence, continuous dependence of estimates on data, etc.) can be obtained.

The data $\{z_l\}$ available (which, in general, will involve longitudinal or time evolution data) determines the nature of the problem. The most classical problem (which we shall refer to as a *Type I* problem) is one in which individual longitudinal data is available for members in the population. In this case there is a wide statistical literature (in the context of hierarchical modeling, mixing distributions, mixed or random effects, mixture models, etc.) [28, 34, 35, 36, 37, 48, 49, 50, 52, 57, 58, 61] which provides theory and methodology for estimating not only individual parameters but also population level parameters and allows one to investigate both intra-individual and inter-individual variability in the population and data. In what we shall refer to as *Type II* problems one has only *aggregate* or population level longitudinal data available. This is common in marine, insect, etc., *catch and release* experiments [22] where one samples at different times from the same population but cannot be guaranteed of observing the same set of individuals at each sample time. This type of data is also typical in experiments where the organism or population member being studied is sacrificed in the process of making a single observation (e.g., certain physiologically based pharmacokinetic (PBPK) modeling [25, 39, 55] and whole organism transport model [22]). In this case one may still have dynamic (i.e., time course) models for individuals, but no individual data is available. Finally, the third class of problems which we shall refer to as *Type III* problems involves dynamics which depend explicitly on the probability distribution P itself. In this case one only has dynamics (*aggregate dynamics*) for the expected value

$$\bar{x}(t) = \int_\Omega x(t;\omega) dP(\omega)$$

of the state variable. No dynamics are available for individual trajectories $x(t;\omega)$ for a given $\omega \in \Omega$. Such problems arise in viscoelasticity and electromagnetics as well as biology [6, 15, 16, 24, 42]. While the approximation methods we discuss in this dissertation are applicable to all three types of problems, we shall illustrate the computational results in the context of size-structured marine populations where the inverse problems are of *Type II.*

Finally, we note that in the problems considered here, one cannot sample directly from the probability distribution being estimated and this again is somewhat different from the usual case treated in some of the statistical literature, e.g., see [63, 64] and the references cited therein.

In Chapter 2 we discuss the Growth Rate Distribution model, a generalization

of the Sinko-Streifer model for size-structured populations. We also outline the standard parametric approach as well as two computationally efficient approximation schemes (one based on delta functions and one based on piecewise linear splines) for the Growth Rate Distribution model in both forward and inverse problem calculations.

We present formulations for the standard parametric inverse problem as well as the non-parametric inverse problems in Chapter 3. As explained in detail in Chapters 2 and 3, the original estimation problem for the infinite-dimensional parameter (probability distribution) is reduced to a finite-dimensional problem via the approximation methods studied here. Confidence intervals based on standard errors can be used to quantify the uncertainty associated with estimates of the finite-dimensional parameters from each method. Thus, we also give a summary of the finite-dimensional asymptotic distribution theory for which we seek a function space analogue.

In order to gain information on the capabilities of the approximation methods, we test the methods first with simulated data before applying the schemes to experimental data. We compare the approximation methods based on the numerical results from our computations with simulated data in Chapter 4 with respect to convergence and sensitivity of the estimates obtained as well as stability of the inverse problem. We then summarize results from both the parametric and non-parametric versions of the estimation problem for the growth rate distribution of a size-structured mosquitofish population.

In Chapter 5 we discuss an extension of the asymptotic standard error theory for finite-dimensional OLS estimators to functional parameters. We explain how to construct functional confidence bands based on the confidence intervals we obtain for the finite-dimensional parameters used in approximating the parameter of interest (probability distribution) in the class of estimation problems that we consider in this dissertation. Computational results with simulated data are presented to demonstrate how one can use these confidence bands to quantify the uncertainty in estimated probability distributions.

We apply the inverse problem computational methodology that we have developed to experimental data for a size-structured shrimp population cultivated in marine raceways in Chapter 6. We outline a computational approach to aid in the design of experiments to validate the Growth Rate Distribution model. Parameter estimation results for the growth rate distribution given experimental data collected based on our design analysis demonstrate the usefulness of our methodology in understanding the early growth dynamics of shrimp. Finally, we offer some concluding remarks and future directions for this work in Chapter 7.

Chapter 2

Mathematical Model for Size-Structured Populations

2.1 Introduction

In this chapter we present the Growth Rate Distribution model, a modification of the Sinko-Streifer model, which we use to describe size-structured mosquitofish populations. Mosquitofish have been used in the place of pesticides as a way to control mosquito populations in rice fields by eating the water-borne mosquito larvae [30]. Biologists desire to correctly predict the growth and decline of the mosquitofish population in order to determine the optimal densities of mosquitofish to use in rice fields to control mosquito populations. Thus, a mathematical model that accurately describes the mosquitofish population would be beneficial in this application as well as in other problems involving population dynamics and age/size-structured population data.

Based on data collected from rice fields, a reasonable mathematical model would have to predict two key features characteristic of mosquitofish population data: dispersion and bifurcation of the population density over time [7, 13, 14]. The data shown in Figure 2.1 exhibits these two features; that is, as time progresses, the population density spreads out (dispersion) and the unimodal density becomes a bimodal density (bifurcation).

There have been various types of mathematical models developed in the literature to model population dynamics [46]. One such model, the Sinko-Streifer model, groups individuals together based on shared characteristics or behaviors. The Sinko-Streifer [62]

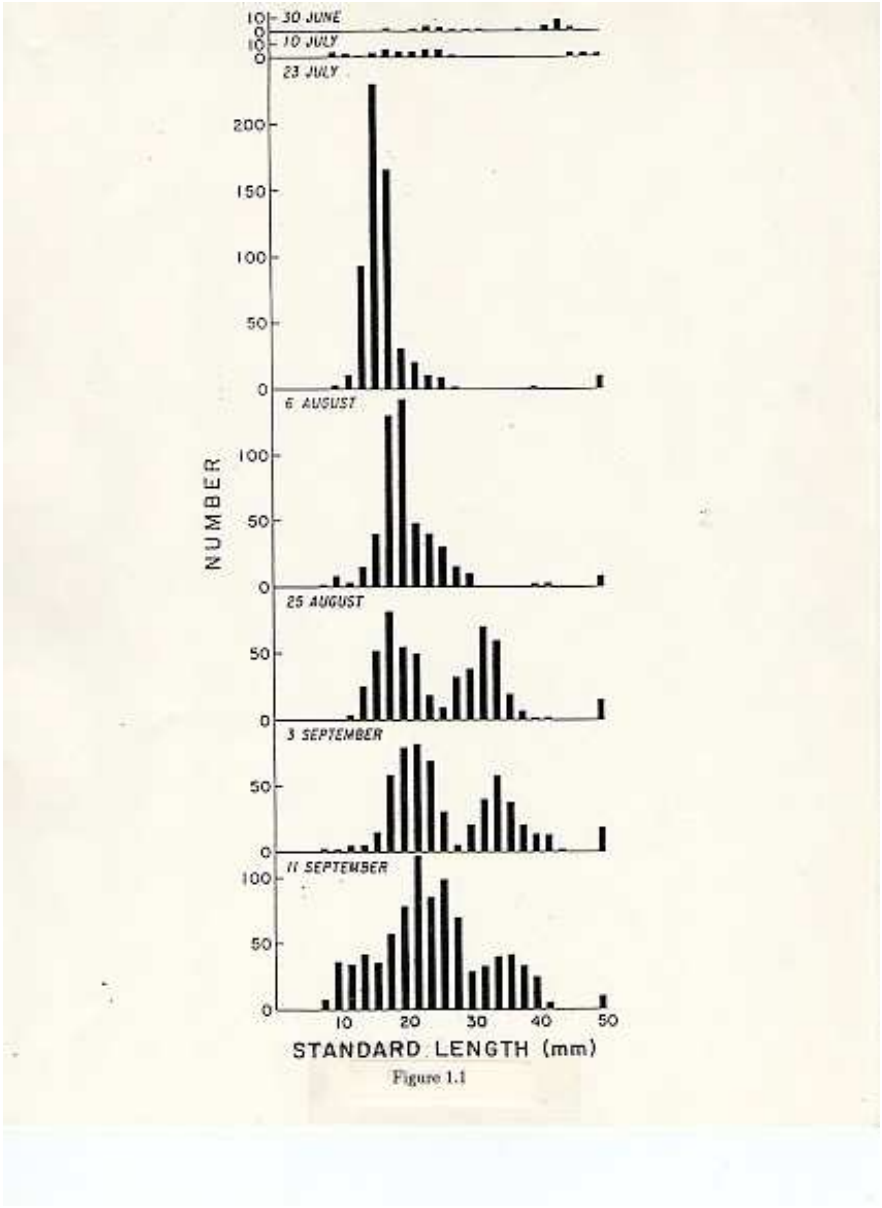

Figure 2.1: Example of mosquitofish population data collected in rice fields from [7].

model has been used widely to describe various age and size-structured populations [2, 3, 46]. However, it has been shown that the Sinko-Streifer model is not rich enough to predict features typically observed in mosquitofish population data (see Figure 2.1). It was suggested in [7] that these features are a result of variability in the growth rates of the mosquitofish. Based on these ideas, the Growth Rate Distribution model was proposed in [7] and developed in [13] to account for this variability by imposing a probability distribution on the growth rates in the Sinko-Streifer model. We discuss the Growth Rate Distribution model in the next section. In Section 2.3 we present computationally efficient approximation methods for the Growth Rate Distribution model that can be used in the estimation of probability distributions on parameters in size-structured population models.

2.2 Growth Rate Distribution Model: Modified Sinko-Streifer Model

Before discussing the Growth Rate Distribution (GRD) model, which was first introduced in [7] and developed in [13], we describe one version of the Sinko-Streifer (SS) model. The Sinko-Streifer model (also known as the McKendrick-von Foerster model) is used for modeling age and size-structured populations because it takes into account that individuals in a population may have different characteristics or behaviors. The Sinko-Streifer model for size-structured mosquitofish populations is given by the following partial differential equation

$$\begin{aligned} v_t(t,x;g) + (g(t,x)v(t,x;g))_x &= -\mu(t,x)v(t,x;g), \quad \underline{x} < x < \bar{x}, \quad t > 0 \\ v(0,x;g) &= v_0(x;g) \\ R(t) = g(t,\underline{x})v(t,\underline{x};g) &= \int_{\underline{x}}^{\bar{x}} K(t,\xi)v(t,\xi;g)d\xi. \end{aligned} \tag{2.1}$$

We note that $v(t,x;g)$ represents the size or population density (with units of number per size class) of individuals with growth rate g at time t with size (length) x. The number of mosquitofish in the population at time t with sizes between $\underline{x}$ and $\bar{x}$ is

$$N_{SS}(t) = \int_{\underline{x}}^{\bar{x}} v(t,x;g)dx.$$

The growth rate $g(t,x)$ of individual mosquitofish in the population is given by the deterministic growth model

$$\frac{dx}{dt} = g(t,x). \tag{2.2}$$

The mortality rate of the mosquitofish is given by $\mu(t,x)$, while $K(t,x)$ represents the fecundity kernel. The initial condition at $t = 0$ is represented by the size density function $v_0(x;g)$. The boundary condition $R(t)$ at $x = \underline{x}$, which is in terms of $K(t,x)$, represents the recruitment or birth rate. At $x = \bar{x}$ the additional physical boundary condition

$$g(t,\bar{x}) = 0$$

ensures the maximum size of the mosquitofish is $\bar{x}$.

All individual mosquitofish of the same size x at the same time t are assumed to have the same growth rate $g(t,x)$ in the SS model. However, with this assumption, solutions to (2.1) do not exhibit the dispersion and bifurcation of the mosquitofish population density observed in the data as noted earlier. The Sinko-Streifer model cannot be used as is to model the mosquitofish population because it does not predict dispersion or bifurcation of the population in time under biologically reasonable assumptions [7]. To predict the features of dispersion and bifurcation typical of the mosquitofish population, the SS model was modified so that the individual growth rates of the mosquitofish vary across the population and individuals with the same growth rate are assumed to be in the same subpopulation [7, 13, 14]. This assumption is more reasonable biologically based on further analysis of the growth dynamics of the mosquitofish. Male mosquitofish typically grow to a maximum size of 30mm, while female mosquitofish usually grow to a maximum size of 60mm. Therefore, it is reasonable to assume that male and female mosquitofish grow at different rates with their individual growth rates depending on additional parameters. The Growth Rate Distribution model assumes that the population is actually made up of several subpopulations, where individuals in the same subpopulation have the same growth rate and hence a distribution of growth rates across the population. The population density $u(t,x;\mathcal{P})$ in the GRD model is given by

$$u(t,x;\mathcal{P}) = \int_{g\in\mathcal{G}} v(t,x;g)d\mathcal{P}(g), \tag{2.3}$$

where $v(t,x;g)$ is the solution to (2.1) with growth rate $g(t,x)$, $\mathcal{G}$ is the collection of admissible growth rates, and $\mathcal{P}$ is a probability measure on $\mathcal{G}$. In this model the number of mosquitofish in the population between sizes $\underline{x}$ and $\bar{x}$ at time t is given by

$$N_{GRD}(t) = \int_{\underline{x}}^{\bar{x}} u(t,x;\mathcal{P})dx.$$

By modifying the SS model so that the growth rates of the mosquitofish vary among the individuals in the population (instead of being the same for all mosquitofish in the popu-

lation), the GRD model is obtained and has been shown to exhibit both dispersal in time of the mosquitofish population and the development of a bimodal density from a unimodal density [7, 8, 13, 14].

For our considerations, we assume that the admissible growth rates are size-dependent only and have the form

$$g(x;b,\gamma) = \begin{cases} b(\gamma - x), & \underline{x} \leq x \leq \gamma, \\ 0, & \text{otherwise}, \end{cases}$$

where b represents the intrinsic growth rate and $\gamma = \bar{x}$ represents the maximum size of the mosquitofish. This choice is based on work in [7], where the idea of other properties related to the growth rates varying among the mosquitofish is discussed. Under the assumption of varying intrinsic growth rates and maximum sizes, we in fact assume that b and γ are both random variables taking values in the compact sets $\mathfrak{B}$ and Γ, respectively. A reasonable assumption is that both are bounded closed intervals. The collection of admissible growth rates is then characterized as

$$\mathcal{G} = \{g(x;b,\gamma) | b \in \mathfrak{B}, \gamma \in \Gamma\},$$

where both $\mathfrak{B}$ and Γ are bounded closed intervals (i.e., compact sets in $\mathbb{R}$). Thus, $\mathcal{G}$ is also compact in, for example, $C[\underline{x}, X]$, where $X = \max(\Gamma)$. Then $\mathcal{P}(\mathcal{G})$ is compact in the Prohorov metric and we are in the framework outlined in Chapter 1. In the remainder of this chapter as well as the following outline of the approximation methods in Section 2.3 and inverse problem formulations in Chapter 3, we chose a growth rate parameterized by the intrinsic growth rate b with $\gamma = 1$, leading to a one parameter family of varying growth rates g among the individuals in the population. However, these ideas easily extend to the two parameter family as well. We also assume here that $\mu(t,x) = 0$ and $K(t,x) = 0$ in order to focus on only the distribution of growth rates; however, distributions could just as well be placed on μ and K.

Before discussing the approximation methods, we derive the initial condition driven solution to the SS model. Since the functions $g(t,x), \mu(t,x)$, and $K(t,x)$ are known explicitly, we can determine the solution via the method of characteristics [38]. Figure 2.2 depicts the (t,x) plane separated into two regions by the curve $G(x) = T(x;0;\underline{x}) = X(t;0,\underline{x})$ passing through $(0,\underline{x})$, where $X(t;\hat{t},\hat{x})$ is the function mapping $t \to x$ and $T(x;\hat{t},\hat{x})$ is its inverse mapping $x \to t$. The characteristic curve passing through $(\hat{t},\hat{x})$ in the (t,x) plane is denoted

by $(t, X(t;\hat{t},\hat{x}))$ and satisfies

$$\frac{d}{dt}X(t;\hat{t},\hat{x}) = g(t, X(t;\hat{t},\hat{x}))$$
$$X(\hat{t};\hat{t},\hat{x}) = \hat{x}.$$

The solution to the SS model is valid only along the characteristic curve and is given by the initial condition driven solution for $t \leq G(x)$ and by the recruitment driven solution for $t > G(x)$. The initial condition driven solution for $t \leq G(x)$ is given by

$$v(t,x;g) = v_0(X(0;t,x);g)\exp\left\{-\int_0^t [g_x(\xi, X(\xi;t,x)) + \mu(\xi, X(\xi;t,x))]d\xi\right\}, \tag{2.4}$$

and the recruitment driven solution for $t > G(x)$ is given by

$$v(t,x;g) = \frac{R(T(\underline{x};t,x))}{g(T(\underline{x};t,x),\underline{x})}\exp\left\{-\int_{T(\underline{x};t,x)}^t [g_x(\xi, X(\xi;t,x)) + \mu(\xi, X(\xi;t,x))]d\xi\right\}. \tag{2.5}$$

Since we assume there are no births into the population ($K(t,x) = 0$), then $R(t) = 0$ for all t and the solution to the SS model in the region $t > G(x)$ is given by $v(t,x;g) = 0$.

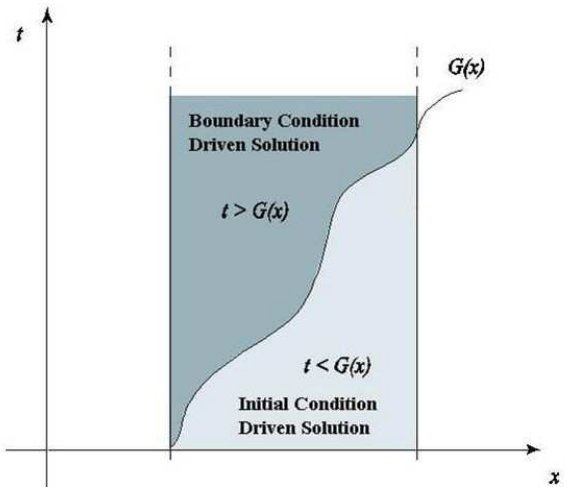

Figure 2.2: Regions in the (t,x) plane defining the solution to the Sinko-Streifer model.

We first determine the equation for the characteristic curve by solving

$$\frac{d}{dt}X(t;\hat{t},\hat{x}) = b(1 - X(t;\hat{t},\hat{x}))$$
$$X(\hat{t};\hat{t},\hat{x}) = \hat{x}.$$

The solution is given by

$$X(t;\hat{t},\hat{x}) = 1 - (1-\hat{x})\exp\{b(\hat{t}-t)\}, \tag{2.6}$$

and its inverse is given by

$$T(x;\hat{t},\hat{x}) = \hat{t} - \frac{1}{b}\ln\left(\frac{1-x}{1-\hat{x}}\right). \tag{2.7}$$

Therefore, $G(x) = T(x;0,\underline{x})$ is given by

$$G(x) = -\frac{1}{b}\ln\left(\frac{1-x}{1-\underline{x}}\right).$$

Substituting the appropriate parameter values in (2.4), we obtain the initial condition driven solution

$$\begin{aligned} v(t,x;g) &= v_0(X(0;t,x);g)\exp\left\{-\int_0^t [g_x(\xi,X(\xi;t,x)) + \mu(\xi,X(\xi;t,x))]d\xi\right\} \\ &= v_0(1-(1-x)e^{bt};g)\exp\left\{-\int_0^t -b\,d\xi\right\} \\ &= v_0(1-(1-x)e^{bt};g)e^{bt} \end{aligned}$$

for $t \le G(x) = -\frac{1}{b}\ln\left(\frac{1-x}{1-\underline{x}}\right)$. Our computations with the GRD model (2.3) are very straightforward since we have an explicit solution for the SS model (2.1) for the examples we consider.

2.3 Approximation Methods

We are interested in using the GRD model to describe size-structured marine populations. Given some data, we would like to determine the growth rate distribution P^* that gives the best fit of the underlying model to the data. However, the general optimization problem which we outline in Chapter 3 involves both an infinite-dimensional state space (u) and an infinite-dimensional parameter space (the space $\mathcal{P}$ of probability measures) which poses problems if one wants to use this model in inverse problem calculations. Thus, approximation methods leading to efficient computations are important for this purpose. We now briefly discuss the different approximation methods that we consider in the inverse problem for the estimation of growth rate distributions of size-structured populations.

2.3.1 Parametric Approach: PAR(M, N)

In the first method that we consider for this problem, we use the standard parametric approach based on the assumption that we have a priori knowledge about the exact form of the probability distribution on the growth rates of the mosquitofish. If we assume that the probability distribution P is (absolutely) continuous, then $\frac{dP(b)}{db} = p(b)$, where $p(b)$ represents the probability density function (pdf) corresponding to P. The population density from the GRD model (2.3) is then given by

$$u(t, x; \theta) = \int_{\mathfrak{B}} v(t, x; g(x; b))p(b; \theta)db, \tag{2.8}$$

where θ represents the parameters that are associated with the a priori probability density and distribution. We denote this approach by PAR(M, N), where M is one less than the number of parameters in θ and N is the number of quadrature nodes used in approximating the integral above with the composite trapezoidal rule [56]. We set M to one less than the number of parameters in θ so when using the asymptotic standard error theory which is outlined in Chapter 3 the correct factor is used in our computations.

Results for one of the computational examples in Chapter 4 were obtained with simulated data (representative of typical mosquitofish data observed in rice fields such as that in Figure 2.1) generated with a bi-Gaussian distribution on the growth rates. The bimodality usually seen in the mosquitofish data has been attributed to the fact that male and female mosquitofish grow to different maximum sizes. Previous simulations [7] demonstrated that an assumption of a bi-Gaussian distribution on the growth rates leads to both dispersion and bifurcation, qualitative features that are observed in mosquitofish data. The following bi-Gaussian probability density function p, which is an average of two Gaussian distributions, was used to create one of the data sets used in our computational tests

$$p(b; \mu_{b_1}, \sigma^2_{b_1}, \mu_{b_2}, \sigma^2_{b_2}) = \frac{\exp\left\{-\frac{(b-\mu_{b_1})^2}{2\sigma^2_{b_1}}\right\}}{2\sqrt{2\pi\sigma^2_{b_1}}} + \frac{\exp\left\{-\frac{(b-\mu_{b_2})^2}{2\sigma^2_{b_2}}\right\}}{2\sqrt{2\pi\sigma^2_{b_2}}}, \tag{2.9}$$

where the parameters (μ_{b_1}, μ_{b_2}) and $(\sigma^2_{b_1}, \sigma^2_{b_2})$ represent the means and variances, respectively, of a bi-Gaussian distribution on the intrinsic growth rates b. An example of a bi-Gaussian density and distribution are shown in Figure 2.3. In terms of the parametric method PAR(M, N), $\theta = (\mu_{b_1}, \sigma^2_{b_1}, \mu_{b_2}, \sigma^2_{b_2})$, and we set $M = 3$ even though θ consists of four parameters. Thus, the parameter θ is contained in $\mathbb{R}^{M+1}_+$. We remark that being

able to identify the form of the probability distribution a priori is required when using this approach.

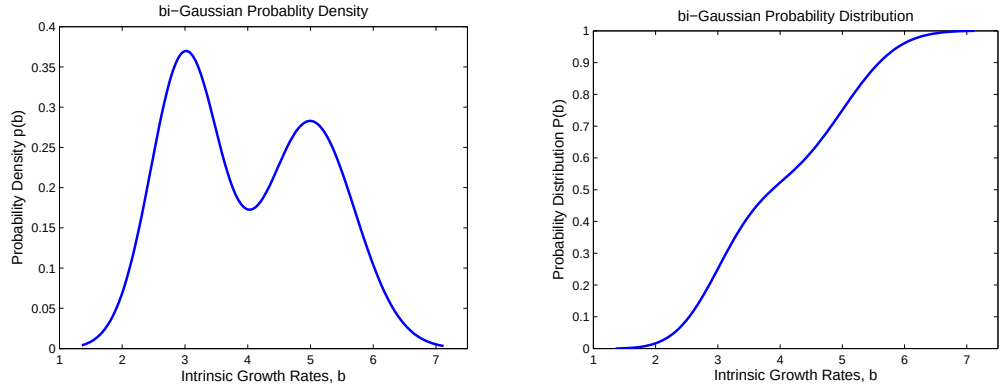

Figure 2.3: Example of a bi-Gaussian probability density and distribution.

2.3.2 Non-parametric Approaches: DEL(M) and SPL(M, N)

As pointed out in the previous section, one must be able to identify the form of the probability distribution using prior knowledge in order to use the parametric approach. Knowledge of the form of the distribution leads to efficient computations when the parameter set is relatively small. However, one may not be able to specify (correctly) the probability distribution a priori which could result in poor fits to data or incorrect conclusions based on good fits to data with a misspecified distribution [23]. Non-parametric approaches do not require any assumptions with respect to the form of the probability distribution. Instead of using a specific distribution in the GRD model (2.3), we use finite approximations to the probability distribution. Based on work in [4] and [24] which we summarize in Chapter 3, we are guaranteed convergence (in the Prohorov metric [6, 29]) of distributions with the families of approximating functions that we now discuss.

The first approach, which has been discussed and used in [13] and [14], involves the use of delta functions or Dirac delta measures. The population is assumed to consist of a finite collection of subpopulations with corresponding probabilities. We assume that the form of the approximating probability distribution $\mathcal{P}^M$ placed on the growth rates is *piecewise constant* on the collection $\mathcal{G}^M = \{g_k^M\}_{k=0}^M$ of admissible growth rates $g_k^M(x; b_k^M) = b_k^M(1-x)$ for $k = 0, 1, \ldots, M$. Figure 2.4 shows a delta function approximation to a bi-Gaussian dis-

tribution. The stem plot is our representation of Dirac measures (delta functions) with atoms at $\{b_k^M\}$ "approximating" the probability density shown in Figure 2.3. Notice the difference in the scale in the plot of the probability density in Figure 2.3 versus the scale in the plot of the approximate probability density shown in Figure 2.4. We will say more about this later. The piecewise constant approximate probability distribution generated with the delta function method is also shown in Figure 2.4 (we note the vertical lines are not a part of the approximation of the probability distribution).

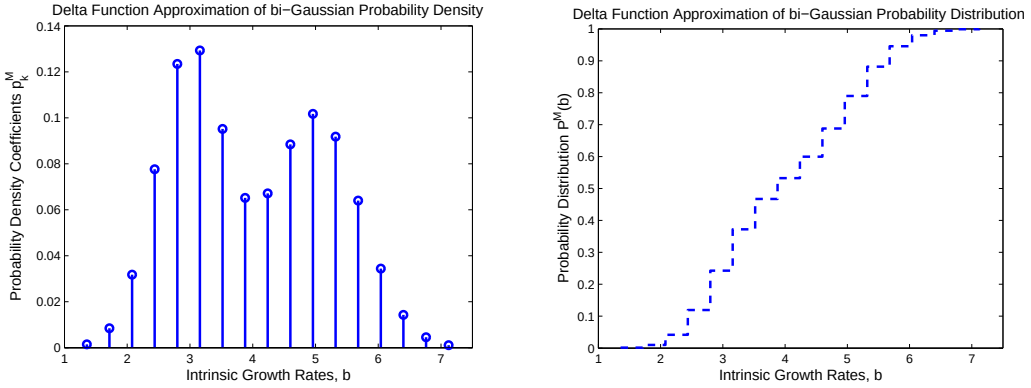

Figure 2.4: Example of a delta function approximation of a bi-Gaussian probability density and distribution.

For each subpopulation with growth rate g_k^M, there is a corresponding probability p_k^M that an individual is in subpopulation k. The population density $u(t, x; P)$ in (2.3) is then approximated by

$$u(t, x; \{p_k^M\}) = \sum_{k=0}^{M} v(t, x; g_k^M) p_k^M, \tag{2.10}$$

where $v(t, x; g_k^M)$ is the subpopulation density from (2.1) with growth rate g_k^M. We denote the delta function approximation method as DEL(M), where $M + 1$ delta functions are used in this approximation.

While it has been shown that DEL(M) provides a nice approximation to (2.3), a better approach might involve techniques that will provide a smoother approximation of (2.3) and faster convergence of the approximated probability distribution in the case of continuous probability distributions on the growth rates. Using piecewise linear splines in the place of delta functions we provide a much smoother approximation of (2.3) when the "true" probability distribution on the growth rates of the mosquitofish is continuous.

Assuming the form of the approximating probability distribution on the growth rates is *continuous,* the second non-parametric approximation scheme involves the use of piecewise linear spline functions to approximate the density $\mathcal{P}'(b) = \frac{d\mathcal{P}}{db}(b) = p(b)$. In Figure 2.5 we see a plot of a spline approximation to the bi-Gaussian distribution in Figure 2.3. We note that the approximate probability density is linear on each interval $[b_k, b_{k+1}]$. We also point out that the scale in the plot of the approximate probability density in Figure 2.5 is the same as the scale in the plot of the "true" probability density in Figure 2.3 which is explained in the following chapters. Moreover, we observe a much smoother approximation of the bi-Gaussian probability distribution when splines are used as noticed in Figure 2.5.

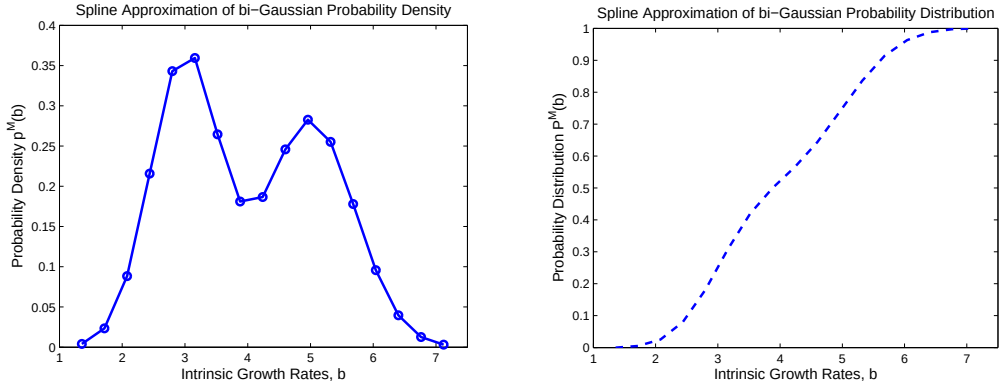

Figure 2.5: Example of a spline function approximation of a bi-Gaussian probability density and distribution.

Denoting this method as SPL(M, N), where $M + 1$ basis elements (splines) are used to approximate the probability distribution on the growth rates, $u(t, x; \mathcal{P})$ from (2.3) is approximated by

$$u(t, x; \{a_k^M\}) = \sum_{k=0}^{M} a_k^M \int_{\mathfrak{B}} v(t, x; g(x; b)) l_k^M(b) db, \tag{2.11}$$

where $g(x; b) = b(1 - x)$ and $p_k^M(b) = a_k^M l_k^M(b)$ is the probability density for individuals in subpopulation k. N represents the number of quadrature nodes used to approximate the integral in (2.11). The composite trapezoidal rule [56] was also used to approximate the integral in (2.11) in the computations that were performed.

In the next chapter we examine the inverse problem for the estimation of the probability distribution P on the growth rates for size-structured populations. We outline the different formulations of the inverse problem for the parametric and non-parametric meth-

ods discussed in this chapter. We also summarize the theoretical background guaranteeing convergence of the approximated distributions considered here.

Chapter 3

Parameter Estimation Problem

We discuss the parameter estimation problem for the growth rate distribution that gives the best fit of the underlying GRD model to a given data set in this chapter. The ordinary least squares (OLS) problem involves determining the probability distribution P^* that minimizes the cost functional

$$J(P) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_x} |u(t_i, x_j; P) - \hat{u}_{ij}|^2 \tag{3.1}$$

over the space of probability measures $\mathcal{P}$, where $u(t_i, x_j; P)$ is the solution to (2.3) and $\hat{u}_{ij}$ is the data as described in Chapter 4. Here N_t represents the number of time points and N_x represents the number of size classes. We see that we must use approximation methods in order to solve (3.1) since it involves both an infinite-dimensional state space (u) and an infinite-dimensional parameter space ($\mathcal{P}$). Using the approximation methods discussed in Chapter 2, the original infinite-dimensional optimization problem in (3.1) is reduced to a finite-dimensional problem. After outlining the three formulations of the approximate inverse problem for PAR(M, N), DEL(M), and SPL(M, N), we summarize the asymptotic standard error theory for finite-dimensional OLS estimators which is used to quantify the uncertainty in the estimates obtained from these three methods.

3.1 Formulation of the Parametric Inverse Problem

We recall in the parametric version that the form of the probability distribution is known and is parameterized by the vector $\theta \in \mathbb{R}_+^{M+1}$. Therefore, the OLS problem in (3.1)

is reduced to the following OLS inverse problem

$$\min_{\theta\in\mathbb{R}_+^{M+1}} J(\theta) = \min_{\theta\in\mathbb{R}_+^{M+1}} \sum_{i,j} |u(t_i, x_j; \theta) - \hat{u}_{ij}|^2, \tag{3.2}$$

which we solve for θ. We note that θ is in R_+^{M+1} since the components of θ correspond to the means and variances associated with the probability distribution and must be positive because they are representative of the growth rate g. The OLS problem in (3.2) is a constrained optimization problem, and we use MATLAB **fmincon** to find an estimate $\hat{\theta}$ corresponding to the minimal value $J(\hat{\theta})$. Since this problem is constrained, we use the active-set optimization algorithm option. After determining an optimal value for θ, we then use this value in the a priori probability density function p to generate the estimated probability density. Since the following relation holds between a continuous probability density function and its corresponding probability distribution

$$P(b) = \int_{-\infty}^{b} p(\xi)d\xi,$$

an estimate for the probability distribution P is computed by integrating the estimated probability density p.

3.2 Formulation of the Non-parametric Inverse Problem

In Chapter 1 we briefly outlined the underlying theory which is fundamental to the approximation methods and computational schemes used in determining estimates of the probability distribution. We note that it is important to have a topology on the space of probability measures $\mathcal{P}$ and continuity of $P \to J(P)$ on $\mathcal{P}(\Omega)$ for some set Ω as well as compactness results for the theoretical arguments that are underlying this work. As already stated, we make use of the Prohorov metric [6, 29] which is defined below.

Definition 3.2.1 *Let $\mathcal{P}(\Omega)$ be the set of probability measures on the Borel subsets of Ω, where Ω is a complete metric space with metric d. For any closed subset $S \subset \Omega$ and $\epsilon > 0$, define $S^\epsilon \equiv \{\omega \in \Omega | d(\tilde{\omega}, \omega) < \epsilon, \tilde{\omega} \in \Omega\}$. The Prohorov metric $\rho : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \to \mathbb{R}_+$ is defined by*

$$\rho(P_1, P_2) \equiv \inf\{\epsilon > 0 | P_1[S] \leq P_2[S^\epsilon] + \epsilon, S \text{ closed}, S \subset \Omega\}.$$

Some well known properties of the Prohorov metric include

- $(\mathcal{P}(\Omega), \rho)$ is a complete metric space.
- If Ω is compact, then $(\mathcal{P}(\Omega), \rho)$ is a compact metric space.

While the definition of the Prohorov metric ρ is not very intuitive, the following theorem [29] proves to be very useful.

Theorem 3.2.1 *Given $P_k, P \in \mathcal{P}(\Omega)$, the following convergence statements are equivalent:*

- $\rho(P_k, P) \to 0$.
- $\int_\Omega h\, dP_k(\omega) \to \int_\Omega h\, dP(\omega)$ *for all bounded, uniformly continuous functions* $h : \Omega \to \mathbb{R}$.
- $P_k[A] \to P[A]$ *for all Borel sets* $A \subset \Omega$ *with* $P[\partial A] = 0$.

From this theorem, one notes that convergence in the ρ metric is equivalent to convergence in distribution. When one considers $\mathcal{P}(\Omega) \subset C_B^*(\Omega)$, where $C_B^*(\Omega)$ denotes the topological dual of the space $C_B(\Omega)$ of bounded continuous functions on Ω, then convergence in the ρ topology is equivalent to weak star convergence in $\mathcal{P}(\Omega)$. Furthermore, as a result of the equivalence relations in Theorem 3.2.1, we remark that $\rho(P_k, P) \to 0$ is equivalent to

$$\int_\Omega x_l(\omega) dP_k(\omega) \to \int_\Omega x_l(\omega) dP(\omega),$$

or "convergence in expectation" since the above relation is equivalent to

$$\mathcal{E}[x_l(\omega)|P_k] \to \mathcal{E}[x_l(\omega)|P].$$

This establishes continuity of $P \to J(P)$ in the ρ topology, and the existence of a solution to the original parameter estimation problem can be ascertained from the continuity of $P \to J(P)$ and compactness of $\mathcal{P}(\Omega)$ (which follows from compactness of Ω).

Once existence of a solution is established, computational issues must be addressed since the parameter space (the probability measure space $\mathcal{P}$) is infinite-dimensional. As already noted in the Introduction, we seek methods where the finite-dimensional sets $\mathcal{P}^M(\Omega)$ converge to $\mathcal{P}(\Omega)$. Theoretical arguments exist for both the delta function approximation method and the spline based approximation method providing us with two ways to choose approximation families for this purpose. We summarize those arguments and outline the inverse problem formulation in the next two sections for each non-parametric scheme.

3.2.1 Delta Function Approximation Method: DEL(M)

The convergence argument for the approximating family consisting of delta functions (Dirac delta measures) is given by the following theorem developed in [4].

Theorem 3.2.2 *Let Ω be a complete, separable metric space with metric d, $\mathbb{S}$ be the class of all Borel subsets of Ω and $\mathcal{P}(\Omega)$ be the space of probability measures on $(\Omega, \mathbb{S})$. Let $\Omega_0 = \{\omega_j\}_{j=0}^{\infty}$ be a countable, dense subset of Ω. Then the set of $P \in \mathcal{P}(\Omega)$ such that P has finite support in Ω_0 and rational masses is dense in $\mathcal{P}(\Omega)$ in the ρ metric. That is,*

$$\mathcal{P}_0(\Omega) \equiv \left\{ P \in \mathcal{P}(\Omega) : P = \sum_{j=0}^{k} p_j \delta_{\omega_j}, k \in \mathbb{N}, \omega_j \in \Omega_0, p_j \text{ rational }, p_j \geq 0, \sum_{j=0}^{k} p_j = 1 \right\}$$

is dense in $\mathcal{P}(\Omega)$ relative to ρ where δ_{ω_j} is the Dirac measure with atom at ω_j and $\mathbb{N}$ are nonnegative integers.

Therefore, if we define the following finite-dimensional approximation $\mathcal{P}^M(\Omega)$ to the probability measure space $\mathcal{P}(\Omega)$

$$\mathcal{P}^M(\Omega) = \left\{ P \in \mathcal{P}(\Omega) | P = \sum_{k=0}^{M} p_k^M \delta_{\omega_k}, \sum_{k=0}^{M} p_k^M = 1 \right\},$$

then $\mathcal{P}^M(\Omega)$ is a compact subset of $\mathcal{P}(\Omega)$ in the ρ metric with $\mathcal{P}^M(\Omega) \subset \mathcal{P}^{M+1}(\Omega)$. Furthermore, we note that as $M \to \infty$ elements in $\mathcal{P}^M(\Omega)$ can be used to approximate elements in $\mathcal{P}(\Omega)$ as a result of $\mathcal{P}_0(\Omega) = \bigcup_{M=0}^{\infty} \mathcal{P}^M(\Omega)$ being dense in $\mathcal{P}(\Omega)$. It is important to point out that this theory guarantees convergence of distributions, not densities, in the Prohorov metric.

We note that the original parameter estimation problem in (3.1) is approximated by the least squares problem

$$\begin{aligned} \min_{P \in \mathcal{P}^M(\mathcal{G})} J(P) &= \min_{P \in \mathcal{P}^M(\mathcal{G})} \sum_{i,j} |u(t_i, x_j; P) - \hat{u}_{ij}|^2 \\ &= \min_{P \in \mathcal{P}^M(\mathcal{G})} \sum_{i,j} \left(u(t_i, x_j; P)^2 - 2u(t_i, x_j; P)\hat{u}_{ij} + (\hat{u}_{ij})^2 \right), \end{aligned} \tag{3.3}$$

where $\hat{u}_{ij}$ is the data and $\mathcal{P}^M(\mathcal{G})$ is the finite-dimensional approximation to $\mathcal{P}(\mathcal{G})$. The finite-dimensional approximation $\mathcal{P}^M(\mathcal{G})$ is given by

$$\mathcal{P}^M(\mathcal{G}) = \left\{ P \in \mathcal{P}^M(\mathcal{G}) | P(b) = \sum_{k=0}^{M} p_k^M \delta_{b_k^M}(b), \sum_{k=0}^{M} p_k^M = 1 \right\}, \tag{3.4}$$

where $\delta_{b_k^M}(b)$ is the delta function with an atom at b_k^M for $g_k^M(x; b_k^M) = b_k^M(1-x)$. Furthermore, this least squares inverse problem (3.3) becomes a constrained quadratic programming problem [13, 14]. Letting $\mathbf{p}$ be the vector that contains p_k^M, $0 \leq k \leq M$, when using DEL(M) we define $\mathcal{A}$ as the matrix with entries given by

$$\mathcal{A}_{km} = \sum_{i,j} v(t_i, x_j; g_k^M) v(t_i, x_j; g_m^M),$$

$\mathbf{b}$ as the vector with entries given by

$$\mathbf{b}_k = -\sum_{i,j} \hat{u}_{ij} v(t_i, x_j; g_k^M),$$

and c as the scalar

$$c = \sum_{i,j} (\hat{u}_{ij})^2,$$

where $0 \leq k, m \leq M$. The optimization problem in (3.3) reduces to finding a minimizer of

$$F(\mathbf{p}) \equiv \mathbf{p}^T \mathcal{A} \mathbf{p} + 2\mathbf{p}^T \mathbf{b} + c \tag{3.5}$$

over $\mathcal{P}^M(\mathcal{G})$. We remark that non-negativity constraints on the weights $\{p_k^M\}$ as well as the last constraint in (3.4) are included in the programming of the inverse problem which we solve with the MATLAB function **quadprog**.

3.2.2 Spline Based Approximation Method: SPL(M, N)

In our earlier discussion, we commented on the advantage of using piecewise linear splines versus delta functions in approximations of continuous probability distributions. Banks-Pinter [24] developed the following theorem which is applicable when arguing convergence of $\mathcal{P}^M(\mathcal{G})$ to $\mathcal{P}(\mathcal{G})$ in the spline based approximation scheme.

Theorem 3.2.3 *Let $\mathcal{F}$ be a weakly compact subset of $L^2(\Omega)$, with Ω compact, and let $\mathcal{P}_{\mathcal{F}}(\Omega)$ be the family of probability distribution functions on Ω generated by $\mathcal{F}$ as a set of densities*

$$\mathcal{P}_{\mathcal{F}}(\Omega) = \{P \in \mathcal{P}(\Omega) | P' = p, p \in \mathcal{F}\}.$$

Then $\mathcal{P}_{\mathcal{F}}(\Omega)$ is a ρ compact subset of $(\mathcal{P}(\Omega), \rho)$ where ρ is the Prohorov metric on the set $\mathcal{P}(\Omega)$ of all probability density functions on Ω.

As a result of this theorem, one can choose a piecewise linear spline approximation $p^M(\omega)$ to $p(\omega)$

$$p^M(\omega) = \sum_{k=0}^{M} a_k^M l_k^M(\omega),$$

where $\{a_k^M\}$ are rational numbers with l_k^M representing the piecewise linear spline functions such that $p^M \to p \in L^2(\Omega)$ since $\mathcal{F} \subset L^2(\Omega)$. One can then show convergence of P^M to P in the ρ metric. Moreover, if $\mathcal{F}^M = \{h \in L^2(\Omega) | h(\omega) = \sum_{k=0}^{M} a_k^M l_k^M(\omega)\}$, then $\bigcup_M \mathcal{P}_{\mathcal{F}^M}(\Omega)$ is dense in $\mathcal{P}_{\mathcal{F}}(\Omega)$ in the Prohorov metric when $\mathcal{P}_{\mathcal{F}^M}(\Omega) = \{P \in \mathcal{P}(\Omega) | P = \int p, p \in \mathcal{F}^M\}$. We remark that this theory guarantees convergence of the distributions in the Prohorov metric as well as convergence of the densities in L^2 with appropriately chosen coefficients $\{a_k^M\}$.

The least squares problem in the case of the spline based approximation method is the same as that of the delta function approximation method in (3.3). However, the finite-dimensional approximation $\mathcal{P}^M(\mathcal{G})$ is given by

$$\mathcal{P}^M(\mathcal{G}) = \left\{ P \in \mathcal{P}(\mathcal{G}) | P'(b) = \sum_{k=0}^{M} a_k^M l_k^M(b), \sum_{k=0}^{M} a_k^M \int_{\mathfrak{B}} l_k^M(b) db = 1 \right\} \tag{3.6}$$

for $g(x; b) = b(1 - x)$. Again, the least squares inverse problem reduces to minimizing the constrained quadratic programming problem in (3.5) over $\mathcal{P}^M(\mathcal{G})$. For SPL$(M, N)$, let $\mathbf{p}$ be the vector that contains $a_k^M, 0 \leq k \leq M$. Then the entries in the matrix $\mathcal{A}$ are

$$\mathcal{A}_{km} = \sum_{i,j} \int_{\mathcal{B}} v(t_i, x_j; g)^2 l_k^M(b) l_m^M(b) db,$$

the entries in the vector $\mathbf{b}$ are

$$\mathbf{b}_k = -\sum_{i,j} \hat{u}_{ij} \int_{\mathcal{B}} v(t_i, x_j; g) l_k^M(b) db,$$

and

$$c = \sum_{i,j} (\hat{u}_{ij})^2,$$

where $0 \leq k, m \leq M$. We note again that non-negativity constraints on $\{a_k^M\}$ as well as the last condition in (3.6) are included in the programming of the inverse problem which is carried out with MATLAB **quadprog**.

3.3 Asymptotic Standard Error Theory for Finite-Dimensional Parameters

We have discussed how the original infinite-dimensional inverse problem (3.1) is reduced to a finite-dimensional problem in each of the methods considered (θ in PAR(M, N), $\{p_k^M\}_{k=0}^M$ for DEL(M), and $\{a_k^M\}_{k=0}^M$ for SPL(M, N)). We briefly outline the standard statistical framework for asymptotic distributions of finite-dimensional ordinary least squares (OLS) estimators [36, 40, 45, 59, 60] which we use to determine information about the reliability of the estimates of the finite-dimensional parameters from the approximation methods.

We begin by considering the following nonlinear statistical model for the observation or measurement process

$$Y_j = Y(\tau_j) = f(\tau_j; \theta_0) + \epsilon_j, \quad j = 1, \ldots, n, \tag{3.7}$$

where τ_j is the jth covariate component of a vector in $\mathbb{R}^n$, $f(\tau_j; \theta_0)$ represents the solution to the mathematical model, and θ_0 is a vector in the constraint set $\Theta \subset \mathbb{R}^{M+1}$ that represents the "true" parameter value. We assume that the ϵ_j are *i.i.d.* with mean 0 and constant variance σ_0^2, where $\sigma_0^2 > 0$ represents the "true" variance. Generally, θ_0 and σ_0^2 are not known but are estimated by the parameters θ and σ^2, respectively. Since ϵ_j is a random variable, Y_j is also a random variable with

$$\mathcal{E}[Y_j] = f(\tau_j; \theta_0) \quad \text{and} \quad Var[Y_j] = \sigma_0^2.$$

As already noted, we consider ordinary least squares formulations of the inverse problem. The following OLS estimator (which is also a random variable denoted here by $\theta_{OLS} = \theta_{OLS}(Y)$) is used in the inverse problem for the estimation of θ :

$$\theta_{OLS} \equiv \arg\min_{\theta \in \Theta} \sum_{j=1}^{n} \left(Y_j - f(\tau_j; \theta)\right)^2. \tag{3.8}$$

The distribution of θ_{OLS} is called the sampling distribution and is very important because it provides information about the uncertainty or reliability of the estimates obtained for a particular realization of $\{Y_j\}$. As $n \to \infty$, the sampling distribution for a random variable $\theta_{OLS}(Y)$ is given by the multivariate normal distribution; i.e.,

$$\theta_{OLS}(Y) \sim \mathcal{N}_{M+1}(\theta_0, \sigma_0^2[\mathcal{X}^T(\theta_0)\mathcal{X}(\theta_0)]^{-1}) \approx \mathcal{N}_{M+1}(\theta_0, \Sigma_0^n),$$

where $\mathcal{X}(\theta) = \mathcal{X}^n(\theta) = \frac{\partial F}{\partial \theta}(\theta) = F_\theta(\theta)$ is the $n \times (M+1)$ sensitivity matrix with elements

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\tau_j;\theta)}{\partial \theta_k},$$

and Σ_0^n is an approximate covariance matrix defined below in (3.9). As we noted, θ_0 is generally unknown; however, we can determine an estimate $\hat{\theta}$ for θ_0 using the OLS estimator. For a particular realization (data set) $\{y_j\}$ the estimates $\hat{\theta}$ minimize

$$\sum_{j=1}^{n} \left(y_j - f(\tau_j;\theta)\right)^2.$$

We can also compute an estimate for σ_0^2 (which is also usually unknown) using the following estimate $\hat{\sigma}^2$:

$$\sigma_0^2 \approx \hat{\sigma}^2 = \frac{1}{n-(M+1)} \sum_{j=1}^{n} \left(y_j - f(\tau_j;\hat{\theta})\right)^2.$$

The estimates $\hat{\theta}$ and $\hat{\sigma}^2$ are used in computing an estimate of the covariance matrix Σ_0^n :

$$\Sigma_0^n \approx \hat{\Sigma} = \hat{\sigma}^2 [\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})]^{-1}. \tag{3.9}$$

We can then determine the standard errors for the estimates $\hat{\theta}$ by computing

$$SE(\hat{\theta}_k) = \sqrt{\hat{\Sigma}_{kk}}, \quad k = 0, \ldots, M.$$

Confidence intervals for the finite-dimensional parameter $\hat{\theta}$ are constructed using the standard errors.

Before presenting the computational results demonstrating the confidence intervals, we make a few comments about the covariance matrix $\hat{\Sigma}$. We notice that determining $\hat{\sigma}^2$ is very straightforward once we have computed an estimate $\hat{\theta}$. We simply multiply the residual $J(\hat{\theta})$ by the factor $\frac{1}{n-(M+1)}$ to compute $\hat{\sigma}^2$. We must also compute $\mathcal{X}(\hat{\theta})$ which can be more difficult in general when dealing with nonlinear systems. In some cases, computing $\mathcal{X}$ can be very straightforward when one is able to derive analytically the entries of $\mathcal{X}(\theta)$ based on $f(\tau_j;\theta)$. However, when this is not the case, there are other ways to compute $\mathcal{X}(\theta)$ including using forward differences or solving the corresponding sensitivity equations [12]. We are actually able to determine analytical expressions for the entries of the sensitivity matrix in the computational examples considered here. Furthermore, the asymptotic distributional results are exact when $f(\tau;\theta)$ is linear in θ (see [36, 40, 45]) as opposed to only being an approximation when $f(\tau;\theta)$ is nonlinear in θ since the results in the nonlinear case are based on a linearization of $f(\tau;\theta)$.

The endpoints of the confidence intervals are given by

$$\hat{\theta}_k \pm t_{1-\alpha/2} SE(\hat{\theta}_k), \quad k = 0, \ldots, M$$

and are computed at the $100(1-\alpha)\%$ level where $\alpha \in [0, 1]$. The confidence level parameters are defined so that

$$P\{\hat{\theta}_k - t_{1-\alpha/2} SE(\hat{\theta}_k) < \theta_{0k} < \hat{\theta}_k + t_{1-\alpha/2} SE(\hat{\theta}_k)\} = 1 - \alpha$$

where $t_{1-\alpha/2} \in \mathbb{R}_+$. The value $t_{1-\alpha/2}$ is a distribution value that is determined by the level of significance α chosen [32]. After a level of significance is chosen (for example, $\alpha = 0.05$ for 95% confidence intervals), we determine the corresponding $t_{1-\alpha/2}$ value from a statistical table for the Student's t-distribution. The confidence intervals constructed in this manner provide us with a means of quantifying the uncertainty of the estimates obtained from the estimation procedure constructed from a realization of Y. In the following chapter, we present some computational results in which we have used this asymptotic standard error theory to compute nodal confidence intervals corresponding to the finite-dimensional parameters.

Chapter 4

Computational Example: Size-Structured Mosquitofish Population

We compare computational and statistical results of the delta function approximation method, the spline based approximation method, and the parameterized OLS formulation based on inverse problem calculations performed in MATLAB in this chapter. Before applying the methods to experimental data, we used the methods with simulated data to demonstrate the validity of the theoretical arguments presented in the previous chapter. Convergence as well as sensitivity of the estimates with respect to noise in the data is discussed for both non-parametric approximation methods. We also compare the results from the three schemes for the estimation of the distribution of growth rates with field data for a size-structured mosquitofish population.

4.1 Analysis of Methods with Simulated Data

In order to obtain information on the strengths and weaknesses of the non-parametric approximation methods, we first began by preparing simulated population density data independent of the two schemes DEL(M) and SPL(M, N) used in the inverse problem. Since we were only concerned at this point with the estimation of the growth rate distribution P^*, we let $\mu = 0$ and $K = 0$ in the Sinko-Streifer model. We then chose a "true" distribution

P^* on the size-dependent growth rates $g(x;b)$, where we assumed $g(x;b) = b(1-x)$ for $(\underline{x}, \bar{x}) = (0,1)$ and b is the intrinsic growth rate of the mosquitofish as stated in the previous outline. More specifically, we assumed that the intrinsic growth rate b of the mosquitofish is a random variable with distribution P^*. This allowed us to generate a collection of growth rates $\mathcal{G}_n = \{g_1, g_2, \ldots, g_n\}$ corresponding to a partition $\mathfrak{B}_n = \{b_1, b_2, \ldots b_n\}$, where we took $n = 200$ in order to get a good approximation of

$$u(t,x;P^*) = \int_{\mathcal{G}} v(t,x;g(x;b))dP^*(g)$$

and $\mathcal{G}_n \subset \mathcal{G}$. Our simulated data $u_d(t,x;P^*)$ was then created by first computing the solution $v(t,x;g_i)$ of the Sinko-Streifer model for each individual g_i using the method of characteristics as discussed earlier and then computing

$$u_d(t,x;P^*) = \int_{\mathcal{G}_n} v(t,x;g)dP^*(g)$$

using the Gauss-Legendre integration method [56]. We took 50 uniformly spaced time observations, where the time interval was $[0, 0.5]$. The range of size values $(\underline{x}, \bar{x})$ was normalized to $(0,1)$ and 50 uniformly spaced size values were used in this range for our simulated data. For the initial size density, we used

$$v_0(x;g) = \begin{cases} \sin^2(10\pi x), & 0 \le x \le 0.1, \\ 0, & 0.1 < x \le 1. \end{cases}$$

We generated two data sets for our computations in this section: one with a "true" approximate truncated Gaussian probability distribution on the intrinsic growth rates b and one with a "true" approximate truncated bi-Gaussian distribution on b. We use truncated distributions on the intrinsic growth rates to ensure that the growth rates are always positive. Negative growth rates could result in negative sizes for the mosquitofish which are not biologically reasonable. The particular parameter values and examples used here are based on previous formulations discussed in [7] and [13]. Similar results that were obtained for these examples were also discussed in [8].

4.1.1 Convergence of Estimated Probability Distributions from DEL(M) and SPL(M,N)

The results in this section with simulated data sets with no noise demonstrate convergence of distributions with the families of approximating functions that we discussed

previously. For our first set of data, we placed an approximate truncated Gaussian distribution on b with probability density function $p(b;\mu_b,\sigma_b^2)$

$$p(b;\mu_b,\sigma_b^2)=\frac{1}{\sqrt{2\pi\sigma_b^2}}\exp\left\{-\frac{(b-\mu_b)^2}{2\sigma_b^2}\right\}$$

with mean $\mu_b=4.5$ and variance $\sigma_b^2=0.25$, where $b\in\mathfrak{B}=[\underline{b},\bar{b}]$. We let $\underline{b}=\mu_b-3\sigma_b$ and $\bar{b}=\mu_b+3\sigma_b$ in order to include approximately 99% of the intrinsic growth rates. However, to ensure that this approximate truncated Gaussian distribution was indeed a true distribution, we had to scale the weights used in the Gauss-Legendre integration method so that the following condition was satisfied

$$\int_{\underline{b}}^{\bar{b}}p(b;\mu_b,\sigma_b)db=1, \tag{4.1}$$

where $p(b;\mu_b,\sigma_b)$ is the probability density function corresponding to the "true" probability distribution P^*. Using the set of data generated with this approximate truncated Gaussian distribution, we then performed the inverse problem using DEL(M) and SPL(M,N) to find estimates of the "true" probability density and distribution (under the assumption that the "true" probability distribution P^* was unknown) and then compared these estimates to the known density and distribution. We reiterate that we do not make any assumptions with respect to the specific form of the probability distribution with the non-parametric approaches DEL(M) and SPL(M,N). As discussed earlier in Chapter 3, this inverse problem is simplified to a quadratic programming problem for both DEL(M) and SPL(M,N) when using least squares, where we minimize $\mathbf{p}^T\mathcal{A}\mathbf{p}+2\mathbf{p}^T\mathbf{b}+c$.

Overall, we found that both DEL(M) and SPL(M,N) produced good estimates of the "true" Gaussian growth rate distribution using the simulated data; however, we also found that in some cases SPL(M,N) resulted in poor estimates of the probability distribution when M and N were not chosen correctly. This will be addressed in more detail in the next section. In Figure 4.1, we have the results from the inverse problem using DEL(16), DEL(48) and SPL(16,128). We see from the results in Figure 4.1 that the spline based approximation method converges in distribution much faster than the delta function approximation method for a given M, which we expect since the "true" distribution was smooth and continuous. While the estimates of the probability distribution from DEL(16) have not quite converged, the estimates of the probability distribution from SPL(16,128) have converged completely to the "true" distribution. As M is increased, the estimated

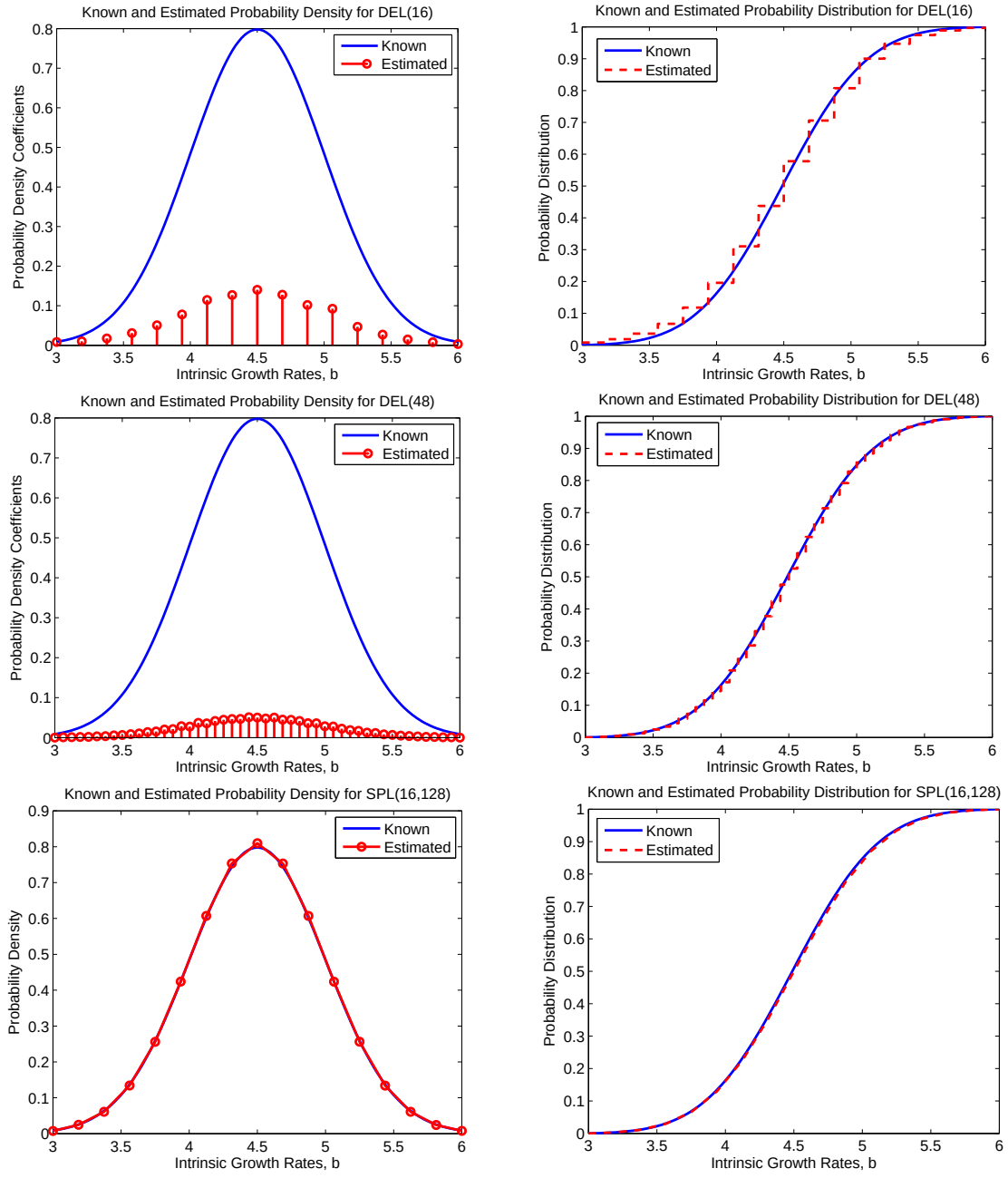

Figure 4.1: Estimated probability densities and distributions for Gaussian example using (a) DEL(16), (b) DEL(48), (c) SPL(16,128).

probability distributions from DEL(M) become better as seen in the results for $M = 48$ in Figure 4.1. We also note that along with convergence in distribution, SPL(M, N) also provides convergence in density, whereas DEL(M) does not provide convergence in density. This may be attributed largely to the difference in the formulations of these two methods. DEL(M) requires

$$\sum_{k=0}^{M} p_k^M = 1,$$

where p_k^M denotes the probability coefficients. On the other hand, SPL(M, N) requires

$$\sum_{k=0}^{M} a_k^M \int_{\mathfrak{B}} l_k^M(b)db = 1,$$

where $p_k^M(b) = a_k^M l_k^M(b)$ is the probability density for an individual in subpopulation k with $l_k^M(b)$ representing the piecewise linear spline functions. Since the "true" density was in fact smooth and continuous, one would not expect convergence in density when using DEL(M) because it is much cruder in its approximation of (4.1). We remark that this agrees fully with the underlying theory for convergence of distributions in the Prohorov metric wherein convergence of densities is not guaranteed with DEL(M) as we discussed in Chapter 3. These different formulations also explain the difference in the scale of the approximated bi-Gaussian probability density from DEL(M) with the "true" bi-Gaussian probability density (noted earlier in Chapter 2).

We also used a second set of data in the inverse problem with an approximate truncated bi-Gaussian distribution on the intrinsic growth rates b. As mentioned before, previous work in [7] suggested that this type of distribution leads to data which exhibits two traits (dispersion and bifurcation) observed in actual mosquitofish field data [7, 13, 14]. To obtain a bi-Gaussian distribution, we took the average of two Gaussian distributions, one with mean $\mu_{b_1} = 3.3$ and variance $\sigma_{b_1}^2 = 0.492$ and the second with mean $\mu_{b_2} = 5.7$ and variance $\sigma_{b_2}^2 = 0.492$. The bi-Gaussian probability density function is given in (2.9). The simulated data was prepared in the same way as described for the Gaussian example except with $b \in \mathfrak{B} = [\underline{b}, \bar{b}] = [\mu_{b_1} - 3\sigma_{b_1}, \mu_{b_2} + 3\sigma_{b_2}]$.

The results for the inverse problem using this set of data are shown in Figure 4.2 for DEL(24), DEL(84), and SPL(24,128). Both methods do a good job of estimating the bi-Gaussian probability distribution with the simulated data. Again, we see that SPL(M, N) converges to the "true" distribution faster than DEL(M). Significantly more basis elements

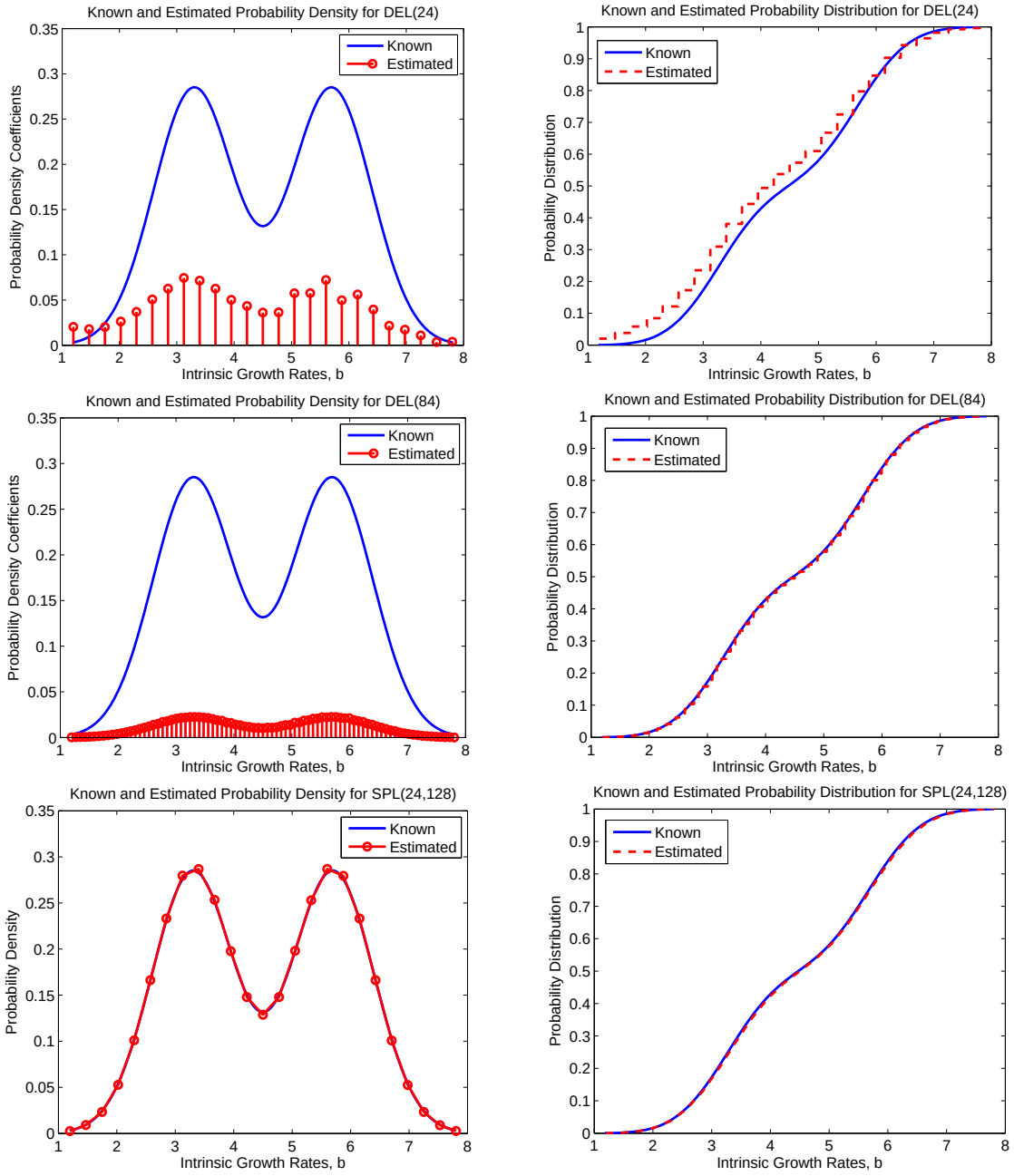

Figure 4.2: Estimated probability densities and distributions for bi-Gaussian example using (a) DEL(24), (b)DEL(84), (c) SPL(24,128).

are required for full convergence of the approximations from DEL(M) to the "true" bi-Gaussian probability distribution. However, it should be noted that more basis elements (larger values of M) were required in both methods to achieve the same level of accuracy in approximating the bi-Gaussian probability distribution in comparison to the Gaussian distribution. As mentioned earlier, the delta function approximation method only results in convergence in distribution, whereas the spline based approximation method results in both convergence in density and distribution.

4.1.2 Stability Issues of SPL(M, N)

In the computational examples discussed in the previous section, SPL(M, N) produced good approximations to both the Gaussian and bi-Gaussian densities and distributions. However, as we stated, obtaining good approximations when using SPL(M, N) was very much dependent on choosing M and N appropriately. In fact, we found if M and N were not chosen carefully, the estimates of the probability distributions from the inverse problem using SPL(M, N) were not very good as a result of the problem becoming unstable. However, by studying the condition number of the matrix $\mathcal{A}$ from the quadratic programming problem, we found that this behavior could be readily explained. There are several different ways in which the condition number $\kappa(\mathcal{A})$ of a matrix $\mathcal{A}$ can be described [56]. In terms of the norm $\|\cdot\|$ of a matrix, we have $\kappa(\mathcal{A}) = \|\mathcal{A}^{-1}\| \cdot \|\mathcal{A}\|$. An equivalent definition for the condition number of a matrix $\mathcal{A}$ when using the 2-norm $\|\cdot\|_2$ is also given as the ratio of the largest singular value to the smallest singular value in the singular value decomposition of the matrix $\mathcal{A}$ (see [56] for further discussion).

What is of most importance is the information one learns from studying the condition number of a matrix. The matrix $\mathcal{A}$ is well-conditioned (well-behaved) if $\kappa(\mathcal{A})$ is relatively small. On the other hand, $\mathcal{A}$ is ill-conditioned (ill-behaved) if $\kappa(\mathcal{A})$ is relatively large. Thus, if $\kappa(\mathcal{A})$ is very large, meaning $\mathcal{A}$ is ill-conditioned, the inverse problem becomes unstable which leads to poor approximations of the probability distribution P. We note that the discussion here is limited to SPL(M, N) based on the fact that the matrix $\mathcal{A}$ for the spline based method can become ill-conditioned for a given M based on the number of quadrature nodes N used in the composite trapezoidal method used for integration purposes as discussed earlier. However, the matrix $\mathcal{A}$ in DEL(M) does not change for a given M due to the way in which the population density and $\mathcal{A}$ is obtained as discussed in the

previous section. Since we must use a quadrature method to compute $\mathcal{A}$ for SPL(M, N), we expect the number of quadrature nodes N used to have a role in determining $\kappa(\mathcal{A})$. In fact, if N is chosen too small for a given M, meaning the quadrature rule gives a very coarse approximation to the actual integration, then we expect $\kappa(\mathcal{A})$ to become larger, the problem to become unstable, and the estimates of the probability distribution to become poor.

To explore the validity of this argument, we computed the condition number of $\mathcal{A}$ for $M = 4$, 20, 36, 68, 84, and 100 and $N = 64$, 128, 256, and 512 in both the Gaussian and bi-Gaussian examples that were used above. Similar results were obtained with different values of M and N and discussed in [8]. The resulting condition numbers of $\mathcal{A}$ for the Gaussian example are given in Table 4.1. As we can see from the values in the table, when $M = 4$, 20, and 36 the condition number $\kappa(\mathcal{A})$ is relatively small for all listed values of N. We found that the inverse problem resulted in good estimates of the Gaussian probability distribution when using SPL$(4, N)$, SPL$(20, N)$ and SPL$(36, N)$. On the other hand, at $M = 68$, we began to see some significant differences in $\kappa(\mathcal{A})$, as shown by the values in Table 4.1. It can be noted that for $M = 68$, 84, and 100, $\kappa(\mathcal{A})$ was very large when using SPL$(M, 64)$. However, there was a significant decrease in the condition number of $\mathcal{A}$ when using SPL(M, N) for $N = 128$, 256, and 512 for these values of M. This difference in condition numbers was also evident in the estimates of the probability distribution from the inverse problem. For $M = 68$, 84, and 100, the estimates when using SPL$(M, 64)$ were worse than those obtained when using SPL(M, N) for all other listed values of N. We note the estimates of the Gaussian probability distribution became better as the value of N was increased for these fixed values of M.

Table 4.1: Condition numbers of $\mathcal{A}$ for Gaussian example when using SPL(M, N).

M	N	$\approx \kappa(\mathcal{A})$
4	64,128,256,512	9.30
20	64,128,256,512	55
36	64,128,256,512	460
68	64	10^{17}
68	128,256,512	10^{5}
84	64	10^{17}
84	128,256,512	10^{6}
100	64	10^{33}
100	128,256,512	10^{6}

For the bi-Gaussian example, we also computed $\kappa(\mathcal{A})$ for the same values of M and N, and the results in this case are given in Table 4.2. The results obtained when using a "true" approximate truncated bi-Gaussian probability distribution were very similar to the results obtained when using a "true" approximate truncated Gaussian probability distribution. At $M = 4$, 20, and 36, $\kappa(\mathcal{A})$ was relatively small (13.35, 57, and 130, respectively, for each value of N). The inverse problem for these values of M was also stable, and the estimates of the probability distribution in both cases were good. However, when $M = 68$, 84, and 100, SPL$(M, 64)$ results in very large condition numbers for $\mathcal{A}$. The estimates for the probability distribution using SPL$(M, 64)$ were very poor in comparison to the estimates obtained from SPL(M, N) for $M = 68$, 84, 100 and $N = 128$, 256, 512. When SPL(M, N) was used in the inverse problem for the estimation of the bi-Gaussian probability distribution for $M = 68$, 84, 100 and for values of N greater than 64, we saw a significant decrease in the condition numbers of $\mathcal{A}$ (see values in Table 4.2) corresponding to better approximations of the probability distribution.

Table 4.2: Condition numbers of $\mathcal{A}$ for bi-Gaussian example when using SPL(M, N).

M	N	$\approx \kappa(\mathcal{A})$
4	64,128,256,512	13.35
20	64,128,256,512	57
36	64,128,256,512	130
68	64	10^{17}
68	128,256,512	755
84	64	10^{17}
84	128,256,512	10^{3}
100	64	10^{19}
100	128,256,512	10^{4}

To summarize these computational results, in both the Gaussian and bi-Gaussian case, when $N > M$, $\kappa(\mathcal{A})$ was relatively low, which resulted in good estimates of the growth rate probability distributions when using SPL(M, N). However, when using SPL(M, N), for $M \sim N$, the condition number of $\mathcal{A}$ was very large, resulting in an ill-posed inverse problem and poor estimates of the probability distributions. For a fixed M, we observed that as the value of N increased, the condition number of $\mathcal{A}$ decreased, which agrees with results in [17, 18]. Therefore, we have shown by these computational efforts that we can "regularize" the inverse problem when using SPL(M, N) by choosing proper ratios of M

and N, which can be described as "regularization by discretization balance." By using a finer discretization in the quadrature method used in SPL(M, N), we were able to obtain better results in the inverse problem involving the estimation of growth rate distributions.

4.1.3 Sensitivity of Estimated Probability Distributions from DEL(M) and SPL(M, N)

We used simulated data with no noise in the previous section to show convergence of the estimated probability distributions to the "true" distribution with both non-parametric approximation methods. However, experimental data is usually corrupted by noise, which can be a result of errors in collecting the data, errors in the instruments and techniques used, etc. Along with verifying that both SPL(M, N) and DEL(M) produce estimates which converge in distribution when simulated data with no noise is used in the inverse problem, we also wanted to make some remarks about the sensitivity with respect to noisy data of the estimates of the probability distributions from the two non-parametric approximation schemes. Thus, we added random absolute noise to the simulated data used in the previous two examples in the following way:

$$\hat{u}(t, x; P^*) = u_d(t, x; P^*) + \eta \cdot \epsilon,$$

where η represents the noise level constant and ϵ represents normally distributed random values with mean 0 and variance 1. We then performed the inverse problem again using $\eta = 0.005, 0.025, 0.05$ corresponding to $1\%, 5\%$, and 10% absolute error, respectively, for both the Gaussian and bi-Gaussian cases.

We begin by discussing the results of the inverse problem using the simulated data with a "true" approximate truncated Gaussian distribution with the various noise level constants. Both approximation methods, DEL(M) and SPL(M, N), performed decently in the inverse problem with the varying percentages of absolute error. With only 1% absolute error, both DEL(M) and SPL(M, N), with M and N chosen appropriately, resulted in estimates that converged to the "true" growth rate probability distribution in very much the same manner as previously noted in the Gaussian example with no noise. The performance of these approximations methods was not greatly affected by the small amount of noise in the data. With a slightly larger percentage of absolute error in the simulated data, SPL(M, N) and DEL(M) were still able to produce good estimates of the probability distribution.

However, the results from the inverse problem using $\eta = 0.025$ began to exhibit some small effects in the estimates obtained from both DEL(M) and SPL(M, N).

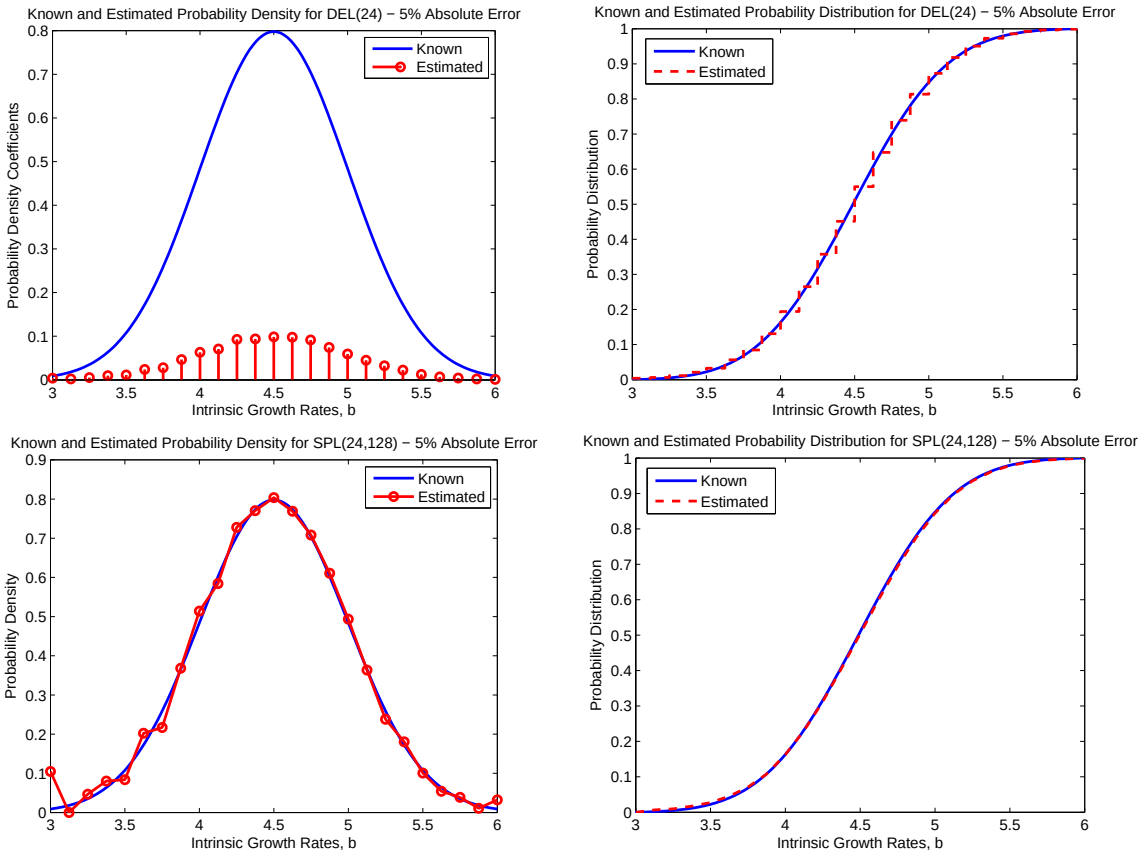

Figure 4.3: Estimated probability densities and distributions for Gaussian example using DEL(24) (upper) and SPL(24,128) (lower) with 5% absolute error.

For example, in Figure 4.3, the approximated probability distributions for DEL(24) and SPL(24,128) with 5% absolute error show slightly overestimated front tails. Moreover, there are small perturbations in the estimated probability density from SPL(24,128) when the data has 5% absolute error in comparison to very smooth estimated probability densities when no noise was present in the data. With very noisy data, $\eta = 0.05$, SPL(M, N) and DEL(M) still perform fairly well. From the results for DEL(24) and SPL(24,128), shown in Figure 4.4, the noisier data resulted in only slightly poorer approximations in comparison to those obtained with 5% absolute noise. As more absolute noise (as opposed to relative noise) was added to the simulated data, the front and end tails of the estimated probability distribution are more affected as a result of the oscillations seen in the estimated probability

densities. Furthermore, the larger amount of noise produced more oscillatory behavior in the approximated probability densities for both DEL(M) and SPL(M, N) which resulted in poorer approximations of the corresponding probability distributions.

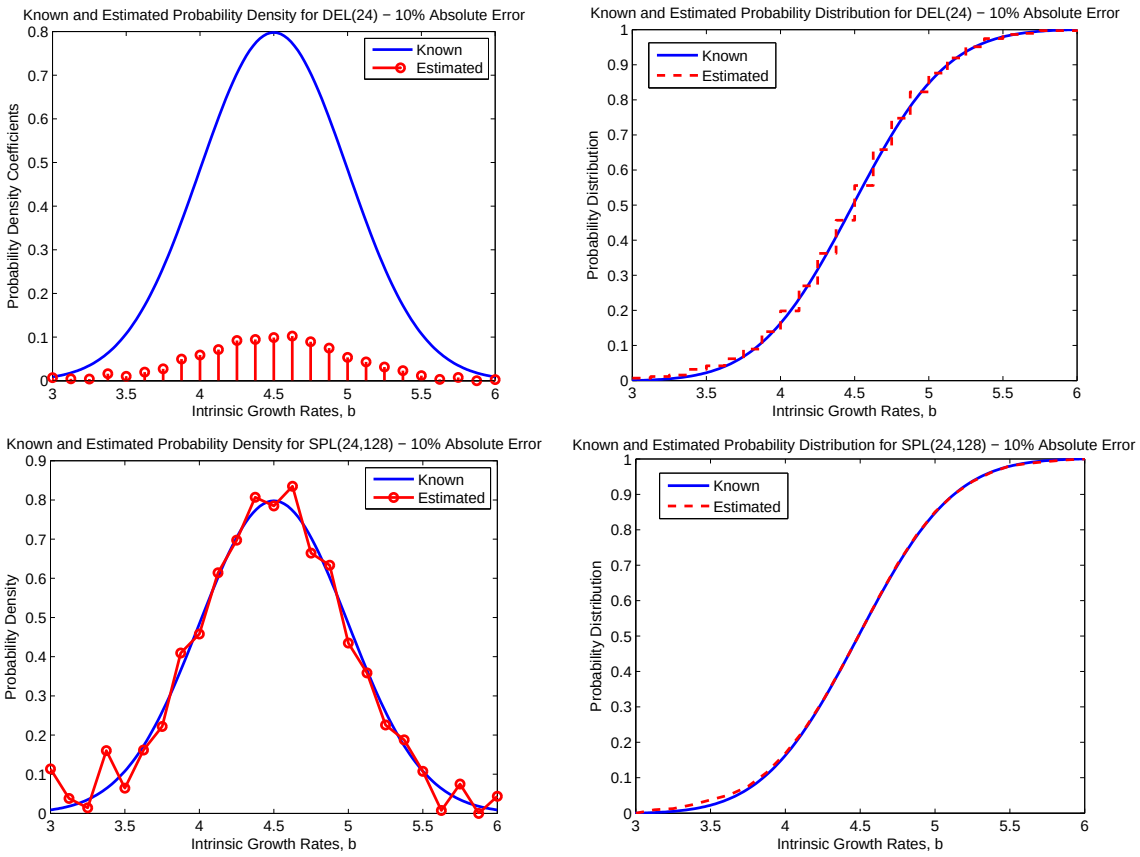

Figure 4.4: Estimated probability densities and distributions for Gaussian example using DEL(24) (upper) and SPL(24,128) (lower) with 10% absolute error.

We also tested the two non-parametric approximation methods for sensitivity to error in the bi-Gaussian data with the same percentages of absolute error considered with the Gaussian data. The results obtained from the inverse problem using DEL(M) and SPL(M, N) with noisy data with a "true" approximate truncated bi-Gaussian distribution were very similar to those obtained when using noisy data with a "true" approximate truncated Gaussian distribution. Overall, the estimated probability distributions from DEL(M) and SPL(M, N) were not largely affected by the various amounts of noise added to the simulated data. Both methods were able to produce good approximations of the probability distributions in the presence of noise. With 1% absolute error in the data, the estimates of

the growth rate distributions from DEL(M) and SPL(M, N) did not change significantly from estimates obtained when there was no noise in the data. We were still able to obtain convergence in distribution (with faster convergence when using SPL(M, N)) with both approximation methods. When the percentage of absolute error in the data was 5%, DEL(M) and SPL(M, N) still performed well by producing good estimates of the bi-Gaussian probability distribution. The small amount of noise had some small effect on the estimates as seen in Figure 4.5; in fact, we see that the front tails in both the estimate from DEL(32) and SPL(32,128) for $\eta = 0.025$ are slightly largely than the tails for the "true" distribution. When even more noise is present in the data, the estimated probability distributions became slightly poorer for a fixed M and N. In Figure 4.6, the estimated probability densities and distributions are shown for DEL(32) and SPL(32,128) for data with 10% absolute error. It is clear from these plots that the estimates from DEL(M) and SPL(M, N) are indeed affected by the noisier data. As in the Gaussian example, we observed some oscillatory behavior in the estimated probability densities from these two approximation methods as the amount of noise present in the data increased. Moreover, the front tails in the estimated probability distributions are overestimated, whereas the end tails are underestimated for both DEL(32) and SPL(32,128).

While we were still able to obtain convergence in distribution using both DEL(M) and SPL(M, N), with M and N chosen carefully, our computational results demonstrated that as more noise was added to both the Gaussian and bi-Gaussian data the estimates of the growth rate distributions from both methods became slightly poorer for a fixed M and N. SPL(M, N) produced probability distribution estimates that converged faster in distribution than DEL(M) using both data with noise and without noise. This behavior, as stated earlier, was expected since the "true" probability distributions in these numerical examples are smooth and continuous. We remark again that while the computational results for SPL(M, N) exhibited convergence in density as well as distribution, convergence in density is generally not guaranteed and is not supported by the general theory in the Prohorov metric [4, 14, 29] as observed in the computational results obtained with DEL(M).

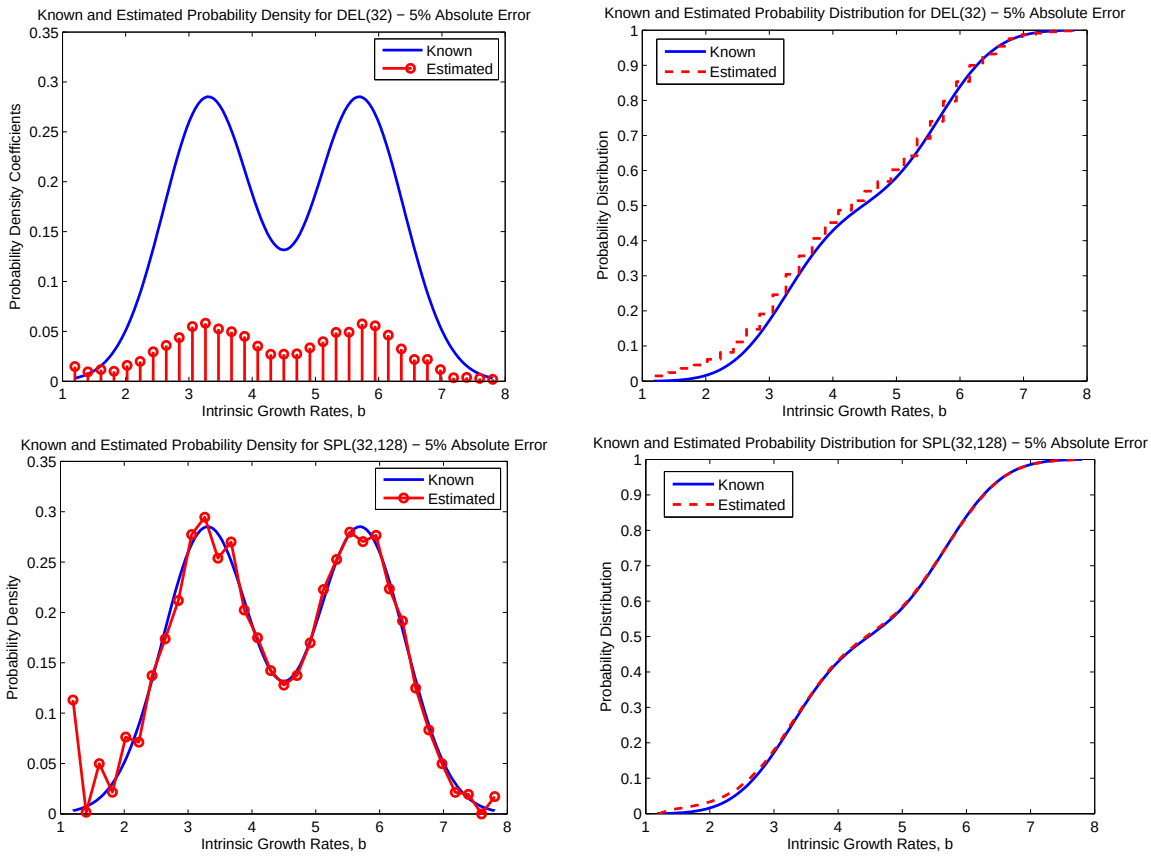

Figure 4.5: Estimated probability densities and distributions for bi-Gaussian example using DEL(32) (upper) and SPL(32,128) (lower) with 5% absolute error.

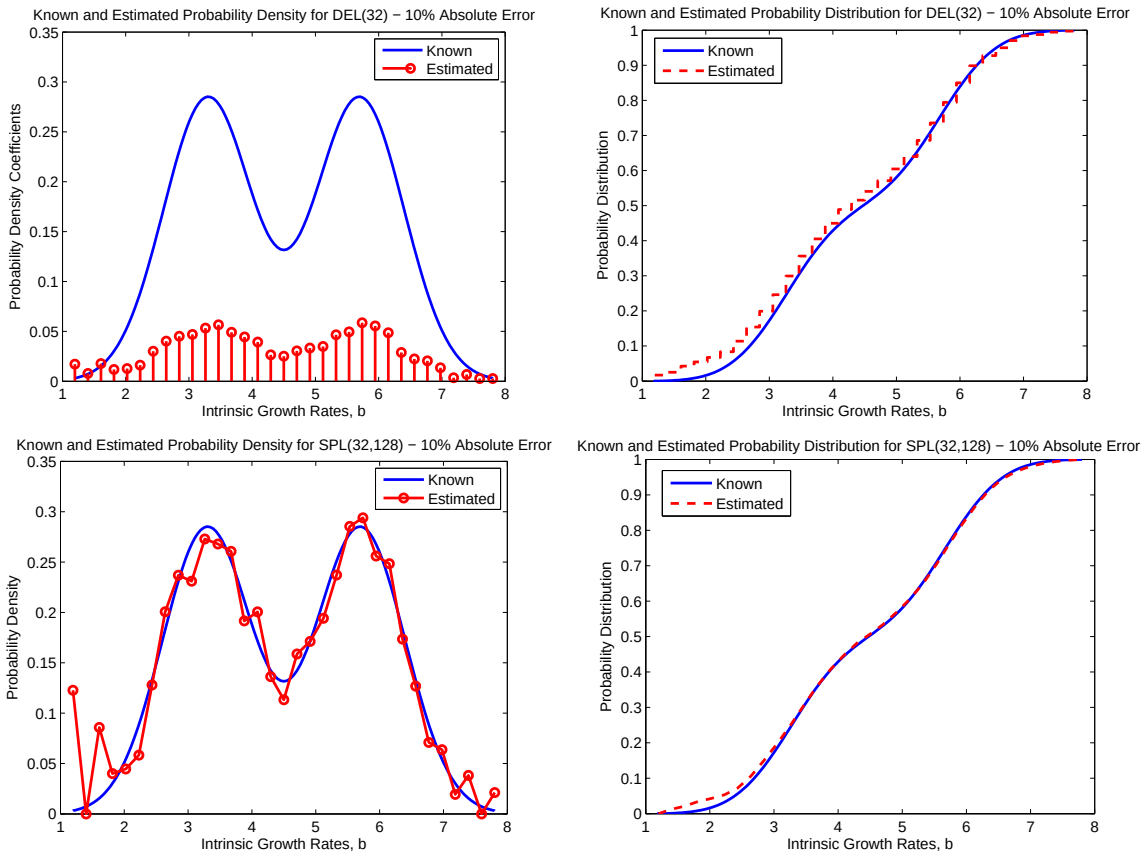

Figure 4.6: Estimated probability densities and distributions for bi-Gaussian example using DEL(32) (upper) and SPL(32,128) (lower) with 10% absolute error.

4.1.4 Statistical Analysis of Non-parametric Approximation Methods

We next considered the sensitivity of the two approximation methods, DEL(M) and SPL(M, N), with respect to noisy data in order to make some comments about the uncertainty associated with the estimated growth rate distributions. We cannot physically observe the entire population; however, we can include some measures (e.g., confidence intervals) on the uncertainty of the estimates obtained from the two non-parametric approximation methods when using only a sample from the population. The following analysis is based on the asymptotic standard error theory for finite-dimensional parameters discussed in Section 3.3.

Since we have already discussed in detail the assumptions of the statistical model, we will define the variables and functions that are used in these examples and refer the reader to Section 3.3 for further explanations. We note that $\{\tau_j\}_{j=1}^n$ corresponds to $(t_l, x_m), l = 1, \ldots, N_t, m = 1, \ldots, N_x$ pairs where N_t corresponds to the number of time values and N_x corresponds to the number of size values used with $N = N_t \cdot N_x$. When using DEL(M), $\theta_0 \approx \theta = \{p_k^M\}_{k=0}^M$ and $\theta_0 \approx \theta = \{a_k^M\}_{k=0}^M$ when using SPL(M, N). Recall that θ_0 represents the "true" parameter value (assumed to exist). For a particular realization or data set $\{y_j\}_{j=1}^n$ we determine estimates of θ_0 by finding solutions to the quadratic programming problem (3.5) discussed in Chapter 3. Furthermore, we point out that

$$f(\tau_j; \theta) = \sum_{k=0}^{M} v(\tau_j; g_k^M) p_k^M \tag{4.2}$$

where $g_k^M(x; b_k^M) = b_k^M(1 - x)$ when considering DEL(M), while

$$f(\tau_j; \theta) = \sum_{k=0}^{M} a_k^M \int_{\mathfrak{B}} v(\tau_j; g) l_k^M(b) db \tag{4.3}$$

where $g(x; b) = b(1 - x)$ for b in some compact set $\mathfrak{B}$ when considering SPL(M, N).

We previously commented on the computation of the sensitivity matrix $\mathcal{X}(\theta)$, which is used to compute the estimated covariance matrix $\hat{\Sigma} = \hat{\sigma}^2[\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})]^{-1}$. As noted earlier, the elements of the $n \times (M + 1)$ matrix $\mathcal{X}(\theta)$ are given by

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\tau_j; \theta)}{\partial \theta_k}.$$

These are actually the sensitivity elements associated with this system. Since both (4.2) and (4.3) are linear in θ, then computing the entries of the sensitivity matrix in this case is

very straightforward. The entries in the sensitivity matrix $\mathcal{X}(\theta)$ for DEL(M) are given by

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\tau_j;\theta)}{\partial \theta_k} = v(\tau_j; g_k^M),$$

where $\theta_k = p_k^M$ is the probability parameter associated with growth rate $g_k^M(x; b_k^M) = b_k^M(1-x)$. For SPL($M,N$), the entries in the sensitivity matrix $\mathcal{X}(\theta)$ are given by

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\tau_j;\theta)}{\partial \theta_k} = \int_{\mathfrak{B}} v(\tau_j; g) l_k^M(b) db$$

where $\theta_k = a_k^M$ and $p_k^M(b) = a_k^M l_k^M(b)$ is the probability density for an individual in subpopulation k. We reiterate that the asymptotic distributional results given earlier are exact in this case (see [36, 40, 45]) as opposed to only being an approximation when $f(\tau_j;\theta)$ is nonlinear in θ.

We present next some findings on the sensitivity of DEL(M) and SPL(M,N) using simulated data generated with a "true" approximate truncated bi-Gaussian distribution. The endpoints of the confidence intervals for the parameter estimates $\hat{\theta}$ are given by

$$\hat{\theta} \pm t_{1-\alpha/2} SE(\hat{\theta}),$$

where $t_{1-\alpha/2}$ is a Student's t-distribution value that is determined by the level of significance that is chosen [32]. We chose to use $\alpha = 0.05$ for a significance level of 95%, which corresponds to $t_{1-\alpha/2} \approx 1.96$ when the number of samples is large, i.e. $n \geq 30$. Based on the confidence intervals, we can make statements about the estimation procedure constructed from a realization of Y. If the resulting confidence intervals are relatively large in relation to $\hat{\theta}$, then we are not very confident about the estimation procedure used to estimate θ_0.

Table 4.3 displays the estimated probability density values and corresponding confidence intervals for DEL(8) while Table 4.4 displays the results for SPL(8,128) in the presence of 5% and 10% absolute error. In Figures 4.7 and 4.8, we see the confidence intervals corresponding to the estimated parameters (p_k^M or a_k^M) with $\alpha = 0.05$ for DEL(8) and SPL(8,128), respectively, with simulated data with both 5% and 10% absolute error. We plot the results of the delta function approximation method without the stems so that the reader can see clearly the lower endpoints of the nodal confidence intervals corresponding to the estimated p_k^M values denoted by the circles in the figures. The simulated data used in these computations was of the form discussed in the previous chapter. Based on the significance level chosen, we can state that we are 95% confident that intervals constructed

Table 4.3: Estimated probability density values and confidence intervals for bi-Gaussian example for DEL(8) with simulated data with 5% and 10% absolute error.

p_k^8	DEL(8) - 5%	DEL(8) - 10%
p_0^8	0.1724 ± 0.0192	0.1750 ± 0.0196
p_1^8	0.1506 ± 0.0171	0.1487 ± 0.0175
p_2^8	0.1665 ± 0.0149	0.1644 ± 0.0153
p_3^8	0.1432 ± 0.0128	0.1386 ± 0.0131
p_4^8	0.0948 ± 0.0115	0.1005 ± 0.0117
p_5^8	0.1084 ± 0.0098	0.1058 ± 0.0100
p_6^8	0.0953 ± 0.0090	0.0946 ± 0.0092
p_7^8	0.0442 ± 0.0080	0.0482 ± 0.0082
p_8^8	0.0246 ± 0.0072	0.0242 ± 0.0073

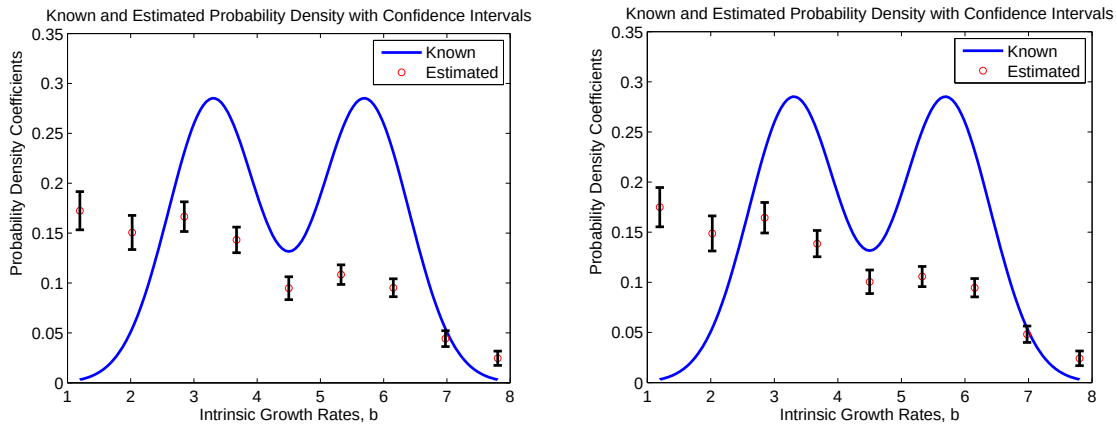

Figure 4.7: Estimated probability densities with confidence intervals for bi-Gaussian example using (a) DEL(8) with 5% absolute error and (b) DEL(8) with 10% absolute error.

Table 4.4: Estimated probability density values and confidence intervals for bi-Gaussian example for SPL(8,128) with simulated data with 5% and 10% absolute error.

a_k^8	SPL(8,128) - 5%	SPL(8,128) - 10%
a_0^8	0.0204 ± 0.0154	0.0461 ± 0.0298
a_1^8	0.0460 ± 0.0096	0.0453 ± 0.0186
a_2^8	0.2447 ± 0.0082	0.2404 ± 0.0159
a_3^8	0.2651 ± 0.0073	0.2506 ± 0.0142
a_4^8	0.0896 ± 0.0065	0.1121 ± 0.0127
a_5^8	0.2737 ± 0.0058	0.2592 ± 0.0112
a_6^8	0.2475 ± 0.0051	0.2508 ± 0.0098
a_7^8	0.0389 ± 0.0045	0.0397 ± 0.0087
a_8^8	0.0000 ± 0.0063	0.0039 ± 0.0122

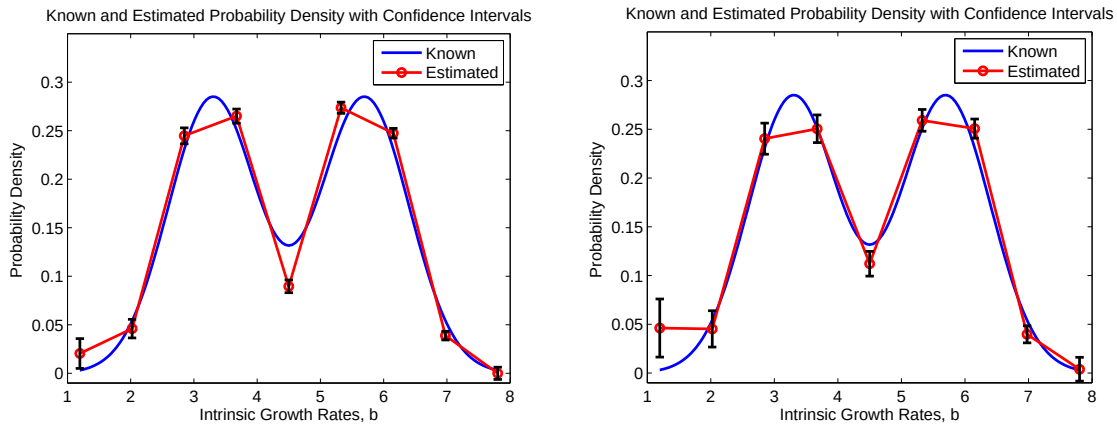

Figure 4.8: Estimated probability densities with confidence intervals for bi-Gaussian example using (a) SPL(8,128) with 5% absolute error and (b) SPL(8,128) with 10% absolute error.

using the estimation procedures with DEL(M) and SPL(M, N) would "cover" θ_0. We note that for the fixed value $M = 8$ the confidence intervals corresponding to $\alpha = 0.05$ when using DEL(M) are relatively small in relation to the estimated $\hat{\theta}$ for data with both 5% and 10% absolute error. Moreover, we see from Figure 4.7 that the resulting confidence intervals for DEL(8) with 5% and 10% absolute error are approximately the same. Thus, for $M = 8$, the delta function approximation method appears to be insensitive to noisy data. In comparison, we note for the fixed value $M = 8$ when using SPL($M, 128$) the resulting confidence intervals are relatively larger for data with 10% absolute error in comparison to those for data with 5% absolute error. Thus, the confidence associated with the estimator procedure based on SPL(M, N) appears to decrease as the percentage of absolute error increases. However, the confidence intervals are still relatively small in relation to the estimates $\hat{\theta}$. Therefore, as a result of these computations, we would infer that for this fixed value of M, the spline based approximation method appears to be very slightly sensitive to very noisy data.

The estimated probability density values and corresponding confidence intervals for DEL(16) and SPL(16,128) in the presence of 5% and 10% absolute error are given in Tables 4.5 and 4.6, respectively. In Figures 4.9 and 4.10, we see the confidence intervals corresponding to the parameter estimates with $\alpha = 0.05$ for DEL(16) and SPL(16,128) with simulated data with both 5% and 10% absolute error. The endpoints of the confidence intervals are constructed in the same way as discussed earlier. We can again state that we

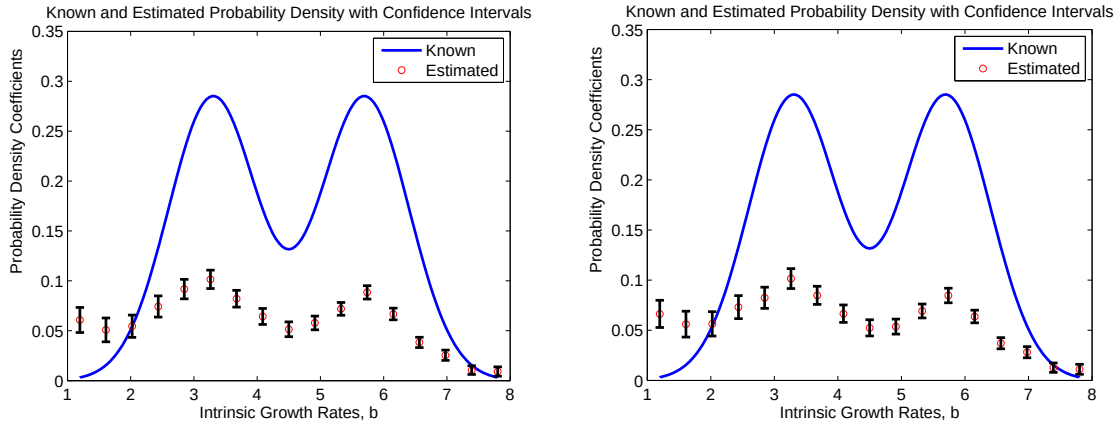

Figure 4.9: Estimated probability densities with confidence intervals for bi-Gaussian example using (a) DEL(16) with 5% absolute error and (b) DEL(16) with 10% absolute error.

Table 4.5: Estimated probability density values and confidence intervals for bi-Gaussian example for DEL(16) with simulated data with 5% and 10% absolute error.

p_k^{16}	DEL(16) - 5%	DEL(16) - 10%
p_0^{16}	0.0608 ± 0.0261	0.0663 ± 0.0136
p_1^{16}	0.0508 ± 0.0119	0.0561 ± 0.0129
p_2^{16}	0.0545 ± 0.0112	0.0564 ± 0.0121
p_3^{16}	0.0743 ± 0.0106	0.0731 ± 0.0115
p_4^{16}	0.0918 ± 0.0098	0.0824 ± 0.0106
p_5^{16}	0.1015 ± 0.0092	0.1016 ± 0.0099
p_6^{16}	0.0821 ± 0.0084	0.0848 ± 0.0091
p_7^{16}	0.0643 ± 0.0080	0.0666 ± 0.0087
p_8^{16}	0.0575 ± 0.0075	0.0524 ± 0.0081
p_9^{16}	0.0578 ± 0.0069	0.0524 ± 0.0081
p_{10}^{16}	0.0719 ± 0.0064	0.0537 ± 0.0075
p_{11}^{16}	0.0884 ± 0.0067	0.0847 ± 0.0073
p_{12}^{16}	0.0668 ± 0.0058	0.0637 ± 0.0063
p_{13}^{16}	0.0383 ± 0.0052	0.0371 ± 0.0056
p_{14}^{16}	0.0254 ± 0.0052	0.0281 ± 0.0056
p_{15}^{16}	0.0106 ± 0.0043	0.0127 ± 0.0047
p_{16}^{16}	0.0092 ± 0.0047	0.0111 ± 0.0050

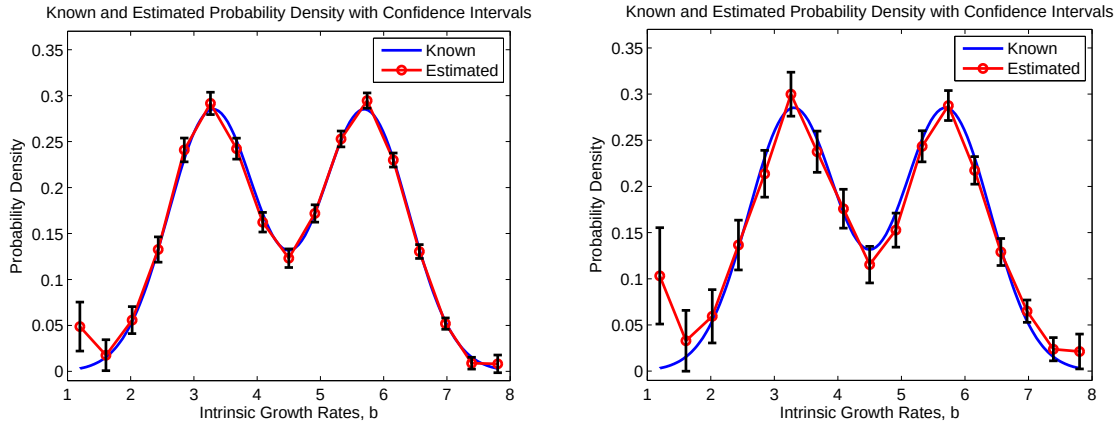

Figure 4.10: Estimated probability densities with confidence intervals for bi-Gaussian example using (a) SPL(16,128) with 5% absolute error and (b) SPL(16,128) with 10% absolute error.

Table 4.6: Estimated probability density values and confidence intervals for bi-Gaussian example for SPL(16,128) with simulated data with 5% and 10% absolute error.

a_k^{16}	SPL(16,128) - 5%	SPL(16,128) - 10%
a_0^{16}	0.0488 ± 0.0266	0.1032 ± 0.0521
a_1^{16}	0.0176 ± 0.0168	0.0328 ± 0.0329
a_2^{16}	0.0558 ± 0.0147	0.0594 ± 0.0288
a_3^{16}	0.1326 ± 0.0138	0.1365 ± 0.0269
a_4^{16}	0.2409 ± 0.0129	0.2137 ± 0.0253
a_5^{16}	0.2916 ± 0.0121	0.2999 ± 0.0238
a_6^{16}	0.2423 ± 0.0114	0.2376 ± 0.0223
a_7^{16}	0.1623 ± 0.0107	0.1759 ± 0.0210
a_8^{16}	0.1231 ± 0.0101	0.1153 ± 0.0198
a_9^{16}	0.1718 ± 0.0094	0.1527 ± 0.0184
a_{10}^{16}	0.2529 ± 0.0086	0.2434 ± 0.0169
a_{11}^{16}	0.0692 ± 0.0069	0.2876 ± 0.0163
a_{12}^{16}	0.2300 ± 0.0076	0.2174 ± 0.0149
a_{13}^{16}	0.1304 ± 0.0075	0.1290 ± 0.0146
a_{14}^{16}	0.0519 ± 0.0062	0.0650 ± 0.0121
a_{15}^{16}	0.0089 ± 0.0064	0.0237 ± 0.0126
a_{16}^{16}	0.0082 ± 0.0096	0.0213 ± 0.0188

are 95% confident that intervals constructed using the estimation procedures with DEL(M) and SPL(M, N) would "cover" θ_0. From Figure 4.9 we see that the confidence intervals corresponding to $\alpha = 0.05$ when using DEL(M) for $M = 16$ are relatively small in relation to $\hat{\theta}$ for data with 5% and 10% absolute error much like those computed for $M = 8$. It can also be noted in this case that the confidence intervals are approximately the same for both sets of data. We arrive at the same conclusion for $M = 16$ as we did for $M = 8$; that is, the delta function approximation method appears to be insensitive to noisy data. We now look at the results for $M = 16$ when using SPL($M, 128$) with data with 5% and 10% absolute error. We observe for $M = 16$ as we did for $M = 8$ that the resulting confidence intervals for SPL(16,128) are larger when the data is noisier. As discussed in the case for $M = 8$, the spline based approximation method also appears to be slightly sensitive to noisy data for $M = 16$ since the confidence intervals are relatively small in relation to the estimates of the a_k^M values.

To summarize, based on the standard error analysis discussed in this section and computational results (those presented here as well as those obtained for $M = 4$) we can conclude that DEL(M) appears to be insensitive to noisy data. Moreover, we can state that we are confident about the estimated growth rate distributions obtained using this method. We also conclude that SPL(M, N) appears to be slightly sensitive to noisy data. Furthermore, we would feel certain about the estimated growth rate distributions obtained using SPL(M, N) with data with small amounts of noise; however, we would infer that larger amounts of noise in the data would lead to larger confidence intervals and less certainty in the associated estimated growth rate distributions obtained using SPL(M, N).

4.2 Application of Methods to Inverse Problem with Experimental Data

In this section we present and discuss the results of the inverse problem for the estimation of growth rate distributions in size-structured mosquitofish populations using the standard parametric approach, the delta function approximation method, and the spline based approximation method. We use field data collected from rice paddies in the place of the simulated data. Since the actual growth rate distribution of the mosquitofish observed in the experiment is unknown, we must compare the field data to the estimated population

data produced by the estimated growth rate distribution from each method in order to compare the efficacy of these methods.

In the inverse problem calculations in this section, we assume that the growth rate of the mosquitofish is now parameterized by both the intrinsic growth rates b and the maximum size γ, where $g(x;b,\gamma) = b(\gamma - x)$. The collection of growth rates is now given by $\mathcal{G} = \{g(x;b,\gamma)|b \in \mathfrak{B}, \gamma \in \Gamma\}$, where both $\mathfrak{B}$ and Γ are bounded closed intervals. We assume here, as we also assumed in our earlier computations, that $\mu = 0$ and $K = 0$, so that we can focus on the growth rate distribution only, but distributions could be placed on μ and K as well. Mortality and fecundity were not thought to be important features of the experimental data of [7, 14].

4.2.1 Experimental Data

The field data that we are using in the inverse problem was collected in an experiment described in [14]. On June 28, 1982, four rice paddies were stocked with mosquitofish. In order to measure emigration, an outflow trap was placed on each paddy. Fifteen traps were used per paddy, and weekly measurements were taken. The length of the mosquitofish ranged from 0 to 40mm, with the mosquitofish being grouped into size classes of 2mm for a total of 20 size classes. The aggregate type data collected on Day 195, Day 202, Day 209, and Day 216 is used in the inverse problem calculations discussed in the following subsections (see Figure 4.11). We define the size distribution frequency for size class j at time t_i as $f_j^i = n_{m,j}^i/N_m$, where $n_{m,j}^i$ is the number of mosquitofish measured at time t_i in size class j and N_m is the total number of mosquitofish measured. Our experimental data was given in this form and is plotted versus the normalized model responses in all figures in this section. For the model responses, the number in size class j at time t_i is given by $\int_{x_j}^{x_{j+1}} u(t_i,x;P)dx$ which, for $\Delta x_j = x_{j+1} - x_j$ small, may be approximated by $u(t_i,x_j;P)\Delta x_j$. The corresponding data is denoted by $d_j(t_i)$ and is given by $N_m f_j^i$. Thus in the inverse problem (3.1) we use data $\hat{u}_{ij} = \frac{d_j(t_i)}{\Delta x_j}$. The total population of mosquitofish is divided into 512 subpopulations. In [14] the discretizations for the intrinsic growth rates b are defined as

$$b_j = 0.2 + \frac{1}{31} \cdot 4.8 \cdot j, \quad j = 0, 1, \ldots, 31,$$

while the discretizations for the maximum sizes γ are defined as

$$\begin{aligned}\gamma_0 &= \frac{16}{38}, \quad \gamma_1 = \frac{22}{38}, \quad \gamma_2 = \frac{24}{38}, \\ \gamma_k &= \frac{16}{38} + \frac{1}{15} \cdot \frac{22}{38} \cdot k, \quad k = 3, \ldots, 15.\end{aligned}$$

As discussed in [14], the discretization for b_j was chosen arbitrarily, while the values for γ_k were chosen as a result of analyzing the collected data. The Day 195 data is interpolated and used an approximation for the initial size density $v_0(x; g)$. Since this data set is used as an approximation for $v_0(x; g)$, it cannot be used in the estimation of the growth rate distributions. Therefore we are left with only three data sets to use in the inverse problem computations. Hence, the experimental data consists of a total of 60 data points since we have 20 size classes per day for three days.

4.2.2 Model Fits with PAR(M, N_1, N_2)

Based on previous work in [13] and [14] and our own numerical simulations with simulated data, we know that a bi-Gaussian growth rate distribution results in population density data with the two key features of dispersion and bifurcation. The experimental data that we use in these computations exhibits these features as well, so we suspect that the underlying growth rate distribution is bi-Gaussian. With that in mind, we chose to use the standard parametric approach with a bi-Gaussian probability density function in the growth rate distribution (GRD) model. The bi-Gaussian probability density function $p(b, \gamma)$ for the two parameter family of growth rates is given by

$$p(b, \gamma; \bar{b}_1, \sigma^2_{b_1}, \bar{b}_2, \sigma^2_{b_2}, \bar{\gamma}_1, \sigma^2_{\gamma_1}, \bar{\gamma}_2, \sigma^2_{\gamma_2}) = \left(\frac{\exp\left\{ \frac{-(b-\bar{b}_1)^2}{2\sigma^2_{b_1}} \right\}}{2\sqrt{2\pi\sigma^2_{b_1}}} + \frac{\exp\left\{ \frac{-(b-\bar{b}_2)^2}{2\sigma^2_{b_2}} \right\}}{2\sqrt{2\pi\sigma^2_{b_2}}} \right) \left(\frac{\exp\left\{ \frac{-(\gamma-\bar{\gamma}_1)^2}{2\sigma^2_{\gamma_1}} \right\}}{2\sqrt{2\pi\sigma^2_{\gamma_1}}} + \frac{\exp\left\{ \frac{-(\gamma-\bar{\gamma}_2)^2}{2\sigma^2_{\gamma_2}} \right\}}{2\sqrt{2\pi\sigma^2_{\gamma_2}}} \right),$$

where we have assumed b and γ to be independent bi-Gaussian random variables in compact sets $\mathfrak{B} = [0.2, 5]$ and $\Gamma = [\frac{16}{38}, 1]$, respectively, based on the values of b_j and γ_k defined in [14]. The parameters $(\bar{b}_1, \bar{b}_2)$ and $(\sigma^2_{b_1}, \sigma^2_{b_2})$ represent the means and variances, respectively, of a bi-Gaussian distribution on the intrinsic rates b, while the parameters $(\bar{\gamma}_1, \bar{\gamma}_2)$ and $(\sigma^2_{\gamma_1}, \sigma^2_{\gamma_2})$ represent the means and variances of a bi-Gaussian distribution on the maximum sizes γ. We define $\theta = \left(\bar{b}_1, \sigma^2_{b_1}, \bar{b}_2, \sigma^2_{b_2}, \bar{\gamma}_1, \sigma^2_{\gamma_1}, \bar{\gamma}_2, \sigma^2_{\gamma_2}\right)$. Since we are now considering a two parameter

family of growth rates, the GRD model in the standard parametric approach that was discussed in Chapter 2 is now given by

$$u(t,x;\theta) = \int_{\mathfrak{B}} \int_{\Gamma} v(t,x;g(x;b,\gamma))p(b,\gamma;\theta)d\gamma db,$$

where again the admissible growth rate function is $g(x;b,\gamma) = b(\gamma - x)$. The addition of the parameter γ introduces another integral that must be approximated as well. This approach is denoted as PAR(M, N_1, N_2) where $M+1$ is the number of parameters in θ to be estimated and N_1 and N_2 represent the number of quadratures used in the composite trapezoidal rule [31] to approximate the double integral with respect to b and γ, respectively. We estimate θ by solving the ordinary least squares problem in (3.2). Once the least squares problem has been solved, we can use the optimal $\hat{\theta}$ in the bi-Gaussian probability density function $p(b,\gamma;\theta)$ to determine the estimated population density $u(t,x;\hat{\theta})$.

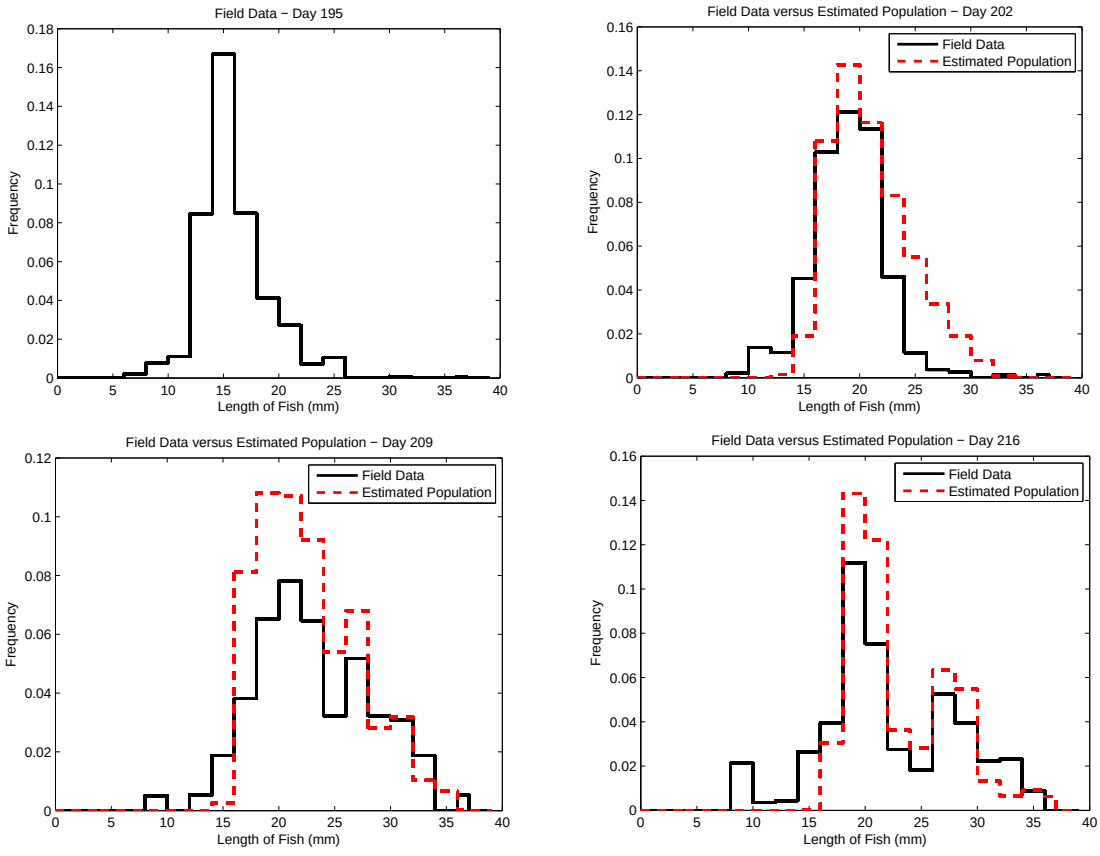

Figure 4.11: Field data versus estimated population for PAR(7,35,35).

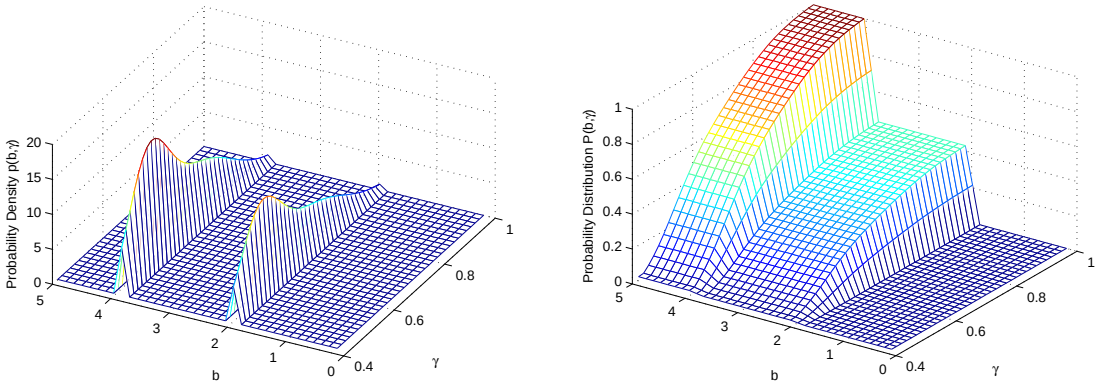

Figure 4.12: (a) Estimated probability density and (b) estimated probability distribution for PAR(7,35,35).

The optimal results for the inverse problem using PAR(7,35,35) are shown in Figure 4.11. The optimal parameters $\hat{\theta} = (1.9749,\ 0.0388 \times 10^{-3},\ 3.9132,\ 0.2228 \times 10^{-3},\ 0.5265,\ 0.0122,\ 0.7208,\ 0.0372)$, with a residual $J(\hat{\theta}) = 0.0074$, were determined in 20.5490 seconds. However, we see from the model fits-to-data in Figure 4.11 that the estimated population density data does not fit the experimental data well. The model does a poor job of predicting the population data for the smaller lengths observed. The bi-Gaussian estimated probability density and probability distribution generated by the optimal $\hat{\theta}$ are shown in Figure 4.12. For a fixed value of γ, we see clearly two distinct modes in the probability density of b. Since the model with the assumption of a bi-Gaussian distribution on the growth rates does not do a very good job of fitting the data, we also used the non-parametric approaches in the inverse problem with the experimental data. The results of these calculations are in the next two sections.

4.2.3 Model Fits with DEL(M_1, M_2)

We introduced the delta function approximation method, DEL(M), earlier when the growth rate is parameterized by b only. Since we are now considering a growth rate parameterized by b and γ, the approximated population density for $u(t, x; P)$ in (2.3) is now given by

$$u(t, x; \{p_{jk}\}) = \sum_{j=0}^{M_1} \sum_{k=0}^{M_2} v(t, x; g_{jk}(x; b_j, \gamma_k))p_{jk},$$

where $v(t, x; g_{jk}(x; b_j, \gamma_k))$ is the subpopulation density from (2.1) with growth rate $g_{jk} = b_j(\gamma_k - x)$ and p_{jk} is the corresponding probability that an individual has growth rate g_{jk}. We will use the notation DEL(M_1, M_2) to denote the delta function approximation method in this case, where $M_1 + 1$ is the number of intrinsic growth rates b_j and $M_2 + 1$ is the number of maximum sizes γ_k used in the approximation.

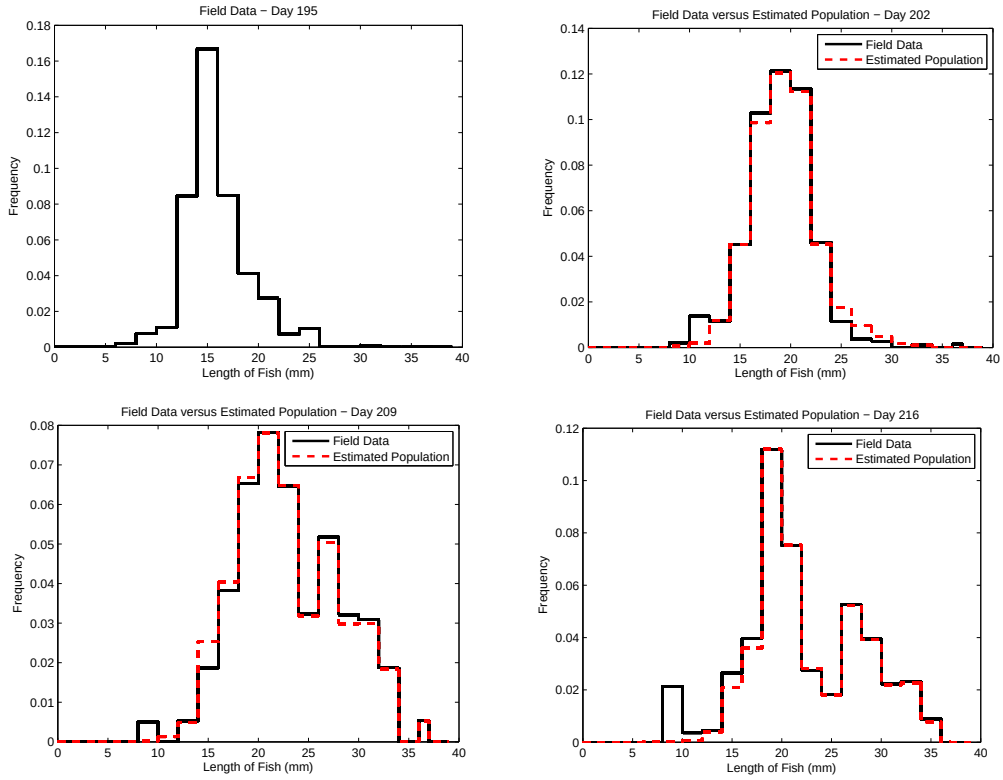

Figure 4.13: Field data versus estimated population for DEL(31,15).

The addition of a second parameter in the growth rate g of the mosquitofish does not change the least squares inverse problem

$$J(P) = \sum_{i,j} |u(t_i, x_j; P) - \hat{u}_{ij}|^2 = \sum_{i,j} \{u(t_i, x_j; P)^2 - 2u(t_i, x_j; P)\hat{u}_{ij} + (\hat{u}_{ij})^2\}$$

that we want to minimize over $P \in \mathcal{P}^{M_1 \times M_2}(\mathcal{G})$ where $\hat{u}_{ij}$ is the data and $\mathcal{P}^{M_1 \times M_2}(\mathcal{G})$ is the finite-dimensional approximation to $\mathcal{P}(\mathcal{G})$, which simplifies to a quadratic programming problem for an appropriately defined

$$F(\mathbf{p}) \equiv \mathbf{p}^T \mathcal{A}\mathbf{p} + 2\mathbf{p}^T \mathbf{b} + c.$$

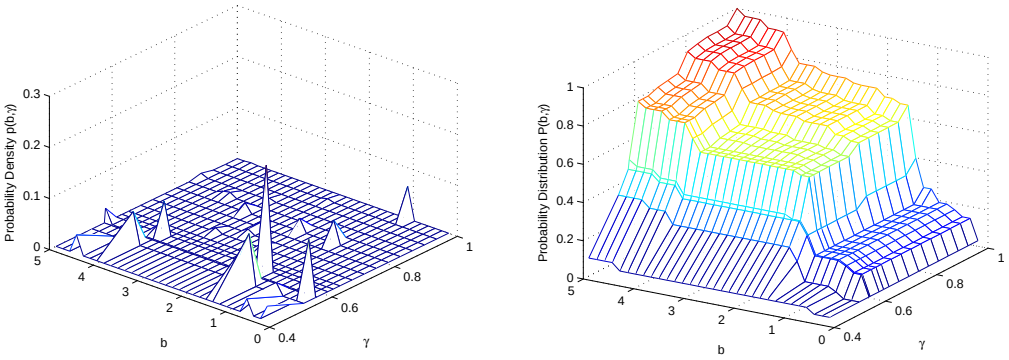

Figure 4.14: (a) Estimated probability density and (b) estimated probability distribution for DEL(31,15).

In Figure 4.13, we have the results from the inverse problem using DEL(31,15). These results were obtained in 514.3600 seconds, and the corresponding residual $J = 8.3169 \times 10^{-4}$. We see from the results shown in Figure 4.13 that the estimated population density from DEL(31,15) gives a much better fit to the field data in comparison to the optimal results obtained from PAR(7,35,35). The two key features of the data, dispersion and bifurcation, are both exhibited in the estimated population. The corresponding estimated probability density and distribution are shown in Figure 4.14. While no useful information can be obtained from Figure 4.14(a), the estimated probability distribution in Figure 4.14(b) appears to be bi-Gaussian in b for a fixed γ as well as in γ for a fixed b.

4.2.4 Model Fits with SPL(M_1, N_1, M_2, N_2)

The spline based approximation method, SPL(M, N), was also introduced earlier for the one parameter family of growth rates. For the two parameter family of growth rates that we are now considering, the approximated population density $u(t, x; P)$ to (2.3) is given by

$$u(t, x; \{a_{jk}\}) = \sum_{j=0}^{M_1} \sum_{k=0}^{M_2} a_{jk} \int_{\mathfrak{B}} \int_{\Gamma} v(t, x; g(x; b, \gamma)) l_j(b) l_k(\gamma) d\gamma db,$$

where $g(x; b, \gamma) = b(\gamma - x)$ and $p_{jk}(b, \gamma) = a_{jk} l_j(b) l_k(\gamma)$ is the probability density for individuals in population subgroup jk with l_j and l_k representing piecewise linear spline functions. We note that $\mathfrak{B} = [0.2, 5]$ and $\Gamma = [\frac{16}{38}, 1]$ for this method as well. The notation that we employ here is SPL(M_1, N_1, M_2, N_2), where $M_1 + 1$ and $M_2 + 1$ are the number of basis

elements used to approximate the growth rate probability distribution with respect to b and γ, respectively, and N_1 and N_2 represent the number of quadrature nodes used in the composite trapezoidal rule [31] for double integrals to approximate the integral in the expression above with respect to b and γ.

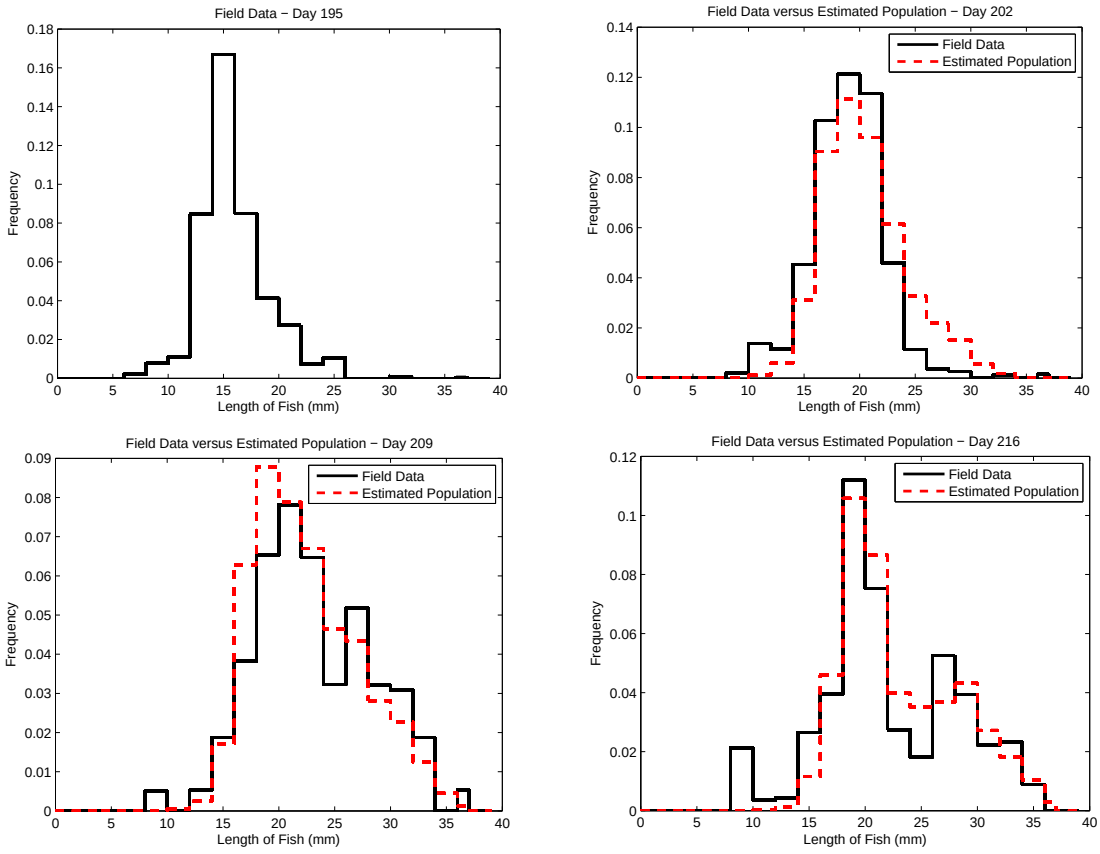

Figure 4.15: Field data versus estimated population for SPL(4,35,8,35).

The best results obtained using SPL(M_1, N_1, M_2, N_2) for the estimation of the growth rate distribution of the field data are shown in Figure 4.15 for SPL(4,35,8,35). We note that the results obtained using SPL(4,35,4,35) and SPL(4,35,6,35) were approximately the same as those obtained using SPL(4,35,8,35) for Day 202 and Day 209. However, the estimated population data obtained using SPL(4,35,8,35) gave a much better fit to the data for Day 216 than the results obtained using SPL(4,35,4,35) and SPL(4,35,6,35). The corresponding residual, J, for SPL(4,35,8,35) is 0.0054, and these results were obtained in 1.8822×10^3 seconds, or approximately 32 minutes. The estimated probability density

and distribution are shown in Figure 4.16. The resulting estimated probability distribution appears to be bi-Gaussian in γ.

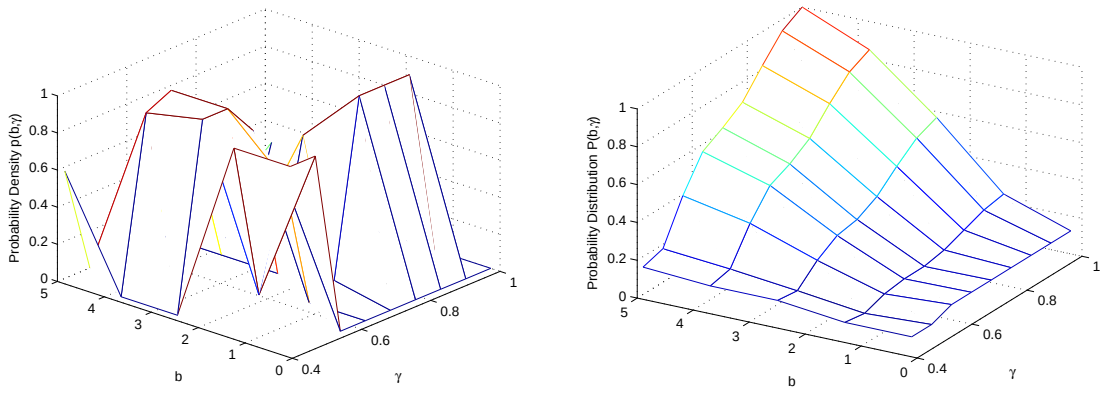

Figure 4.16: (a) Estimated probability density and (b) estimated probability distribution for SPL(4,35,8,35).

In comparison to the estimated population data produced by the estimated growth rate distribution from DEL(31,15), the estimated population data produced by the estimated growth rate distribution from SPL(4,35,8,35) does not give as good a fit to the field data, as is seen in Figure 4.15. The delta function approximation method does a better job of fitting the given field data in a more efficient way in comparison to the spline based approximation method. The optimal cost from SPL(4,35,8,35) is comparable to that obtained from PAR(7,35,35). We see that the spline based approximation method does a better job of estimating the frequencies for the smaller size classes and the peaks in the field data than the parametric version. While the results from the spline based approximation method and the standard parametric approach are very similar, the computational time required by PAR(7,35,35) is much lower than the computational time required by SPL(4,35,8,35).

4.2.5 Confidence Intervals for Parameter Estimates from PAR(M, N_1, N_2) and SPL(M_1, N_1, M_2, N_2)

As we did previously, we would like to also present here some results on the uncertainty associated with the estimated growth rate distributions determined by the inverse problem. The treatment in this section with the field data is very similar to the treatment previously carried out with the simulated data. With respect to the optimal results given

in this section, we were only able to perform a statistical analysis for PAR(7,35,35) and SPL(4,35,8,35). We are unable to perform this analysis for DEL(31,15) because the field data consists of 60 data points and the number of parameters determined by DEL(31,15) is 512; thus, the analysis in this case is invalid. Since we have explained in detail the underlying statistical model that we are considering, we will omit these details here and define the functions and variables that are used with respect to PAR(M, N_1, N_2) and SPL(M_1, N_1, M_2, N_2). First, we remark that $\{\tau_i\}_{i=1}^n$ corresponds to $(t_l, x_m), l = 1, \ldots, 3, m = 1, \ldots, 20$ pairs since, as previously noted, the field data that we use in the inverse problem consists of three days and twenty size classes (hence $n = 60$), $\theta = (\bar{b}_1, \sigma_{b_1}^2, \bar{b}_2, \sigma_{b_2}^2, \bar{\gamma}_1, \sigma_{\gamma_1}^2, \bar{\gamma}_2, \sigma_{\gamma_2}^2)$ when using PAR(M, N_1, N_2), and $\theta = \{a_{jk}\}_{j,k=0}^{j=M_1,k=M_2}$ when using SPL(M_1, N_1, M_2, N_2). We also note when considering PAR(M, N_1, N_2),

$$f(\tau_i; \theta_0) \approx \int_{\mathfrak{B}} \int_{\Gamma} v(\tau_i; g) p(b, \gamma; \theta) d\gamma db,$$

where $\theta_0 \approx \theta = \left(\bar{b}_1, \sigma_{b_1}^2, \bar{b}_2, \sigma_{b_2}^2, \bar{\gamma}_1, \sigma_{\gamma_1}^2, \bar{\gamma}_2, \sigma_{\gamma_2}^2\right)$. When considering SPL$(M_1, N_1, M_2, N_2)$,

$$f(\tau_i; \theta_0) \approx \sum_{j=0}^{M_1} \sum_{k=0}^{M_2} a_{jk} \int_{\mathfrak{B}} \int_{\Gamma} v(\tau_i; g) l_j(b) l_k(\gamma) d\gamma db$$

where $\theta_0 \approx \theta = \{a_{jk}\}$. We will define M from our previous outline to be one less than the number of parameters in θ for PAR(M, N_1, N_2) and $(M_1 + 1) \cdot (M_2 + 1) - 1$ for SPL(M_1, N_1, M_2, N_2). For the standard parametric version PAR(M, N_1, N_2), the entries in the $n \times (M + 1)$ sensitivity matrix $\mathcal{X}(\theta)$ are given by

$$\mathcal{X}_{im}(\theta) = \frac{\partial f(\tau_i; \theta)}{\partial \theta_m} = \int_{\mathfrak{B}} \int_{\Gamma} v(\tau_i; g) \frac{\partial p(b, \gamma; \theta)}{\partial \theta_m} d\gamma db,$$

and we can analytically compute $\frac{\partial p(b,\gamma;\theta)}{\partial \theta_m}, m = 0, \ldots, M$, since $p(b, \gamma; \theta)$ is the known bi-Gaussian pdf. The entries in the sensitivity matrix $\mathcal{X}(\theta)$ for SPL(M_1, N_1, M_2, N_2) are again straightforward and are given by

$$\mathcal{X}_{im}(\theta) = \frac{\partial f(\tau_i; \theta)}{\partial \theta_m} = \int_{\mathfrak{B}} \int_{\Gamma} v(\tau_i; g) l_j(b) l_k(\gamma) d\gamma db$$

for $m = j \cdot (M_2 + 1) + k$, where $j = 0, \ldots, M_1$ and $k = 0, \ldots, M_2$. Using these facts, we were able to estimate the covariance matrix Σ which is used in determining the standard errors and confidence intervals for $\hat{\theta}$. As we stated before, the endpoints of the confidence intervals are given by

$$\hat{\theta} \pm t_{1-\alpha/2} SE(\hat{\theta}),$$

where $t_{1-\alpha/2}$ is a distribution value given in the Student's t-distribution table (determined by the level of significance chosen [32]). We also chose to use a significance level of $\alpha = 0.05$ corresponding to the 95% confidence intervals computed in this section. From the statistical table for the t-distribution, $\alpha = 0.05$ corresponds to a $t_{1-\alpha/2}$ value of approximately 1.96 when the number of observations or data points is large, i.e., $n \geq 30$, which is true in this case.

In Table 4.7, the optimal parameter values for θ with the corresponding confidence intervals are given for PAR(7,35,35). We see that the confidence intervals computed with respect to $\bar{b}_1, \sigma^2_{b_1}, \bar{b}_2$, and $\sigma^2_{b_2}$, i.e., the parameters associated with the intrinsic growth rate b, are very large in comparison to the optimal estimates obtained for these parameters. Therefore, we cannot be certain when reporting values for these particular parameters when using PAR(7,35,35). However, we notice that the confidence intervals obtained for $\bar{\gamma}_1, \sigma^2_{\gamma_1}, \bar{\gamma}_2$, and $\sigma^2_{\gamma_2}$, i.e., those associated with the maximum size γ, are very small in comparison to the optimal values obtained using this method. Based on this analysis, we feel more certain about the reliability of the estimates obtained for the parameters associated with the maximum size γ because of the much smaller confidence intervals computed for these parameters.

Table 4.7: Estimated parameters for bi-Gaussian $p(b, \gamma)$ and confidence intervals given field data for PAR(7,35,35).

Parameter	$\hat{\theta}_m \pm 1.96SE(\hat{\theta}_m)$
$\bar{b}_1$	1.9749 ± 395.3664
$\sigma^2_{b_1}$	$0.0388 \times 10^{-3} \pm 3.1450$
$\bar{b}_2$	3.9132 ± 139.7600
$\sigma^2_{b_2}$	$0.2228 \times 10^{-3} \pm 1.5226$
$\bar{\gamma}_1$	$0.5265 \pm 0.8471 \times 10^{-4}$
$\sigma^2_{\gamma_1}$	$0.0122 \pm 0.2982 \times 10^{-4}$
$\bar{\gamma}_2$	$0.7208 \pm 0.1426 \times 10^{-3}$
$\sigma^2_{\gamma_2}$	$0.0372 \pm 0.8201 \times 10^{-4}$

For SPL(4,35,8,35), the optimal values for $\{a_{jk}\}$, for $j = 0, 1, \ldots, 4$ and $k = 0, 1, \ldots, 8$, and the corresponding confidence intervals are given in Table 4.8. As clearly seen by the values given in this table, the confidence intervals are very large (by a factor of $\approx 10^5$) in comparison to the estimated values for each component of $\hat{\theta}$. We are not very confident about estimated growth rate distributions obtained using SPL(4,35,8,35) in the inverse problem with this given data set. Moreover, we found that as the quantity $(M_1+1)\cdot(M_2+1)$

became larger (while still remaining smaller than n), $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ became nearly singular, resulting in a very ill-conditioned covariance matrix $\hat{\Sigma}$. This, in turn, resulted in larger confidence intervals which suggested that the estimated growth rate distributions produced by the spline based approximation method were even more unreliable with respect to the given field data. Furthermore, we were unable to compute confidence intervals for the (M_1, M_2) pairs $(4, 10)$ and $(10, 4)$ because the covariance matrix $\hat{\Sigma}$ did not exist in these cases ($\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ was singular to machine accuracy).

Table 4.8: Estimated probability density values $p(b, \gamma)$ and confidence intervals given field data for SPL(4,35,8,35).

a_{jk}	$\hat{\theta}_m \pm 1.96SE(\hat{\theta}_m)$	a_{jk}	$\hat{\theta}_m \pm 1.96SE(\hat{\theta}_m)$	a_{jk}	$\hat{\theta}_m \pm 1.96SE(\hat{\theta}_m)$
a_{00}	$1.0000 \pm 0.4245 \times 10^5$	a_{01}	$1.0000 \pm 0.6404 \times 10^5$	a_{02}	$0.0000 \pm 0.7277 \times 10^5$
a_{03}	$0.0000 \pm 1.0735 \times 10^5$	a_{04}	$0.0000 \pm 3.1710 \times 10^5$	a_{05}	$0.0000 \pm 7.1776 \times 10^5$
a_{06}	$0.0000 \pm 8.7419 \times 10^5$	a_{07}	$0.0000 \pm 3.0799 \times 10^5$	a_{08}	$0.0000 \pm 0.2847 \times 10^5$
a_{10}	$1.0000 \pm 0.2878 \times 10^5$	a_{11}	$0.1525 \pm 0.3884 \times 10^5$	a_{12}	$0.7181 \pm 0.0999 \times 10^5$
a_{13}	$0.0000 \pm 0.4497 \times 10^5$	a_{14}	$0.0620 \pm 0.1099 \times 10^5$	a_{15}	$1.0000 \pm 0.5300 \times 10^5$
a_{16}	$1.0000 \pm 0.1159 \times 10^5$	a_{17}	$1.0000 \pm 0.1826 \times 10^5$	a_{18}	$0.0000 \pm 0.0721 \times 10^5$
a_{20}	$0.0000 \pm 0.1106 \times 10^5$	a_{21}	$1.0000 \pm 0.5037 \times 10^5$	a_{22}	$1.0000 \pm 0.4524 \times 10^5$
a_{23}	$0.8275 \pm 0.2581 \times 10^5$	a_{24}	$0.0000 \pm 0.0555 \times 10^5$	a_{25}	$0.6864 \pm 0.1384 \times 10^5$
a_{26}	$0.0000 \pm 0.1537 \times 10^5$	a_{27}	$0.1591 \pm 0.0115 \times 10^5$	a_{28}	$0.0000 \pm 0.0340 \times 10^5$
a_{30}	$0.0000 \pm 0.3470 \times 10^5$	a_{31}	$0.9352 \pm 0.2464 \times 10^5$	a_{32}	$1.0000 \pm 0.3213 \times 10^5$
a_{33}	$0.6130 \pm 0.2102 \times 10^5$	a_{34}	$0.0000 \pm 0.0946 \times 10^5$	a_{35}	$0.0000 \pm 0.0683 \times 10^5$
a_{36}	$0.4924 \pm 0.0181 \times 10^5$	a_{37}	$0.2598 \pm 0.0278 \times 10^5$	a_{38}	$0.0000 \pm 0.0011 \times 10^5$
a_{40}	$0.5786 \pm 0.7387 \times 10^5$	a_{41}	$0.0000 \pm 0.1237 \times 10^5$	a_{42}	$0.0000 \pm 0.1336 \times 10^5$
a_{43}	$0.0000 \pm 0.2196 \times 10^5$	a_{44}	$0.0000 \pm 0.1207 \times 10^5$	a_{45}	$0.0000 \pm 0.1020 \times 10^5$
a_{46}	$0.0000 \pm 0.0372 \times 10^5$	a_{47}	$0.0000 \pm 0.0172 \times 10^5$	a_{48}	$0.1682 \pm 0.0005 \times 10^5$

Overall, we found that the estimated growth rate distributions using DEL(31,15) produced the best fit to the field data in comparison to both SPL(4,35,8,35) and PAR(7,35,35). The model fits obtained with PAR(7,35,35) were very much comparable to those obtained with SPL(4,35,8,35) in a more efficient manner in terms of computational time. From the statistical analysis carried out for the estimates obtained from SPL(4,35,8,35) and PAR(7,35,35), we observed that for this inverse problem the estimates obtained from the spline based approximation method are not very reliable (as a result of the very large confidence intervals). We are more certain about the parameters related to γ when using PAR(7,35,35) than the parameters related to b. Furthermore, based on the fit-to-data and required computational time, the delta function approximation provided the best estimates

of the growth rate distribution for the mosquitofish field data given in this example; however, as previously noted, we are unable to make any remarks about the reliability of these estimates.

Chapter 5

Quantifying Uncertainty in Estimated Probability Distributions

In the previous chapter, we demonstrated how to construct nodal confidence intervals for the finite-dimensional parameters (i.e., $\{p_k^M\}_{k=0}^M, \{a_k^M\}_{k=0}^M, \theta$) using the standard asymptotic theory for OLS estimators. The finite-dimensional parameters that we determined by solving the inverse problem were at the level of the probability density. As shown in Chapter 4, we constructed estimates of the parameter of interest in our original problem (the probability distribution) by using the estimates of the probability density obtained from the inverse problem. While we can use the standard error theory that has already been established to quantify the uncertainty associated with the estimates of the finite-dimensional parameters, we cannot apply this same theory to the estimated probability distributions which are in an infinite-dimensional setting. Since asymptotic standard error theory does not exist for problems with functional parameters, we would like to develop the mathematical and asymptotic statistical theory for OLS problems where the parameter of interest is a probability distribution. In this chapter we derive the mean and variance of the OLS estimator for the functional linear regression model and discuss the results of a Monte Carlo simulation study providing computational evidence of "true" functional confidence bands for infinite-dimensional parameters. We also outline how to construct confidence bands for estimated probability distributions by extending the nodal confidence intervals derived from

the standard error theory for finite-dimensional OLS problems. Computational results with simulated data with PAR(M, N), DEL(M), and SPL(M, N) are presented illustrating the concept of confidence bands.

5.1 Function Space Analogue of Asymptotic Standard Error Theory

5.1.1 Least Squares Estimation for Functional Linear Regression Models

We consider the following statistical model for the scalar observation process Y_j

$$Y(\tau_j) = \int_0^T k(\tau_j, s)\beta_0(s)ds + \epsilon_j, \quad j = 1, \ldots, n, \tag{5.1}$$

where the "true" parameter $\beta_0(s)$ is a scalar function in $L^2(0, T)$ (assumed to exist) and $k(\tau_j, s)$ is also a scalar function in $L^2(0, T)$. We also assume that the ϵ_j are *i.i.d.* with $\mathcal{E}[\epsilon_j] = 0$ and $Var[\epsilon_j] = \sigma_0^2 > 0$. We note that (5.1) can be rewritten as

$$Y_j = \int_0^T k_j(s)\beta_0(s)ds + \epsilon_j, \quad j = 1, \ldots, n.$$

Let $X : L^2(0, T) \to \mathbb{R}^n$ be defined as

$$X\beta = \int_0^T k(s)\beta(s)ds, \tag{5.2}$$

where $k(s)$ is a $n \times 1$ vector of scalar functions. Then the statistical model in (5.1) for the random variable Y can be expressed as

$$Y = X\beta_0 + \epsilon, \tag{5.3}$$

where $\epsilon \in \mathbb{R}^n$ is a random variable with $\mathcal{E}[\epsilon] = 0$ and $Var[\epsilon] = \sigma_0^2 I_n$ (I_n represents the $n \times n$ identity matrix). Since Y is also a random variable, the expected value of Y is $X\beta_0$ and the variance of Y is $\sigma_0^2 I_n$. As is the case in finite dimensions, β_0 and σ_0^2 are usually unknown. Therefore, we would like to construct an ordinary least squares (OLS) estimator B of β_0 in order to compute an estimate $\hat{\beta}$ for a realization (data set) $\{y_j\}$ that can be used in approximating the sampling distribution of the OLS estimator B which is also a random variable. As already noted in the Introduction, the sampling distribution is of importance because of its role in quantifying the uncertainty associated with estimates $\hat{\beta}$.

The least squares problem that we consider is formulated as

$$J(\beta) = (Y - X\beta)^T(Y - X\beta), \tag{5.4}$$

which we seek to minimize for $\beta \in L^2(0,T)$. We can determine a solution to (5.4) by taking the derivative of $J(\beta)$ with respect to β and setting it equal to zero. The derivative of J with respect to β is actually the directional derivative of J at β in the direction φ, where $\varphi \in L^2(0,T)$ [12]. Thus, for all $\varphi \in L^2(0,T)$,

$$\begin{aligned} 0 &= \frac{\partial J(\beta)}{\partial \beta}[\varphi] \\ &= (Y - X\beta)^T \frac{\partial (Y - X\beta)}{\partial \beta}[\varphi] + \frac{\partial (Y - X\beta)^T}{\partial \beta}[\varphi](Y - X\beta). \end{aligned}$$

Before carrying out these computations, we derive the following useful relations.

First, we compute the adjoint X^* which is a mapping from $\mathbb{R}^n$ to $L^2(0,T)$. The adjoint operator is defined by the equation

$$\langle X\beta, \xi \rangle_{\mathbb{R}^n} = \langle \beta, X^*\xi \rangle_{L^2(0,T)}, \qquad \text{where } \xi \in \mathbb{R}^n.$$

Therefore,

$$\begin{aligned} \langle X\beta, \xi \rangle_{\mathbb{R}^n} &= \xi \bullet X\beta \\ &= \xi^T X\beta \\ &= \xi^T \int_0^T k(s)\beta(s)ds \\ &= \int_0^T \xi^T k(s)\beta(s)ds \\ &= \int_0^T \beta(s)k^T(s)\xi \, ds \\ &= \langle \beta, X^*\xi \rangle_{L^2(0,T)}, \end{aligned}$$

which implies

$$X^*\xi = k^T(s)\xi.$$

As a result of (5.2), we note

$$(X\beta)^T = \int_0^T \beta(s)k^T(s)ds.$$

Using this relation, we have the following for $\varphi \in L^2(0,T)$

$$\frac{\partial (X\beta)^T}{\partial \beta}[\varphi] = \int_0^T \varphi(s)k^T(s) = (X\varphi)^T.$$

Thus,

$$\frac{\partial(X\beta)^T}{\partial\beta}[\varphi](Y - X\beta) = (X\varphi)^T(Y - X\beta) = \langle X\varphi, Y - X\beta\rangle_{\mathbb{R}^n}.$$

We now return to the problem of solving $\frac{\partial}{\partial\beta}J(\beta)[\varphi] = 0$ for all $\varphi \in L^2(0,T)$. For all $\varphi \in L^2(0,T)$, we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial\beta}J(\beta)[\varphi] \\
&= (Y - X\beta)^T\frac{\partial(Y - X\beta)}{\partial\beta}[\varphi] + \frac{\partial(Y - X\beta)^T}{\partial\beta}[\varphi](Y - X\beta) \\
&= -(Y - X\beta)^T\frac{\partial(X\beta)}{\partial\beta}[\varphi] - \frac{\partial(X\beta)^T}{\partial\beta}[\varphi](Y - X\beta) \\
&= -\langle Y - X\beta, X\varphi\rangle_{\mathbb{R}^n} - \langle X\varphi, Y - X\beta\rangle_{\mathbb{R}^n} \\
&= -\langle X\varphi, Y - X\beta\rangle_{\mathbb{R}^n} - \langle X\varphi, Y - X\beta\rangle_{\mathbb{R}^n} \\
&= -2\langle X\varphi, Y - X\beta\rangle_{\mathbb{R}^n} \\
&= -2\langle \varphi, X^*(Y - X\beta)\rangle_{L^2(0,T)}.
\end{aligned}$$

Since

$$-2\langle \varphi, X^*(Y - X\beta)\rangle_{L^2(0,T)} = 0 \quad \forall\varphi \in L^2(0,T),$$

then

$$X^*(Y - X\beta) = 0,$$

which gives us the normal equations

$$X^*X\beta = X^*Y. \tag{5.5}$$

If X^*X is invertible, then the least squares estimator is given by

$$B = (X^*X)^{-1}X^*Y, \tag{5.6}$$

and the least squares estimate $\hat{\beta}$ for a particular realization y is

$$\hat{\beta} = (X^*X)^{-1}X^*y.$$

We note that this is the function space analogue of the normal equations for finite-dimensional parameters derived in [59].

When $X\beta = Y$ has more than one solution, we can use Theorem 1 of Section 6.1 in Luenberger [51] to determine the solution of the least squares problem. The theorem is stated as follows:

Theorem 5.1.1 *Let G and H be Hilbert spaces and let $A \in \mathcal{B}(G, H)$ with range closed in H. Then the vector x of minimum norm satisfying $Ax = y$ is given by $x = A^*z$ where z is any solution of $AA^*z = y$.*

Since $X \in \mathcal{B}(L^2(0,T), \mathbb{R}^n)$, the range of X is closed because it is a subspace of the finite-dimensional space $\mathbb{R}^n$. Hence, by the theorem given above, the solution to our least squares problem is given by

$$B = X^*Z,$$

where Z is any solution of

$$XX^*Z = Y.$$

If XX^* is invertible, then the least squares estimator can be expressed as

$$B = X^*(XX^*)^{-1}Y \tag{5.7}$$

with the least squares estimate for a specific data set y

$$\hat{\beta} = X^*(XX^*)^{-1}y.$$

We can rewrite the solutions in (5.6) and (5.7) in terms of the pseudoinverse $X^\dagger$ of X. In the first case, if X^*X is invertible, we note the following property in [51]

$$X^\dagger = (X^*X)^{-1}X^*,$$

so the solution to the least squares problem is

$$B = (X^*X)^{-1}X^*Y = X^\dagger Y.$$

Likewise, if XX^* is invertible, then $X^\dagger = X^*(XX^*)^{-1}$ (see [51]), so the solution to the least squares problem is

$$B = X^*(XX^*)^{-1}Y = X^\dagger Y.$$

Lastly, we show that the expected value of the estimator in (5.6) and (5.7) is equal to the "true" parameter β_0. When X^*X is invertible, then

$$\begin{aligned} \mathcal{E}[B] &= \mathcal{E}[(X^*X)^{-1}X^*Y] \\ &= (X^*X)^{-1}X^*X\beta_0 \quad \text{since } \mathcal{E}[Y] = X\beta_0 \\ &= \beta_0. \end{aligned}$$

By a similar computation when XX^* is invertible, we see that

$$\begin{aligned}\mathcal{E}[B] &= \mathcal{E}[X^*(XX^*)^{-1}Y] \\ &= X^*(XX^*)^{-1}X\beta_0 \quad \text{since } \mathcal{E}[Y] = X\beta_0 \\ &= X^*(X^*)^{-1}X^{-1}X\beta_0 \\ &= \beta_0.\end{aligned}$$

Therefore, in both cases B is an unbiased estimator of β_0.

The two cases examined above are special cases of the more general problem of finding a solution to $Y = X\beta$. We now show that the estimator $B = X^\dagger Y$, where $X^\dagger$ is the pseudoinverse of X, is an unbiased estimator of β_0 in general. The following definition given in [51] will be useful for the general case where G and H are Hilbert spaces and $A \in \mathcal{B}(G, H)$ with $\mathcal{R}(A)$ closed.

Definition 5.1.1 *Among all vectors $x_1 \in G$ satisfying*

$$\|Ax_1 - y\| = \min_x \|Ax - y\|,$$

*let x_0 be the unique vector of minimum norm. The **pseudoinverse** $A^\dagger$ of A is the operator mapping y into its corresponding x_0 as y varies over H.*

Using this definition with $X \in \mathcal{B}(L^2(0,T), \mathbb{R}^n)$ and again noting that $\mathcal{R}(X)$ is closed since it is a subspace of $\mathbb{R}^n$, B is the unique function of minimum norm among all $\beta_1 \in L^2(0,T)$ of

$$\|X\beta_1 - Y\| = \min_{\beta \in L^2(0,T)} \|X\beta - Y\|.$$

Furthermore, the pseudoinverse operator $X^\dagger$ maps Y into its corresponding B as Y varies over $\mathbb{R}^n$.

We use the following theorem from Section 3.4 in Luenberger [51] to determine the solution of $Y = X\beta$.

Theorem 5.1.2 *If M is a closed linear subspace of a Hilbert space H, then $H = M \oplus M^\perp$ and $M = M^{\perp\perp}$.*

As a result of this theorem, we can write $L^2(0,T)$ as the direct sum of the nullspace of X and the orthogonal complement of the nullspace of X

$$L^2(0,T) = \mathcal{N}(X) \oplus \mathcal{N}(X)^\perp,$$

and $\mathbb{R}^n$ can be written as the direct sum of the range of X and the orthogonal complement of the range of X

$$\mathbb{R}^n = \mathcal{R}(X) \oplus \mathcal{R}(X)^{\perp}.$$

The pseudoinverse $X^{\dagger}$ is defined as the operator that maps elements from $\mathcal{R}(X)$ into $\mathcal{N}(X)^{\perp}$ and elements from $\mathcal{R}(X)^{\perp}$ to the null vector θ. Therefore, $X^{\dagger}$ is the inverse operator of X restricted to $\mathcal{N}(X)^{\perp}$.

By Theorem 5.1.2 we can write any $Y \in \mathbb{R}^n$ as

$$Y = \hat{Y} + Y_p, \quad \text{where } \hat{Y} \in \mathcal{R}(X) \text{ and } Y_p \in \mathcal{R}(X)^{\perp}. \tag{5.8}$$

Thus,

$$X^{\dagger}Y = X^{\dagger}(\hat{Y} + Y_p) = X^{\dagger}\hat{Y},$$

so $\hat{Y}$ is the best approximation to $Y \in \mathcal{R}(X)$. We let

$$B = X^{\dagger}Y.$$

Then, using (5.8) we see that

$$B = X^{\dagger}Y = X^{\dagger}(\hat{Y} + Y_p) = X^{\dagger}\hat{Y}.$$

Since $\hat{Y} \in \mathcal{R}(X)$, then $X^{\dagger}$ maps $\hat{Y}$ into its corresponding B by Definition 5.1.1 and

$$XB = \hat{Y}.$$

Moreover, we note that $B \in \mathcal{N}(X)^{\perp}$ and is the minimum norm solution of $X\beta_1 = \hat{Y}$.

We now show that B is an unbiased estimator of β_0; that is, the expected value of B is equal to β_0. We assume that $\mathcal{E}[\hat{Y}] = X\beta_0$, where $\hat{Y} \in \mathcal{R}(X)$. Then,

$$\begin{aligned}
\mathcal{E}[B] &= \mathcal{E}[X^{\dagger}Y] \\
&= \mathcal{E}[X^{\dagger}\hat{Y}] \quad \text{since } B = X^{\dagger}Y = X^{\dagger}\hat{Y} \\
&= X^{\dagger}X\beta_0 \quad \text{since } \mathcal{E}[\hat{Y}] = X\beta_0 \\
&= \beta_0,
\end{aligned}$$

since $X^{\dagger}$ is the inverse operator of X restricted to $\mathcal{N}(X)^{\perp}$.

We must also compute the variance of the estimator B in order to specify its sampling distribution. We start with the general case where

$$B = X^{\dagger}Y.$$

Recalling the fact that

$$Var(\epsilon) = \sigma_0^2 I_n,$$

we have

$$\begin{aligned}
Var[B] &= \mathcal{E}[(B-\beta_0)^2] \\
&= \mathcal{E}[(\beta_0 + X^\dagger\epsilon - \beta_0)^2] \quad \text{since } B = X^\dagger Y = X^\dagger(X\beta_0+\epsilon) = \beta_0 + X^\dagger\epsilon \\
&= \mathcal{E}[(X^\dagger\epsilon)^2] \\
&= \mathcal{E}[(X^\dagger\epsilon)(X^\dagger\epsilon)^T] \\
&= \mathcal{E}[X^\dagger\epsilon\epsilon^T(X^\dagger)^*] \\
&= X^\dagger\mathcal{E}[\epsilon\epsilon^T](X^\dagger)^* \\
&= \sigma_0^2 X^\dagger(X^\dagger)^* \quad \text{since } \mathcal{E}[\epsilon\epsilon^T] = Var[\epsilon] = \sigma_0^2 I_n.
\end{aligned}$$

Now, considering the special case when X^*X is invertible, we recall that $X^\dagger = (X^*X)^{-1}X^*$, so

$$\begin{aligned}
Var[B] &= \sigma_0^2 X^\dagger(X^\dagger)^* \\
&= \sigma_0^2 (X^*X)^{-1}X^*((X^*X)^{-1}X^*)^* \\
&= \sigma_0^2 (X^*X)^{-1}X^*X(X^*X)^{-1} \\
&= \sigma_0^2 (X^*X)^{-1}.
\end{aligned}$$

Likewise, we note in the special case when XX^* is invertible, then $X^\dagger = X^*(XX^*)^{-1}$ and

$$\begin{aligned}
Var[B] &= \sigma_0^2 X^\dagger(X^\dagger)^* \\
&= \sigma_0^2 X^*(XX^*)^{-1}(X^*(XX^*)^{-1})^* \\
&= \sigma_0^2 X^*(XX^*)^{-1}(XX^*)^{-1}X \\
&= \sigma_0^2 X^*(X^*)^{-1}X^{-1}(X^*)^{-1}X^{-1}X \\
&= \sigma_0^2 X^{-1}(X^*)^{-1} \\
&= \sigma_0^2 (X^*X)^{-1}.
\end{aligned}$$

The variance of the estimator B depends on the "true" variance σ_0^2 which is generally not known. Therefore, we will need to construct an estimator for the variance σ_0^2. The estimator in the finite-dimensional case depends on the trace operator and the dimension of the parameter to be estimated; however, since the parameter to be estimated in this framework is infinite-dimensional, our computations will require the trace operator in functional

space. Furthermore, we need to make use of the Central Limit Theorem in functional space in order to approximate the sampling distribution of the estimator B based on a realization of Y.

5.1.2 Construction of Confidence Bands from Confidence Intervals

To construct confidence bands for the estimated probability distributions, we use the confidence intervals obtained for the finite-dimensional parameters. We will first discuss how we construct confidence bands when using the standard parametric approach PAR(M, N). When using PAR(M, N), we use an a priori probability density in the GRD model (2.3), which we assume is continuous. After using the standard error theory to compute a confidence interval for θ, we construct a confidence band for the estimated probability distribution by using the endpoints of the confidence interval in the known probability density function (pdf). We note that the 95% confidence region for the estimated probability density is formed by plotting

$$p_- = p(b; \hat{\theta} - 1.96SE(\hat{\theta})) \quad \text{and} \quad p_+ = p(b; \hat{\theta} + 1.96SE(\hat{\theta})),$$

where $\hat{\theta}$ represents the estimates of θ that solve the OLS problem (3.2). Then, using the fact that the probability density function p also represents the derivative of the probability distribution function P, we construct the upper edge of the confidence band for the estimated probability distribution by using the portions of p_- and p_+ that lie above the estimated probability density when this function is increasing (i.e., the slope is positive). When the estimated probability density is decreasing and the slope is negative, the portions of p_- and p_+ that lie below the estimated probability density are used to construct the upper edge of the confidence band. We use this same technique to create the lower edge of the confidence band by using the portions of p_- and p_+ that lie below (above) the estimated probability density when the slope is positive (negative). We integrate over these values and then normalize by an appropriate factor so that the lower and upper edges forming the confidence band are "true" distributions (integrate to 1).

When using the non-parametric approaches, DEL(M) and SPL(M, N), the confidence intervals computed using the standard error theory correspond to the weights, $\{p_k^M\}_{k=0}^M$ and $\{a_k^M\}_{k=0}^M$, used in the approximations. In some cases, the lower endpoints of the confidence intervals for these estimated weights may be negative, which violates the non-negativity condition required of probability densities (see results for SPL(8,128) in

Figure 4.8). Thus, before constructing the confidence band for the estimated probability distribution, we first truncate any negative values to zero in order to have a "true" density. If the estimated probability density is monotone and increasing, the upper (lower) edge of the confidence band for the estimated distribution is constructed by integrating over the upper (lower) confidence interval endpoints and normalizing by an appropriate factor so that the upper (lower) edge of the confidence band is a "true" probability distribution. In the case that the estimated probability density is not monotone (which is the case in the examples considered here), the construction of the confidence bands using DEL(M) and SPL(M, N) again depends on the slope of the estimated probability densities. The technique employed in these cases mimics that described when using PAR(M, N). The upper (lower) edge of the confidence band is created by integrating over the upper (lower) confidence interval endpoints when the slope of the estimated probability density is positive and the lower (upper) confidence interval endpoints when the slope is negative. We again normalize by an appropriate factor so that both edges of the confidence band for the estimated probability distribution are also "true" distributions. We will illustrate these methods in two examples using simulated size-structured mosquitofish population data.

5.2 Computational Results with Gaussian Example

In the results presented in this section, we note that the data was generated using the same parameter values ($\mu_b = 4.5, \sigma_b^2 = 0.25$) given earlier with the known approximate truncated Gaussian probability distribution. The results in this section were all obtained with the same data set with 20% absolute noise. A larger percentage of absolute noise was added to the simulated data for the results in this section so one could differentiate visually between the estimated probability distributions and the confidence bands when using the parametric approach. Results with the confidence bands were also obtained with 10% absolute noise; however, the confidence bands were much tighter around the estimated probability distributions.

5.2.1 Results with PAR(M, N)

Using the method outlined above, we obtained the following results with PAR(1,128) from the inverse problem using the simulated data with 20% absolute noise. The optimal cost for this set of results is 14.0615, while the estimate of $\hat{\sigma}^2$ is 0.0056. We computed

the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$, which is used in calculating the standard errors for the finite-dimensional parameter θ, and obtained a value of 1.2833. Table 5.1 contains the optimal estimate of θ as well as the corresponding 95% confidence intervals. The plots of the known and estimated probability densities along with p_- and p_+ are shown on the left in Figure 5.1, and the plots of the known and estimated probability distributions and corresponding confidence band are shown on the right in Figure 5.1. We observe that the estimated probability distribution is indeed contained in the small area bounded by the confidence band constructed with the technique outlined above.

Table 5.1: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b^2)$ and confidence intervals for Gaussian example with 20% absolute error when using PAR(1,128).

$\hat{\mu}_b \pm 1.96SE(\hat{\mu}_b)$	$\hat{\sigma}_b^2 \pm 1.96SE(\hat{\sigma}_b^2)$
4.5062 ± 0.0163	0.2774 ± 0.0146

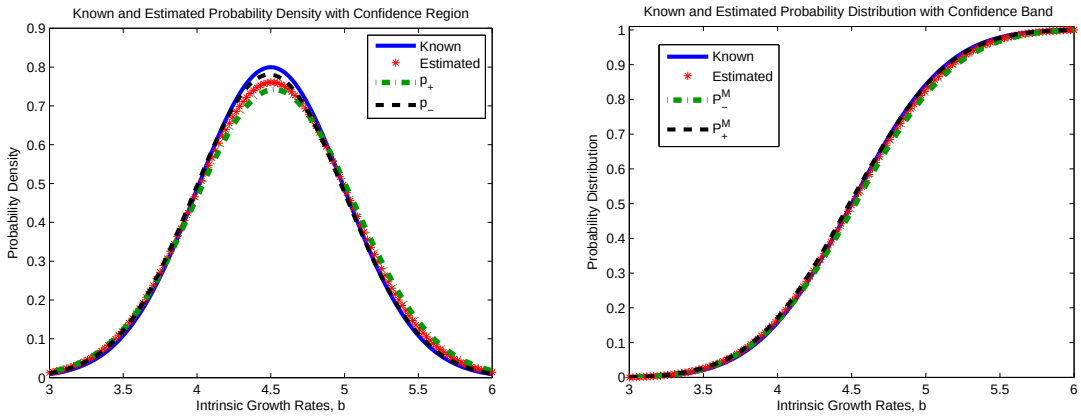

Figure 5.1: Estimated probability density with confidence region and probability distribution with confidence band for Gaussian example using PAR(1,128) with 20% absolute error.

5.2.2 Results with DEL(M)

The next set of results from the parameter estimation problem were obtained using DEL(M) for various values of M. In Table 5.2, the optimal cost $J(\hat{\theta})$, estimated variance $\hat{\sigma}^2$, and condition number $\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$ for M = 4, 8, 12, 16, 24, and 32 are given. The estimated probability densities with the corresponding 95% confidence intervals and the estimated probability distributions with the corresponding confidence bands for these values

of M are in Figures 5.2 through 5.7. We see that the estimated probability distribution converges to the true probability distribution as M is increased. The optimal cost $J(\hat{\theta})$ and the estimated variance $\hat{\sigma}^2$ decrease as the number of parameters M that we estimate increases. While the estimated probability distribution converges to the known distribution as M is increased, the confidence bands appear to converge nicely until M becomes too large and the problem becomes over-parametrized. The increase in the condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ that are shown in Table 5.2 is relatively smaller as M is increased from 4 to 24; however, there is a significant increase in $\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$ as M is increased from 24 to 32. Therefore, we note from the computational results obtained here that the confidence bands appear to be converging nicely while the number of parameters is reasonable.

Table 5.2: Optimal cost values, $\hat{\sigma}^2$, and condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for Gaussian example with 20% absolute error when using DEL(M).

M	$J(\hat{\theta})$	$\hat{\sigma}^2$	$\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$
4	67.8765	0.0272	2.6236
8	35.5805	0.0143	3.0244
12	23.0116	0.0093	3.7123
16	16.0959	0.0065	4.7307
24	13.8528	0.0056	10.3792
32	13.8392	0.0056	44.7902

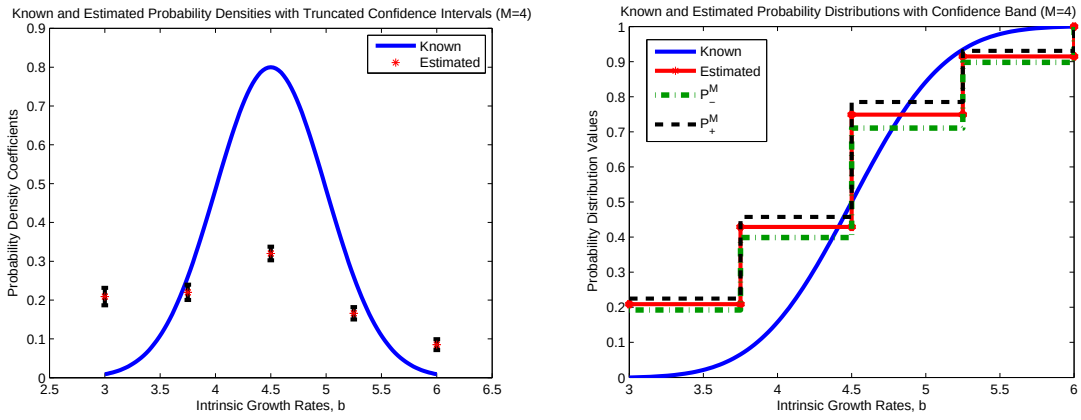

Figure 5.2: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using DEL(4) with 20% absolute error.

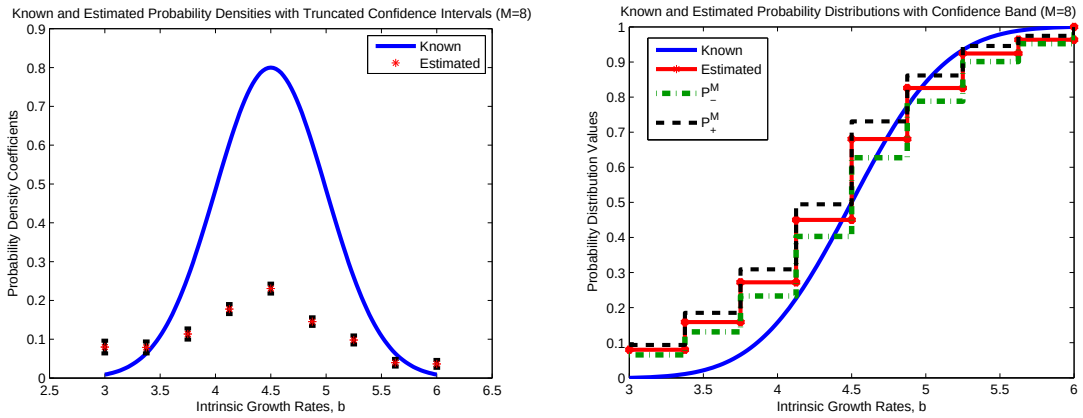

Figure 5.3: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using DEL(8) with 20% absolute error.

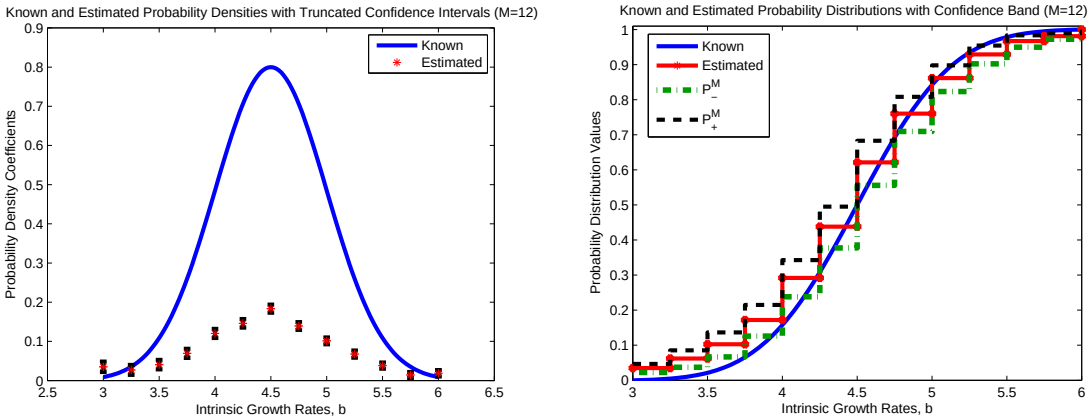

Figure 5.4: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using DEL(12) with 20% absolute error.

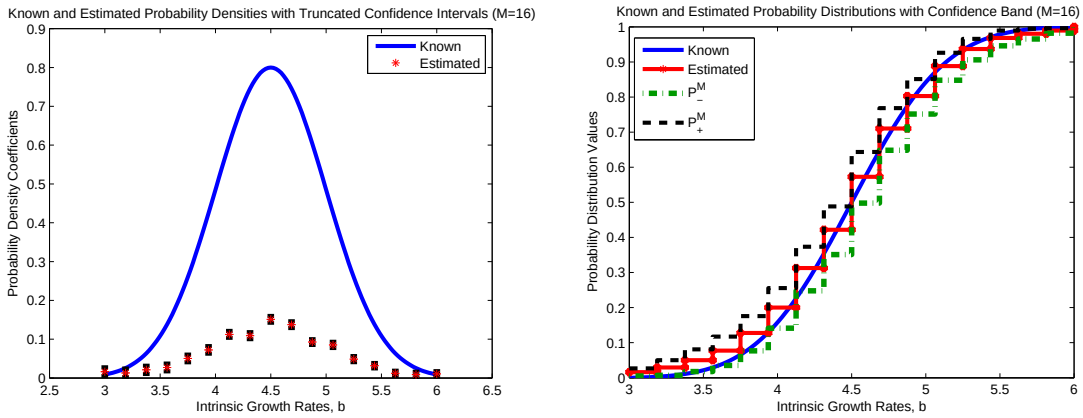

Figure 5.5: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using DEL(16) with 20% absolute error.

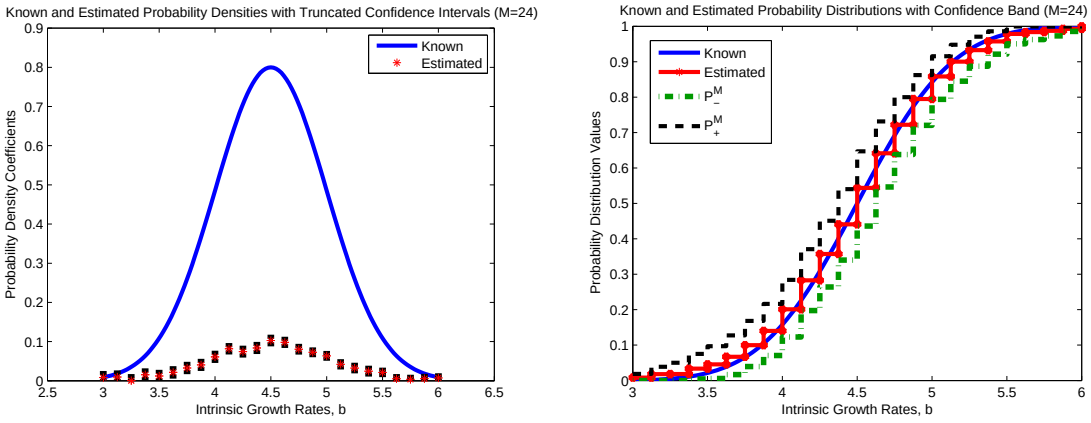

Figure 5.6: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using DEL(24) with 20% absolute error.

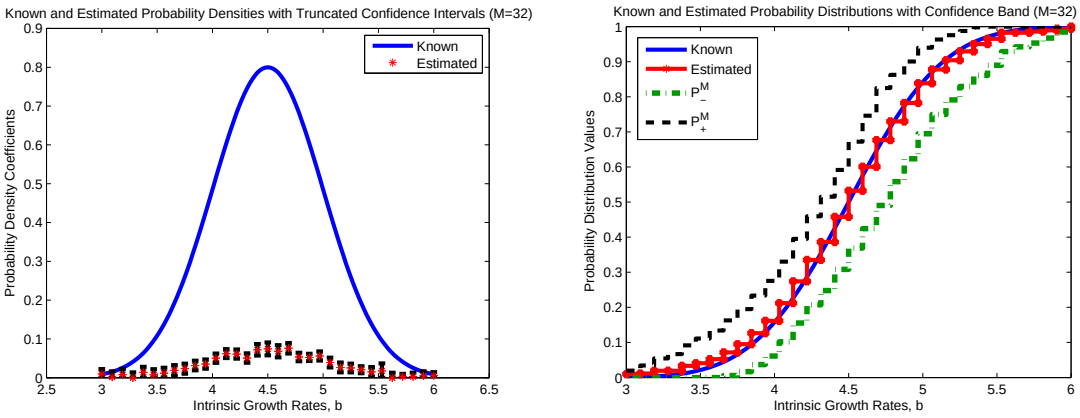

Figure 5.7: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using DEL(32) with 20% absolute error.

5.2.3 Results with SPL(M, N)

The final set of computational results for the Gaussian example were obtained using SPL$(M, 128)$ for $M = 4$, 8, 12, 16, 24, and 32. Since we added absolute noise instead of relative noise to the simulated data used in the inverse problem calculations shown here, the tails of the estimated probability density functions (pdfs) are very poor. We see in Table 5.3 the decreasing optimal cost values as the value of M is increased. The estimates of $\hat{\sigma}^2$ decrease when M is increased from 4 to 8 but have essentially converged as M is increased from 8 up to 32. Figures 5.8 through 5.13 display the estimated probability densities and distributions along with the 95% confidence intervals and confidence bands. We notice that the estimated probability distributions converge to the known probability distribution quickly (for much smaller values of M in comparison to DEL(M)). As the number of parameters is increased, the condition number κ of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ also increases. Moreover, there are very significant increases in $\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$ as M is increased from 16 to 24 and from 24 to 32. Again, we are able to use this behavior in understanding the confidence bands that are obtained for these values of M. The confidence bands initially appear to be converging nicely as M is increased. However, the confidence bands grow larger as M is increased too much and the problem becomes over-parametrized. This example illustrates that we are able to construct confidence bands for the estimated probability distributions using our scheme which appear to converge nicely for appropriately chosen values of M.

Table 5.3: Optimal cost values, $\hat{\sigma}^2$, and condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for Gaussian example with 20% absolute error when using SPL$(M, 128)$.

M	$J(\hat{\theta})$	$\hat{\sigma}^2$	$\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$
4	14.2128	0.0057	9.2919
8	13.9665	0.0056	16.0090
12	13.9244	0.0056	24.4291
16	13.8604	0.0056	37.4285
24	13.8385	0.0056	88.4467
32	13.8167	0.0056	215.7808

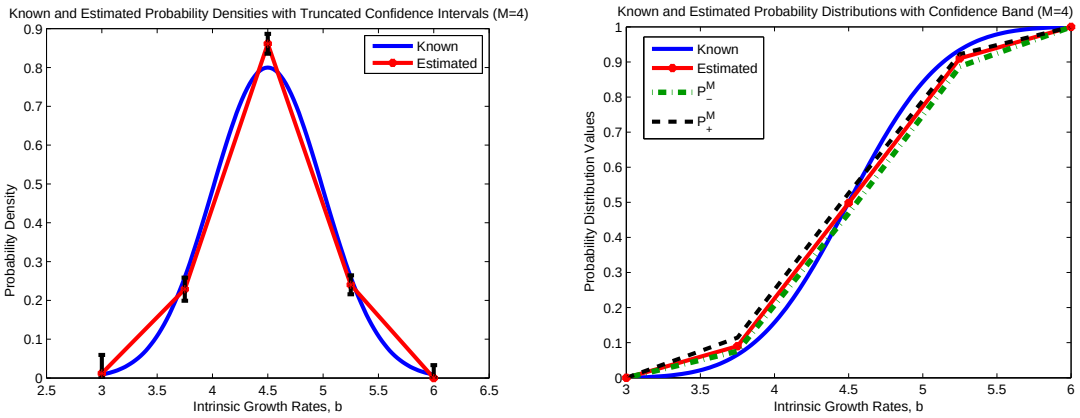

Figure 5.8: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using SPL(4,128) with 20% absolute error.

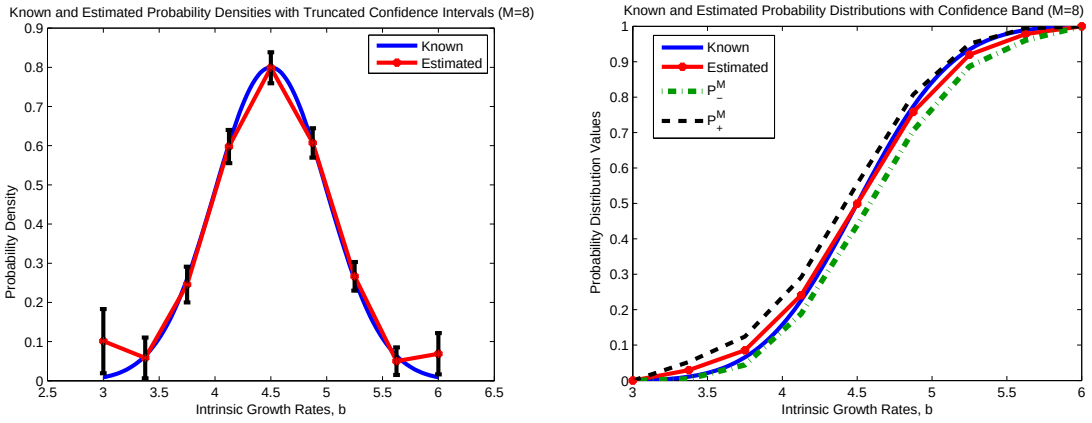

Figure 5.9: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using SPL(8,128) with 20% absolute error.

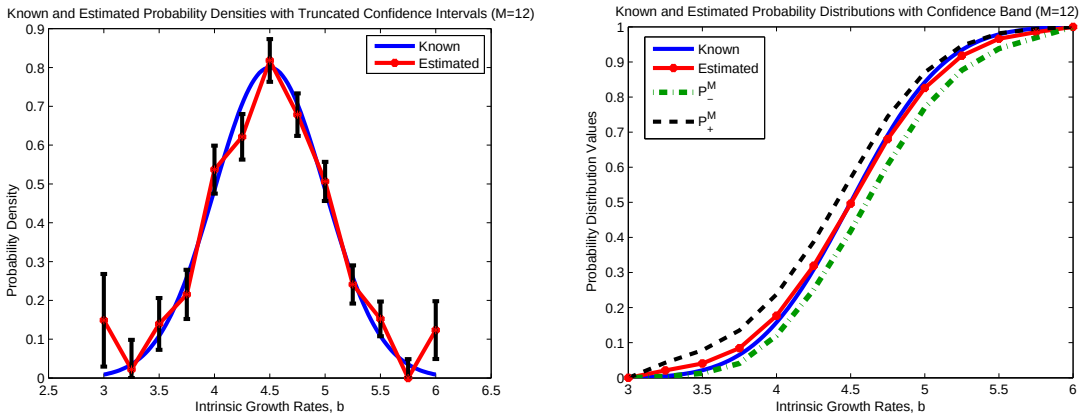

Figure 5.10: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using SPL(12,128) with 20% absolute error.

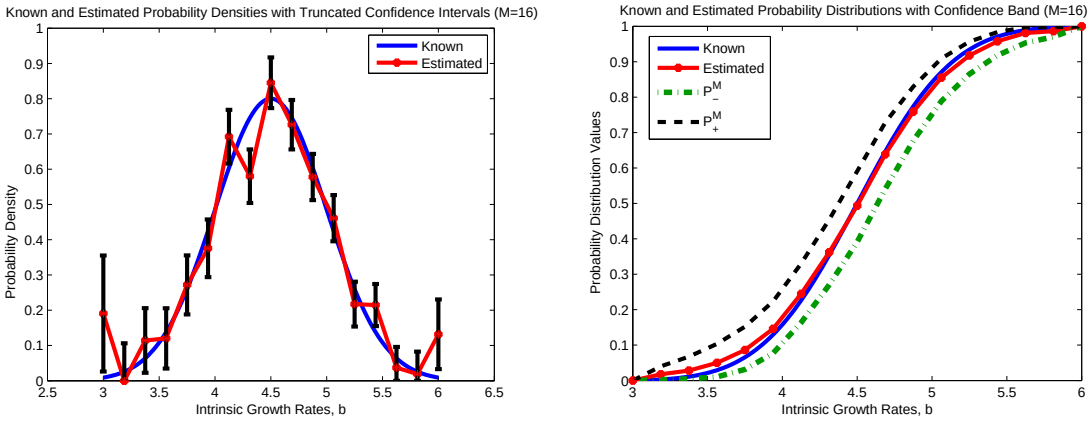

Figure 5.11: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using SPL(16,128) with 20% absolute error.

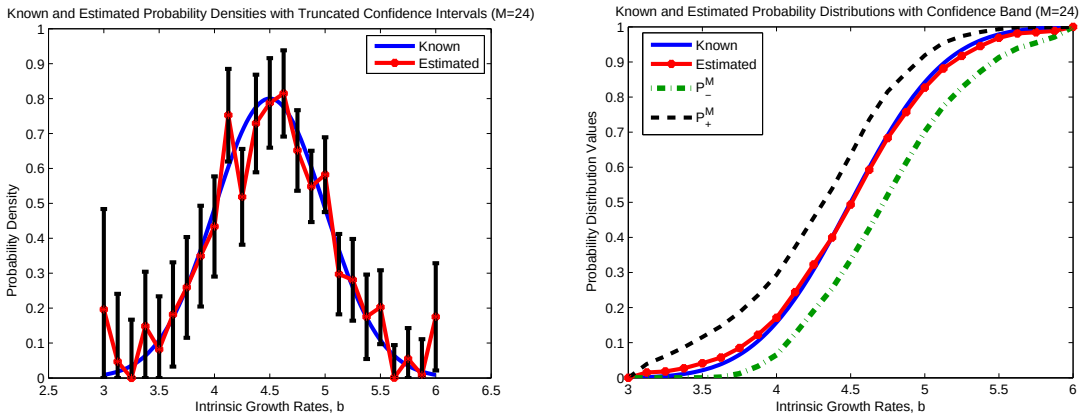

Figure 5.12: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using SPL(24,128) with 20% absolute error.

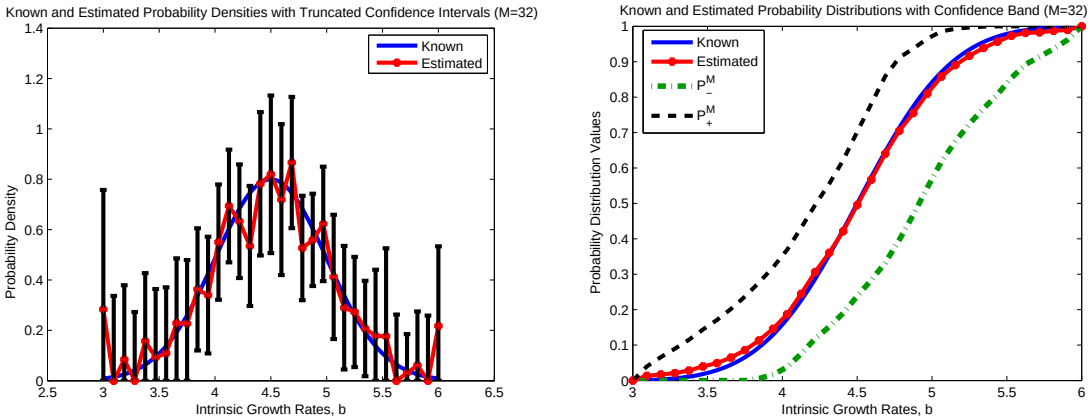

Figure 5.13: Estimated probability density with confidence intervals and probability distribution with confidence band for Gaussian example using SPL(32,128) with 20% absolute error.

5.3 Computational Results with bi-Gaussian Example

In this section, we discuss and present computational results for the inverse problem with data generated using the bi-Gaussian distribution with subpopulation means $\mu_{b_1} = 3.3$ and $\mu_{b_2} = 5.7$ and equal subpopulation variances $\sigma^2_{b_1} = \sigma^2_{b_2} = 0.492$ that we discussed earlier. We also added 20% absolute noise to the simulated data and used the same data set in all of the inverse calculations with PAR(M, N), DEL(M), and SPL(M, N).

5.3.1 Results with PAR(M, N)

The following results were obtained from computations carried out in MATLAB with PAR(3,128). The optimal cost for this set of results is 14.6131, while the estimate of $\hat{\sigma}^2$ is 0.0059. We also computed the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$, which is used in computing the standard errors for the finite-dimensional parameter θ, and obtained a value of 35.0084. Table 5.4 contains the optimal estimate of θ as well as the corresponding confidence intervals based on the standard error theory in Chapter 3. In Figure 5.14 the plots of the known and estimated probability densities along with p_- and p_+ are shown on the left as well as the plots of the known and estimated probability distributions and corresponding confidence band on the right. We note that the estimated probability distribution lies within the confidence band constructed using the technique that we outlined.

Table 5.4: Estimated parameters $(\hat{\mu}_{b_1}, \hat{\sigma}^2_{b_1}, \hat{\mu}_{b_2}, \hat{\sigma}^2_{b_2})$ and confidence intervals for bi-Gaussian example with 20% absolute error when using PAR(3,128).

$\hat{\mu}_{b_1} \pm 1.96SE(\hat{\mu}_{b_1})$	$\hat{\sigma}^2_{b_1} \pm 1.96SE(\hat{\sigma}^2_{b_1})$	$\hat{\mu}_{b_2} \pm 1.96SE(\hat{\mu}_{b_2})$	$\hat{\sigma}^2_{b_2} \pm 1.96SE(\hat{\sigma}^2_{b_2})$
3.2247 ± 0.0697	0.5767 ± 0.2597	5.7050 ± 0.0507	0.5844 ± 0.1840

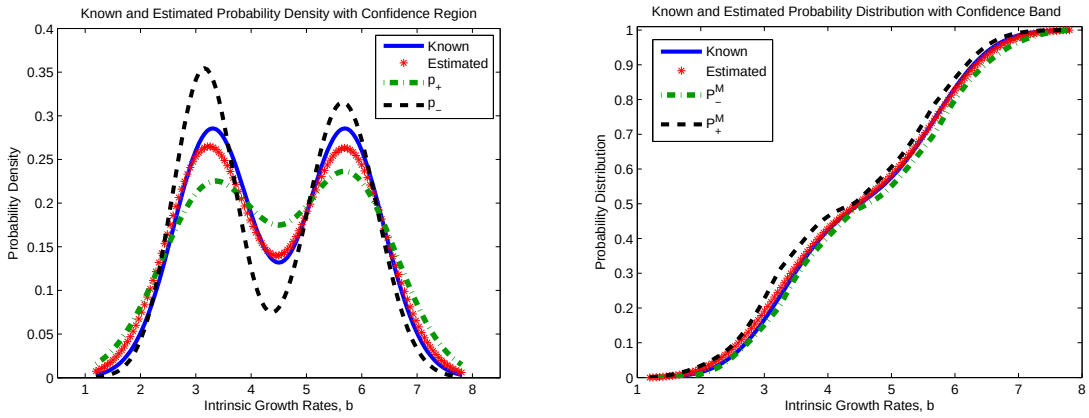

Figure 5.14: Estimated probability density with confidence region and probability distribution with confidence band for bi-Gaussian example using PAR(3,128) with 20% absolute error.

5.3.2 Results with DEL(M)

We now present some of the results obtained using DEL(M) for various values of M in the estimation problem using the bi-Gaussian data set with 20% absolute noise. The optimal cost values, estimates of $\hat{\sigma}^2$, and condition numbers κ of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ can be found in Table 5.5 for $M = 8$, 12, 16, 24, 32, 48, and 64. Figures 5.15 through 5.21 show the estimated probability densities and 95% confidence intervals as well as the estimated probability distributions and confidence bands. As the value of M is increased, we observe the optimal cost and estimate of σ_0^2 decrease, which we expect because we are allowing more degrees of freedom. The estimated probability distribution converges to the known distribution as M is increased. However, we also note from Table 5.5 that as M is increased, $\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$ increases. Once M becomes too large, the problem becomes over-parametrized and ill-conditioned (exhibited by the larger condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$), and we observe the confidence bands become larger. As M is increased from 8 to 32, the confidence bands appear to be converging nicely; however, when M is increased from 32 to 48 and from 48 to 64, we no longer observe nice convergence of the confidence bands. However, by examining the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$, we can better understand the behavior of the confidence bands, which appear to converge nicely until the problem becomes over-parametrized (beyond $M = 32$).

Table 5.5: Optimal cost values, $\hat{\sigma}^2$, and condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for bi-Gaussian example with 20% absolute error when using DEL(M).

M	$J(\hat{\theta})$	$\hat{\sigma}^2$	$\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$
8	40.3022	0.0162	15.6702
12	31.0134	0.0125	16.2206
16	25.8105	0.0104	16.9203
24	19.8538	0.0080	19.0560
32	16.4280	0.0067	22.1191
48	14.4454	0.0059	53.5280
64	14.1326	0.0058	105.5634

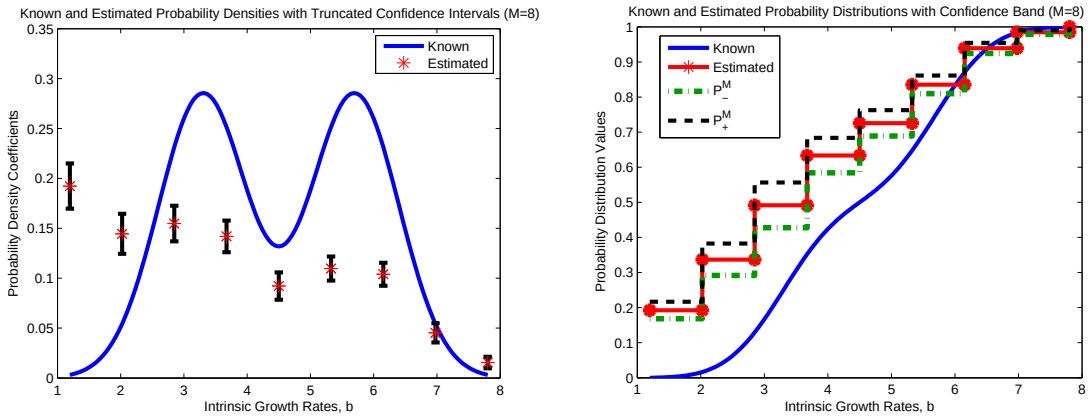

Figure 5.15: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using DEL(8) with 20% absolute error.

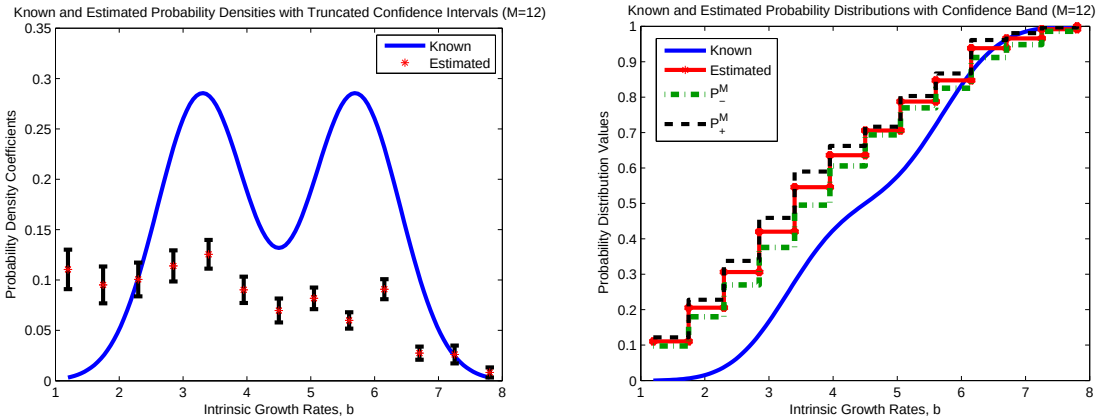

Figure 5.16: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using DEL(12) with 20% absolute error.

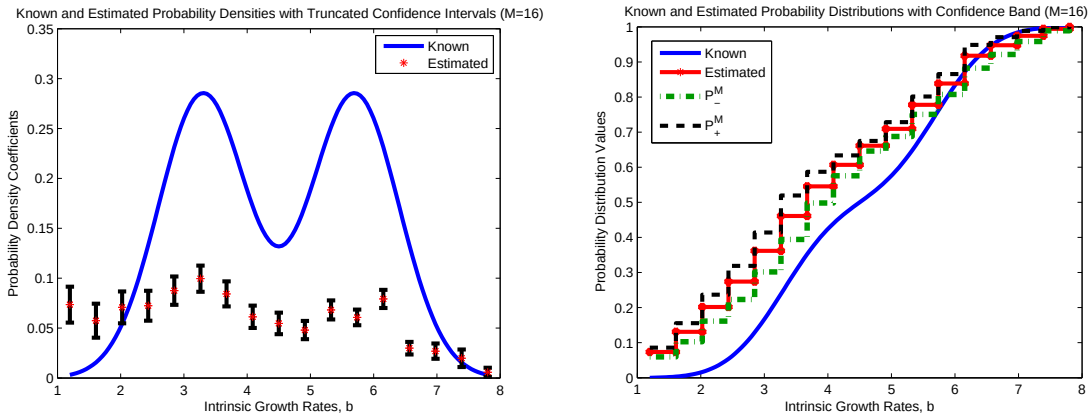

Figure 5.17: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using DEL(16) with 20% absolute error.

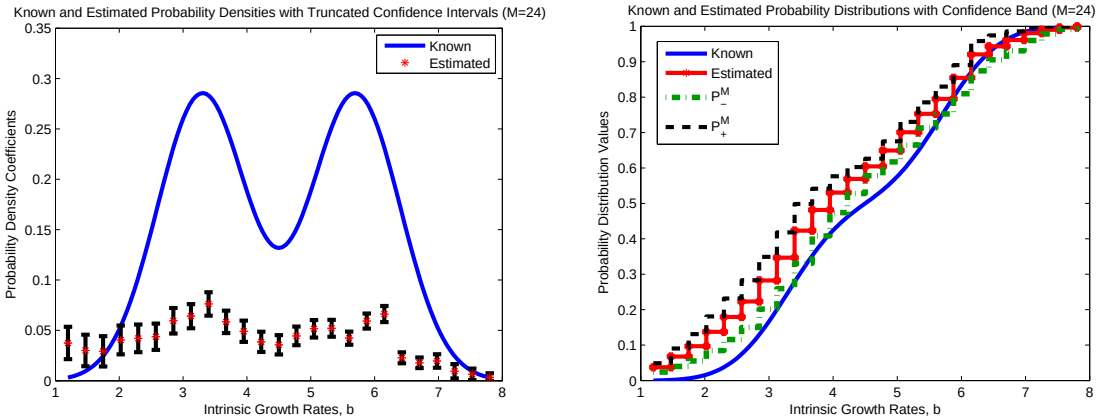

Figure 5.18: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using DEL(24) with 20% absolute error.

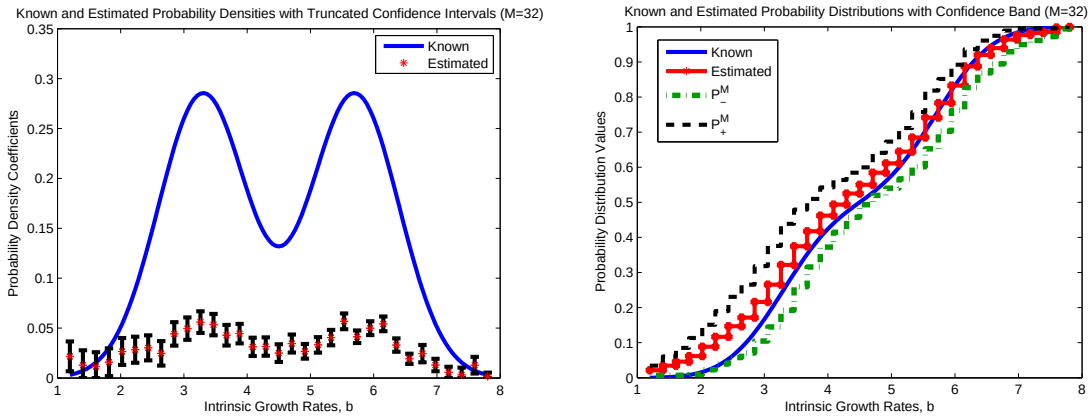

Figure 5.19: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using DEL(32) with 20% absolute error.

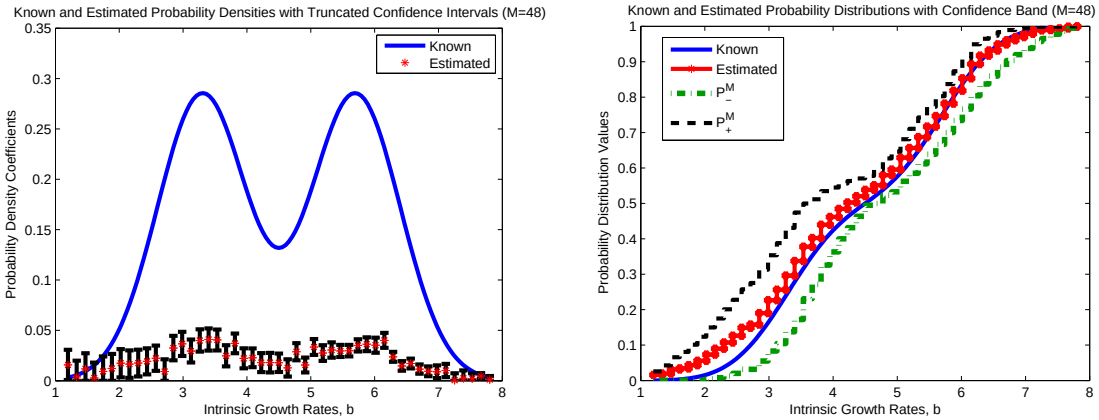

Figure 5.20: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using DEL(48) with 20% absolute error.

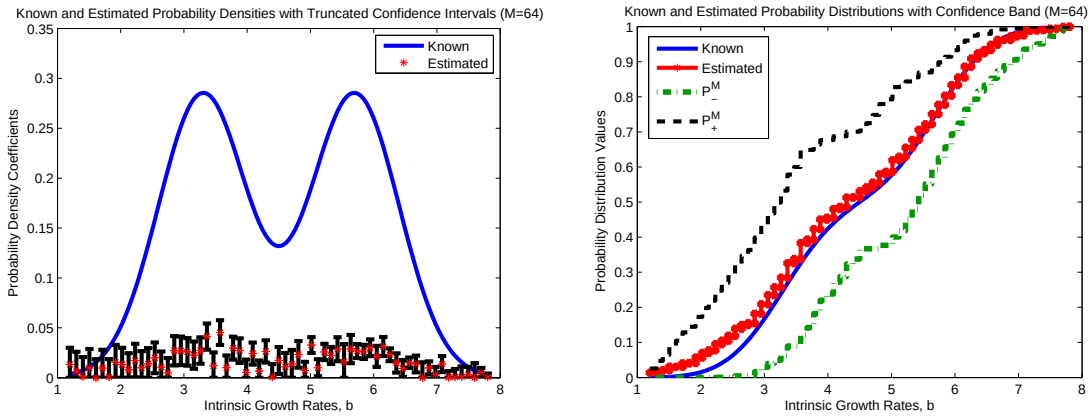

Figure 5.21: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using DEL(64) with 20% absolute error.

5.3.3 Results with SPL(M, N)

We also obtained computational results for the inverse problem with SPL$(M, 128)$ using the same data set with 20% absolute noise for various values of M. For $M = 8$, 12, 16, 24, and 32, we report the optimal cost values $J(\hat{\theta})$, the estimates $\hat{\sigma}^2$, and the condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ in Table 5.6. The figures displaying the estimated probability densities with the nodal 95% confidence intervals as well as the estimated probability distributions with the functional confidence bands for these values of M are shown in Figures 5.22 through 5.26. We observe the same type of behavior in the confidence bands here as noted when using DEL(M).

Table 5.6: Optimal cost values, $\hat{\sigma}^2$, and condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for bi-Gaussian example with 20% absolute error when using SPL$(M, 128)$.

M	$J(\hat{\theta})$	$\hat{\sigma}^2$	$\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$
8	14.6181	0.0059	22.4873
12	14.4995	0.0058	31.3596
16	14.4422	0.0058	40.2284
24	14.4240	0.0058	63.5889
32	14.3928	0.0058	91.0741

As M is increased, there is a small decrease in the optimal cost. The decrease in the estimate of the variance of the system $\hat{\sigma}^2$ is so small that it is not noticeable when reported to only four significant digits. We also note the increase in $\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$ as M is increased, and again, we can use this to explain the behavior we observe in the confidence bands constructed for these values of M. The confidence bands appear to be converging nicely as M is increased from 8 to 16. However, the confidence bands begin to grow larger as M is increased beyond 16 which is also accompanied by a much larger increase in the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for the values of M above 16. Over-parametrization of the inverse problem does not only affect the estimates obtained but the confidence bands as well. However, for appropriately chosen values of M, we observe very nice convergence of the confidence bands constructed using the technique outlined above for the approximation methods DEL(M) and SPL(M, N).

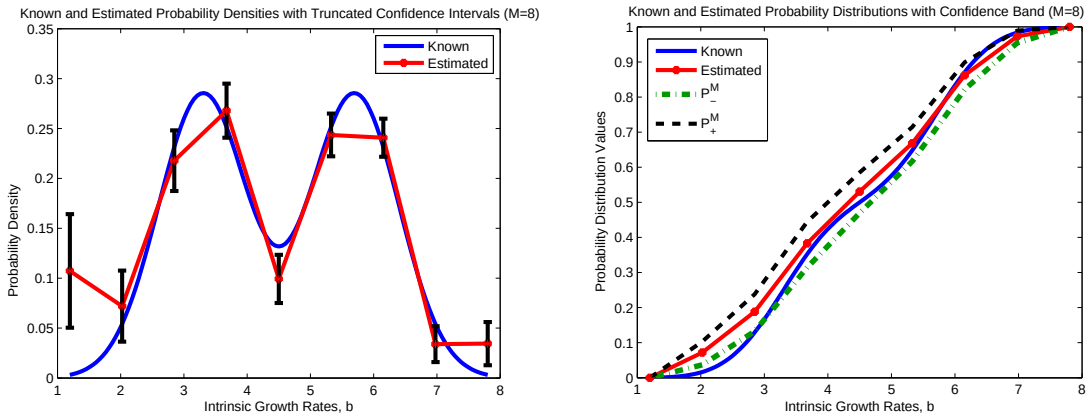

Figure 5.22: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using SPL(8,128) with 20% absolute error.

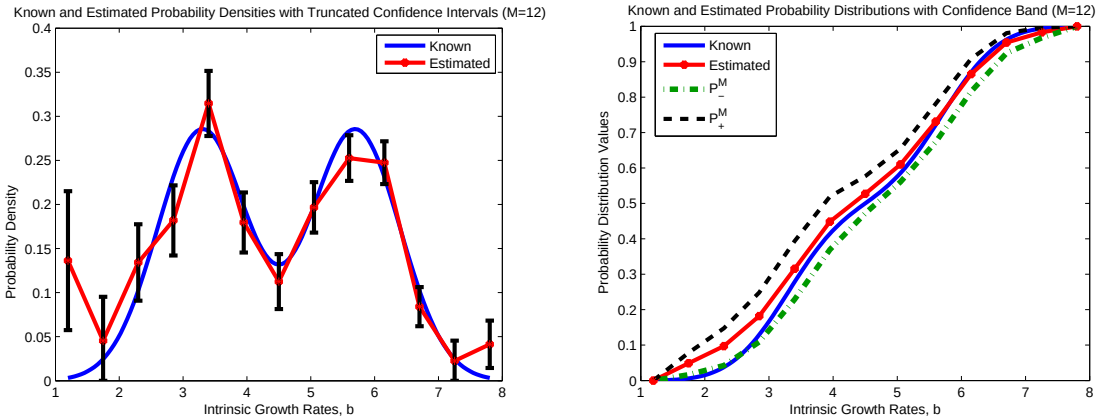

Figure 5.23: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using SPL(12,128) with 20% absolute error.

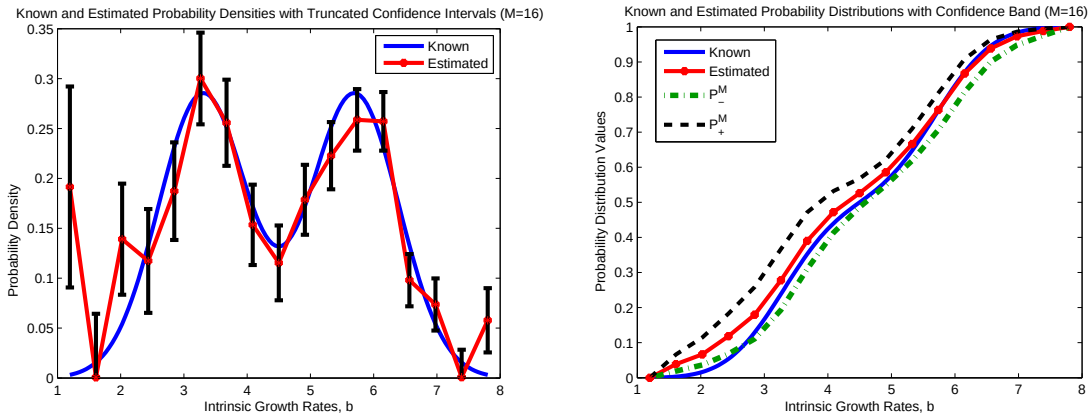

Figure 5.24: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using SPL(16,128) with 20% absolute error.

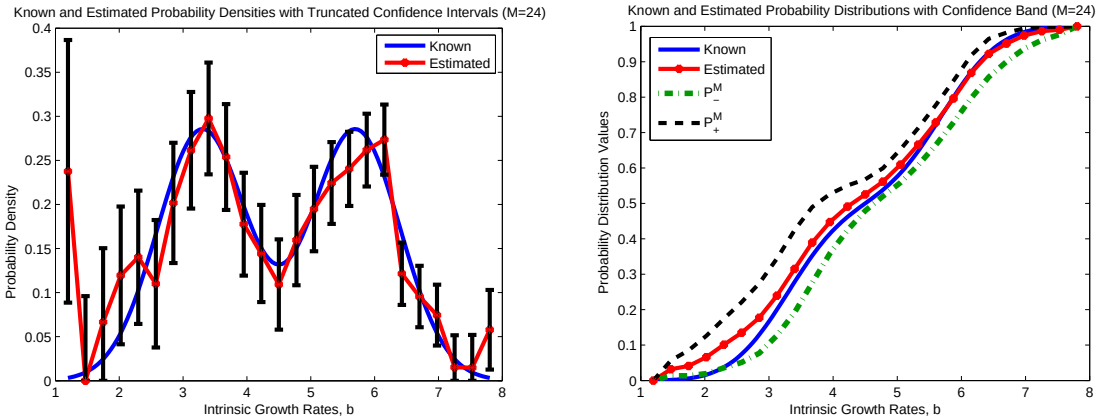

Figure 5.25: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using SPL(24,128) with 20% absolute error.

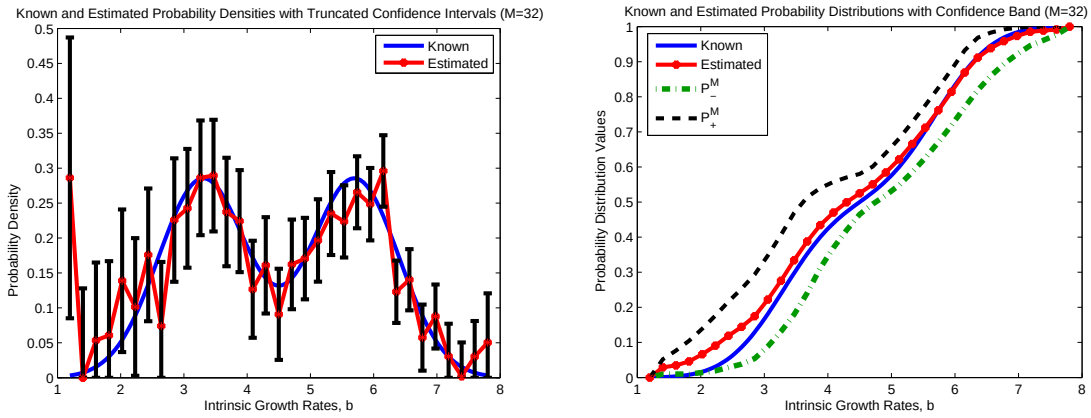

Figure 5.26: Estimated probability density with confidence intervals and probability distribution with confidence band for bi-Gaussian example using SPL(32,128) with 20% absolute error.

Other examples illustrating this behavior can be found in [9]. Moreover, the ideas outlined here can also be adopted to treat many other (non probability density) functional parameter estimation problems including those described in the Introduction.

5.4 Monte Carlo Simulations - Convergence of Confidence Bands

We derived the mean and variance of the OLS estimator for the functional regression model in Section 5.1. We also discussed a scheme to construct confidence bands for probability distributions in function space based on the confidence intervals obtained for the finite-dimensional parameters and presented several numerical examples demonstrating our technique in Sections 5.2 and 5.3. Although we must still determine an estimator for the variance σ_0^2 in the functional statistical model (5.1) and complete the theoretical arguments for the asymptotic sampling distribution for functional parameters, those efforts will be investigated further in future work. We focus instead on providing computational results from Monte Carlo simulations supporting the analysis carried out for the mean of the infinite-dimensional OLS estimator and suggesting the existence of "true" functional confidence bands for infinite-dimensional parameters.

Our Monte Carlo study consisted of a series N_{MC} of inverse problem calculations with the non-parametric approximation methods DEL(M) and SPL(M, N). We generated N_{MC} different realizations $\{y_j\}_{j=1}^n$ of the observation process described by (3.7) by adding randomly sampled noise to the solution of the GRD model (2.3) where the mean and variance of the "true" distribution P^* represented by θ_0 are fixed. Recall the observation process is described by

$$Y_j = f(\tau_j; \theta_0) + \epsilon_j, \quad j = 1, \ldots, n,$$

where $f(\tau_j; \theta_0)$ is the solution of the mathematical model and $\theta_0 \in \Theta \subset R^{M+1}$. For each different data set generated with random absolute noise, we determine a realization of the OLS estimator θ_{OLS}. That is, we compute N_{MC} estimates of the finite-dimensional parameter θ ($\{p_k^M\}_{k=0}^M$ for DEL(M) and $\{a_k^M\}_{k=0}^M$ for SPL(M, N)) used to approximate θ_0 in the statistical model. Each estimate of the finite-dimensional parameter θ at the density level corresponds to an estimate approximating the probability distribution that we seek in the original parameter estimation problem.

As we discussed in Section 3.3, for a single realization we can approximate the sampling distribution of the finite-dimensional OLS estimator θ_{OLS} with a multivariate normal distribution with mean $\hat{\theta}$ and covariance $\hat{\sigma}^2(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))^{-1}$, where $\hat{\theta}$ and $\hat{\sigma}^2$ are estimates based on a given data set. Using the standard asymptotic theory for finite-dimensional OLS estimators, we developed a technique to construct confidence bands for the estimated probability distributions based on the nodal confidence intervals for a specific data set. Since we obtain N_{MC} estimates of the probability distribution from the Monte Carlo simulations, we can determine the sample mean and sample variance of the approximate probability distributions for a fixed value of M when considering DEL(M) and SPL(M, N). The sample mean and sample variance can then be used to construct a confidence band for the average estimated probability distribution based on the results of the Monte Carlo runs. From the results of these two different approaches, we can determine if the sample means of the estimated probability distributions are converging to the "true" probability distribution which would validate our analytical arguments in Section 5.1. The confidence bands constructed from the Monte Carlo simulations would also provide computational evidence as to whether or not the approximate confidence bands are converging to some "true" functional confidence band as M tends to ∞. Furthermore, we can compare the confidence bands based on the Monte Carlo simulations to the confidence bands constructed from the confidence intervals based on the asymptotic standard error theory for finite-dimensional parameters for a single realization or data set in order to gain more insight on the behavior observed and discussed in Sections 5.2 and 5.3. We present results addressing these issues in the following two sections for both a Gaussian and bi-Gaussian example.

5.4.1 Results with Gaussian Example

In this section we present the results of 5000 Monte Carlo simulations with 20% absolute noise added to data (previously described in Chapter 4) generated with a "true" approximate truncated Gaussian distribution with mean $\mu_b = 4.5$ and variance $\sigma_b^2 = 0.25$ on the intrinsic growth rates b. The first set of results shown in Figures 5.27 through 5.29 were computed with the delta function approximation method DEL(M) for $M = 4$, 8, 12, 16, 24, and 32. The estimates of $\{p_k^M\}_{k=0}^M$ and corresponding probability distributions P^M for these values of M were determined by solving the quadratic programming problem in (3.5). We remark that this problem is minimized over the set given in (3.4). After carrying

out 5000 inverse problem calculations, we computed the sample mean and sample variance of the estimated probability distributions for each value of M. The resulting averages of the estimated probability distributions are shown on the left in Figure 5.27. We observe the sample means of the estimated probability distributions converging nicely to the "true" probability distribution as M is increased from 4 to 32. The same behavior is also observed in the resulting estimated probability distributions for the various values of M for a single realization or data set as seen in the plot on the right in Figure 5.27.

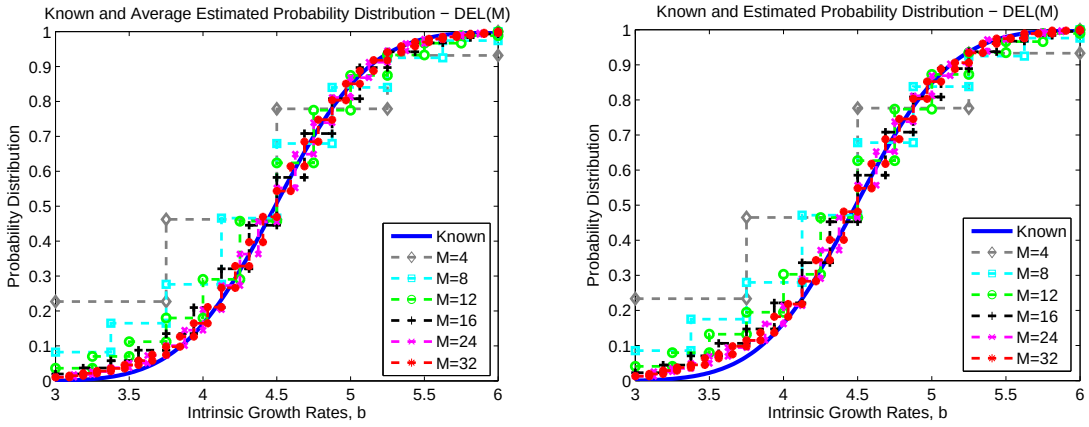

Figure 5.27: (a) Monte Carlo averages of estimated P^M and (b) estimated P^M corresponding to a single realization for Gaussian example using DEL(M) for M = 4, 8, 12, 16, 24, 32.

The upper and lower edges of the 95% confidence bands for the Monte Carlo simulations are constructed by adding and subtracting $1.96 \cdot \sigma_{MC}$ to the sample mean where σ_{MC} is the sample standard deviation or square root of the sample variance from the Monte Carlo runs. The lower and upper edges of the 95% confidence bands based on these results are shown in the top two plots in Figure 5.28. As M is increased from 4 up to 32 the lower edges of the 95% confidence bands (on the left) and the upper edges of the 95% confidence bands (on the right) both appear to be converging nicely to functions representing the edges of the "true" 95% functional confidence band. This behavior is in contrast to the behavior observed in the numerical results in Section 5.2 as well as the lower and upper edges of the confidence bands shown in the bottom of Figure 5.28 constructed with the scheme based on the nodal asymptotic based confidence intervals for a single data set. As we previously noted, the confidence bands constructed from the confidence intervals for the finite-dimensional estimates converge nicely until the problem becomes over-parametrized.

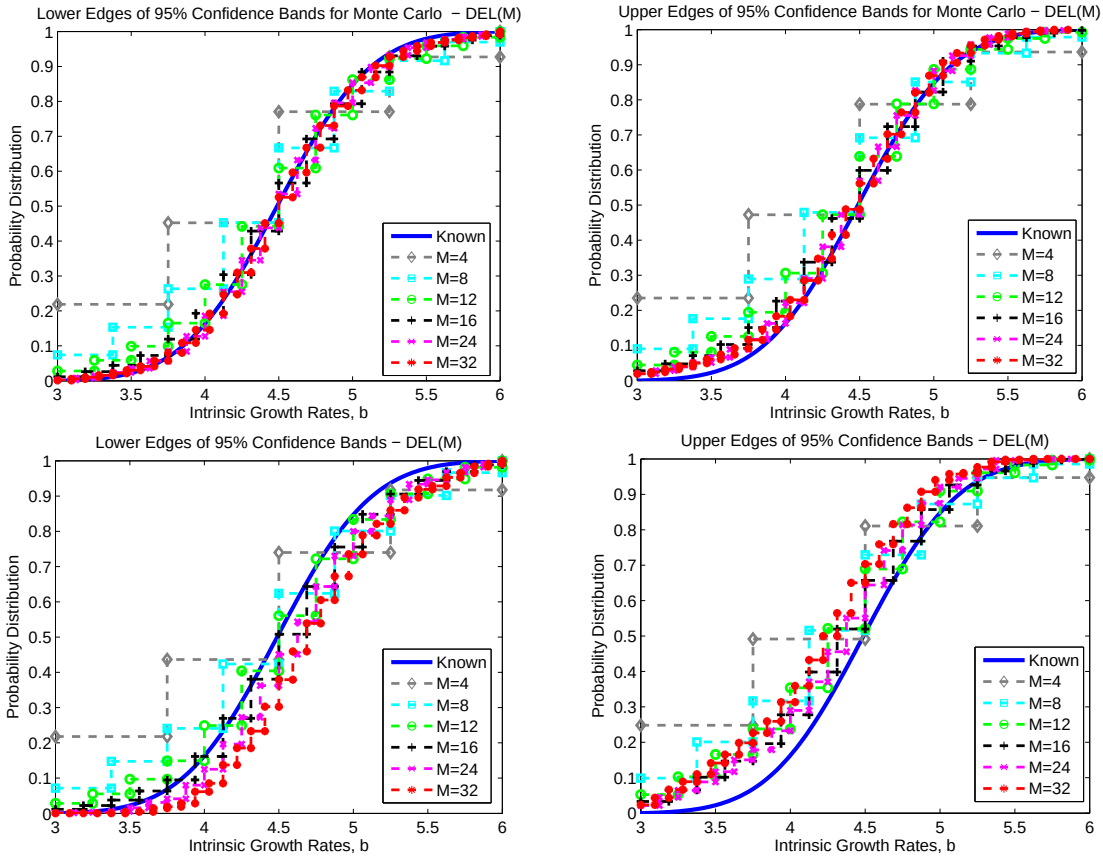

Figure 5.28: Lower and upper edges of 95% confidence bands for Monte Carlo (upper) and nodal based asymptotic theory (lower) for Gaussian example using DEL(M) for $M = 4$, 8, 12, 16, 24, 32.

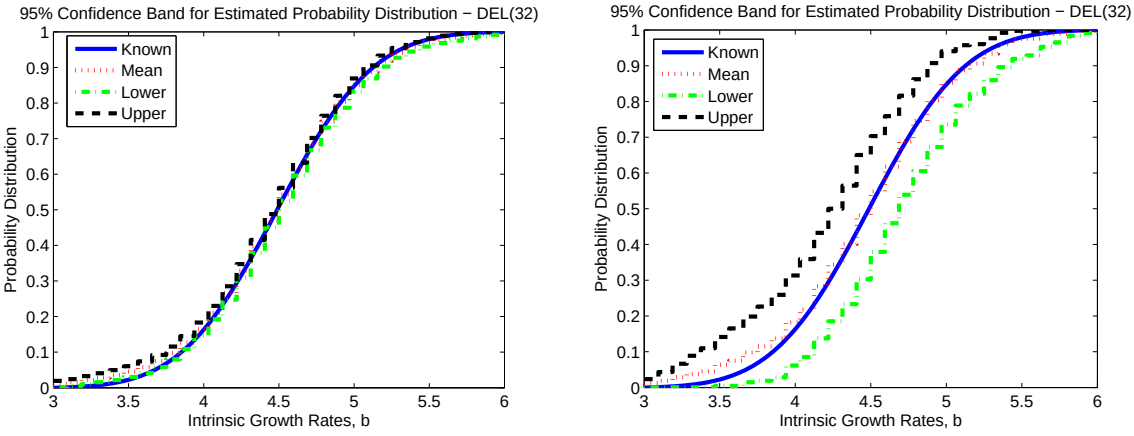

Figure 5.29: 95% confidence band for (a) Monte Carlo and (b) nodal based asymptotic theory for Gaussian example using DEL(32).

Once M is too large, the 95% confidence bands based on the nodal confidence intervals grow larger and the condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ increase as a result of using too many elements which was also noticed in the examples reported in Section 5.2. We see a much tighter confidence band from the Monte Carlo simulations for DEL(32) in Figure 5.29 in comparison to the much larger confidence band based on the confidence intervals for DEL(32) for a single data set. These results suggest that one must be careful when using the confidence bands that are obtained by extending the asymptotic standard error theory in the finite-dimensional case for statistical analysis since wider confidence bands are indicative of less reliable estimates of the probability distribution.

The next set of results in Figures 5.30 through 5.32 were obtained with the spline based approximation method SPL(M, 128) for 5000 Monte Carlo simulations with the same data set used in the calculations with DEL(M). We also used the same values of M = 4, 8, 12, 16, 24, and 32. In this case we minimized (3.5) over the finite-dimensional set in (3.6) and determined estimates for the coefficients $\{a_k^M\}_{=0}^M$ as well as the approximating continuous probability distributions P^M. The results of the 5000 Monte Carlo runs were similar to those obtained with DEL(M). In Figure 5.30 the sample means of the estimated probability distributions converge to the "true" Gaussian probability distribution supporting the theoretical arguments in Section 3.3. We note again (as in previously discussed examples) the much faster convergence of the estimated probability distributions when using the scheme with piecewise linear splines in both plots in Figure 5.30 in comparison to the delta function approximation method.

Using the sample means and sample variances of the Monte Carlo runs, we also constructed the lower and upper edges of the 95% confidence bands that are shown in the top of Figure 5.31. The behavior of the bands shown here is also suggestive of the existence of "true" functional confidence bands. As M is increased from 4 to 32 we observe nice convergence behavior in both the lower and upper edges of the 95% confidence bands (again at a much faster rate than those of DEL(M)). In fact, there is little noticeable difference in the edges for M = 8, 16, 24, and 32. However, when comparing the confidence bands from the Monte Carlo simulations to those constructed from the nodal asymptotic based confidence intervals we see significant differences. Although the confidence bands based on extending the asymptotic standard error theory shown in the bottom of Figure 5.31 appear to converge initially for the smaller values of M, the bands grow noticeably larger when M is increased to 24 and 32. We again remark that this behavior is readily

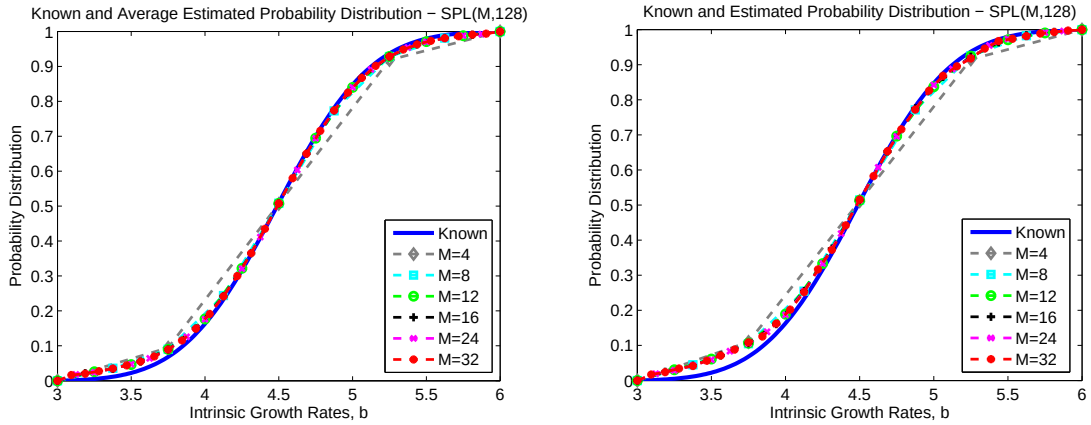

Figure 5.30: (a) Monte Carlo averages of estimated P^M and (b) estimated P^M corresponding to a single realization for Gaussian example using SPL$(M, 128)$ for $M = 4$, 8, 12, 16, 24, 32.

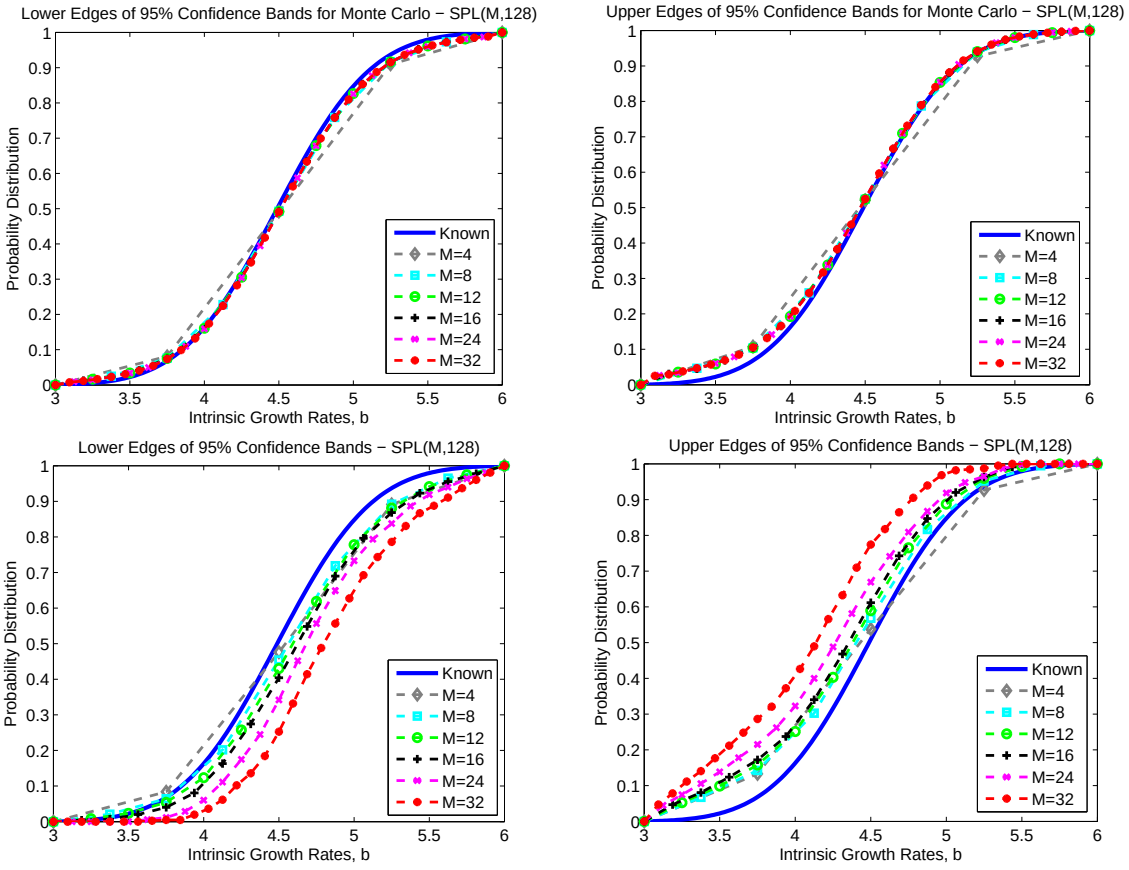

Figure 5.31: Lower and upper edges of 95% confidence bands for Monte Carlo (upper) and nodal based asymptotic theory (lower) for estimated probability distributions for Gaussian example SPL$(M, 128)$ for $M = 4$, 8, 12, 16, 24, 32.

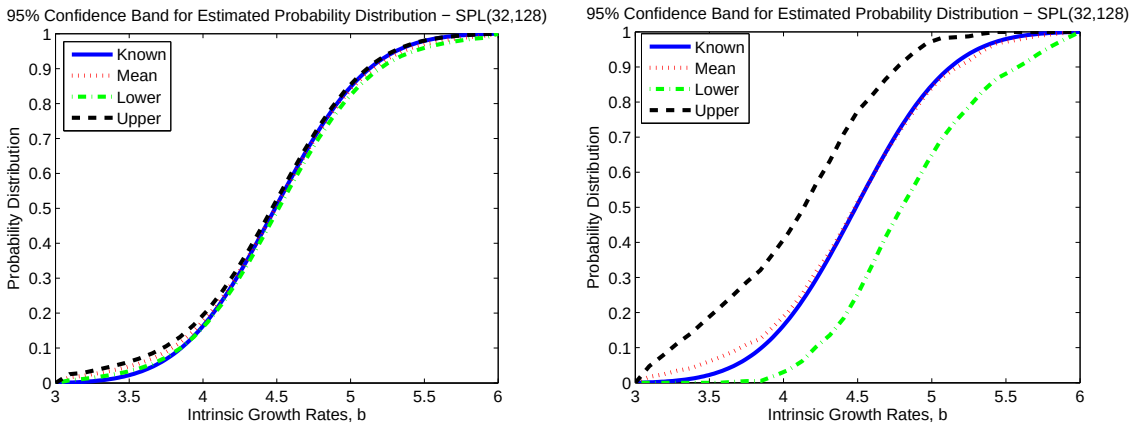

Figure 5.32: 95% confidence band for (a) Monte Carlo and (b) nodal based asymptotic theory for Gaussian example using SPL(32,128).

explained by studying the condition number κ of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ which increased significantly as M was increased from 16 to 24 and 24 to 32 in the previous numerical example in Section 5.2. Furthermore, it is clear from Figure 5.32 that our scheme is highly sensitive to over-parametrization in the inverse problem as noticed by the much larger confidence band for SPL(32,128) created with the technique discussed earlier in this chapter.

5.4.2 Results with bi-Gaussian Example

The results in this section were obtained from 5000 Monte Carlo simulations with 20% absolute noise added to data generated with a "true" approximate truncated bi-Gaussian distribution with subpopulation means $\mu_{b_1} = 3.3$ and $\mu_{b_2} = 5.7$ and equal subpopulation variances $\sigma^2_{b_1} = \sigma^2_{b_2} = 0.492$ on the intrinsic growth rates b which was used in previous examples in Chapter 4. In the first set of results in Figures 5.33 through 5.35, we used DEL(M) with M = 8, 12, 16, 24, 32, 48 and 64. We reiterate that the estimates of $\{p_k^M\}_{k=0}^M$ for DEL(M) and the corresponding probability distributions P^M were found by minimizing the quadratic programming problem in (3.5) over the finite-dimensional approximate probability measure space (3.4). We calculated the sample mean and sample variance of the estimated probability distribution for each value of M after performing the inverse problem 5000 times. The sample means for the values of M considered in this example appear to be converging to the "true" probability distribution based on the plot on the left in Figure 5.33. These computational results also support the analysis carried out in the

previous section. Furthermore, we notice the same convergence behavior in the estimated probability distributions P^M as M is increased from 8 to 64 for a single data set in the right plot in Figure 5.33.

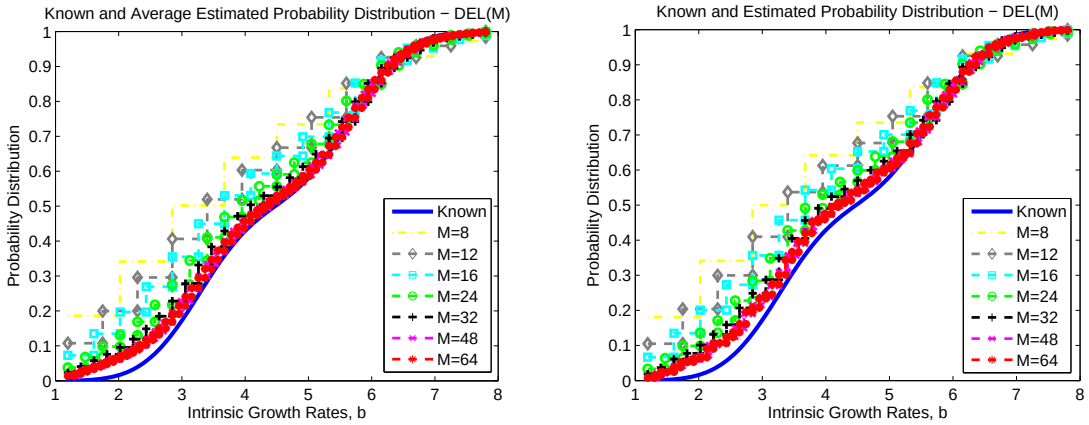

Figure 5.33: (a) Monte Carlo averages of estimated P^M and (b) estimated P^M corresponding to a single realization for bi-Gaussian example using DEL(M) for M = 8, 12, 16, 24, 32, 48, 64.

In the top of Figure 5.34 we see the lower and upper edges of the 95% confidence bands constructed from the sample means and sample standard deviations of the Monte Carlo simulations. We observe the lower edges of the 95% confidence bands (top left) and upper edges of the 95% confidence bands (top right) both converging to some "true" lower and upper edges of a functional confidence band as M is increased from 8 up to 64. The difference in the convergence properties of the edges of these confidence bands in comparison to the edges of the confidence bands constructed from our proposed scheme in the bottom of Figure 5.34 as well as in the computational results in Section 5.3 should also be noted. The confidence bands based on the asymptotic standard error theory for finite-dimensional parameters appear to behave nicely for small values of M; however, the confidence bands grow larger when too many elements are used in the approximation. As a result of over-parametrization of the problem, the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ increases significantly which further highlights the problem of using large values of M as discussed in Section 5.3. The confidence bands from the 5000 Monte Carlo runs are very suggestive of "true" confidence bands despite the behavior of the confidence bands constructed from the nodal confidence intervals (see Figure 5.35 for a comparison of the confidence bands for DEL(64)).

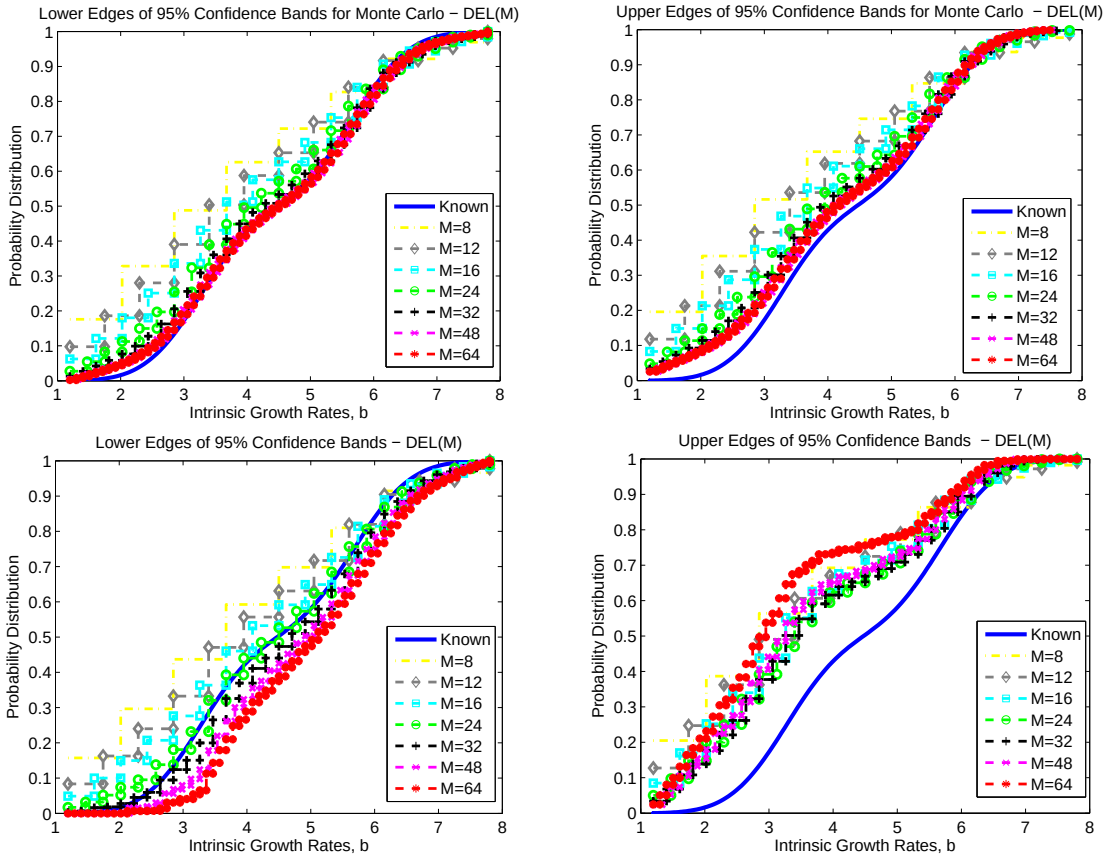

Figure 5.34: Lower and upper edges of 95% confidence bands for Monte Carlo (upper) and nodal based asymptotic theory (lower) for bi-Gaussian example using DEL(M) for $M = 8$, 12, 16, 24, 32, 48, 64.

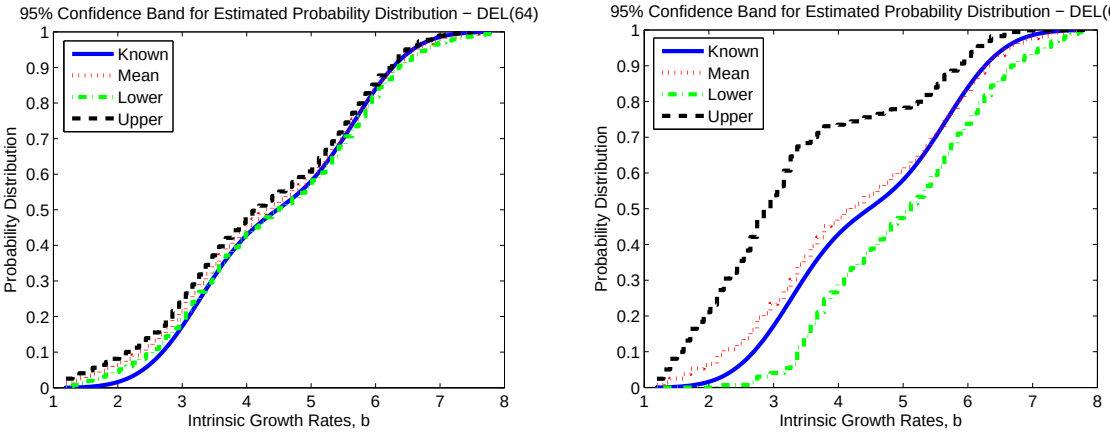

Figure 5.35: 95% confidence band for (a) Monte Carlo and (b) nodal based asymptotic theory for bi-Gaussian example using DEL(64).

The last set of results in this section were obtained from SPL$(M, 128)$ for $M =$ 8, 12, 16, 24, and 32 for 5000 Monte Carlo runs with the same data used in the computations with DEL(M). After determining estimates of the coefficients $\{a_k^M\}_{k=0}^M$ for the approximate probability density and distribution from the quadratic programming problem, we again computed the sample means and sample variances of the estimated probability distributions P^M for the various values of M. It is clear from the results in Figure 5.36 that the sample means of the estimated probability distribution as well as the estimated probability distributions for one data set converge for the values of M considered in this example. Moreover, the spline based approximation method provides much faster convergence in comparison to the delta function approximation method.

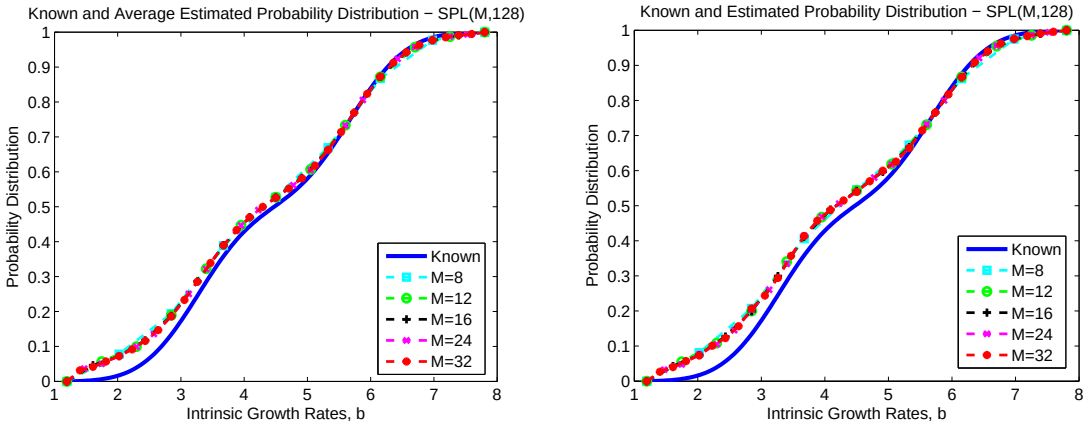

Figure 5.36: (a) Monte Carlo averages of estimated P^M and (b) estimated P^M corresponding to a single realization for bi-Gaussian example using SPL$(M, 128)$ for $M = 8$, 12, 16, 24, 32.

We also observe very nice convergence of the upper and lower edges of the 95% confidence bands from the Monte Carlo simulations in the top of Figure 5.37 for $M = 8$, 12, 16, 24, and 32. As M is increased from 8 up to 32 both the lower and upper edges of the 95% confidence bands appear to be converging to the lower and upper edges of the "true" 95% functional confidence band. In comparison to the lower and upper edges of the confidence bands in the lower two plots in Figure 5.37 and the confidence bands reported previously in Section 5.3, we see again that the results from the Monte Carlo simulations are much different. However, as we have previously pointed out, the confidence bands determined from the nodal asymptotic based confidence intervals are very much dependent upon using the appropriate number of elements in the approximation of the probability

distribution. When the value of M used in the parameter estimation problem is too large, the widths of the resulting confidence bands are much larger than when small values of M are used. These larger confidence bands are illustrative of the growing condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ which is used in determining the confidence intervals for the finite-dimensional parameters. Therefore, one must be aware of the effects that over-parametrization can have on the confidence bands constructed from the nodal confidence intervals such as that seen when comparing the confidence band from the Monte Carlo runs with those based on the finite-dimensional asymptotic standard error theory for $M = 32$ in Figure 5.38.

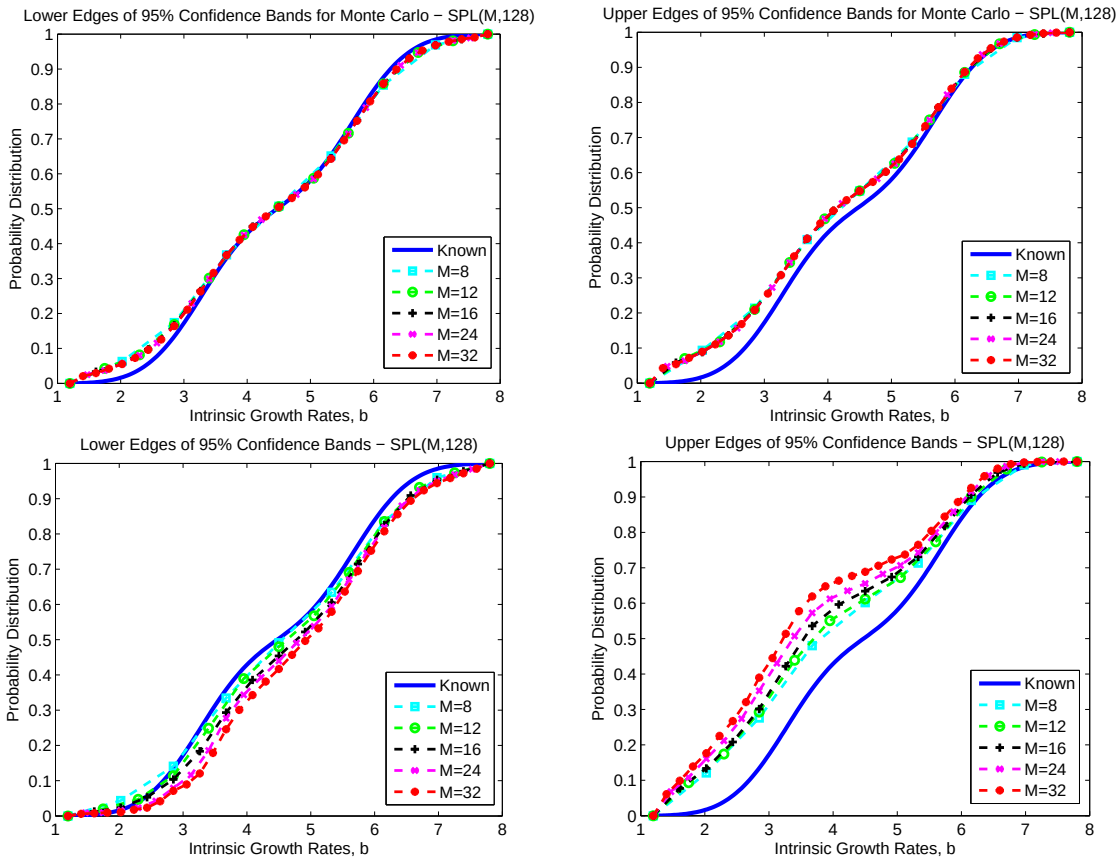

Figure 5.37: Lower and upper edges of 95% confidence bands for Monte Carlo (upper) and nodal based asymptotic theory (lower) for bi-Gaussian example SPL$(M, 128)$ for M = 8, 12, 16, 24, 32.

In summary, the results from both the Gaussian and bi-Gaussian examples demonstrate that the sample means of the estimated probability distributions from the Monte Carlo runs converge to the "true" probability distributions as $M \to \infty$, thereby confirming

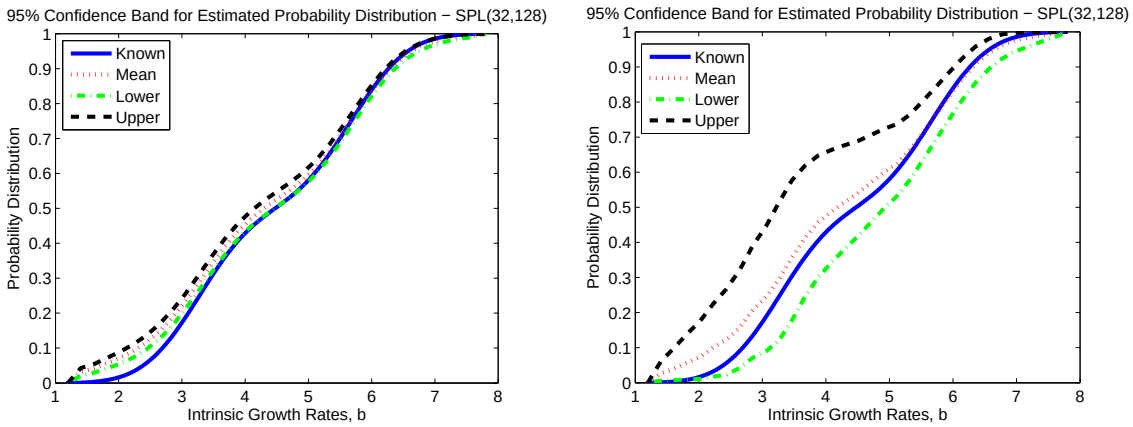

Figure 5.38: 95% confidence band for (a) Monte Carlo and (b) nodal based asymptotic theory for bi-Gaussian example using SPL(32,128).

the theoretical analysis for the OLS estimator for the functional linear regression model. We also observe convergence of the lower and upper edges of the 95% confidence bands from these simulations which suggests the existence of "true" functional confidence bands. In comparing these results to those obtained based on our technique of extending the nodal asymptotic based confidence intervals constructed in finite dimensions, we found that the confidence bands from our scheme are very much dependent on the number of elements used in the approximations. Over-parametrization of the parameter estimation problem results in larger confidence intervals which in turn results in larger confidence bands suggesting less reliable estimates of the probability distribution. However, the computational results presented in this chapter illustrate the usefulness of our technique in quantifying the uncertainty in the estimation of functional parameters as long as one is also aware of the effects that over-parametrization can have on the estimates from the inverse problem as well as the measures of uncertainty associated with these estimates.

Chapter 6

Computational Example: Size-Structured Shrimp Population

6.1 Problem Description

In this chapter we present results for the inverse problem for the estimation of growth rate distributions in size-structured shrimp populations. The efforts discussed here are motivated by previous work on a hybrid model of the shrimp biomass/countermeasure production system developed and discussed in [5]. One of the goals of this joint project with Advanced Bionutrition Corporation (ABN) was the development of a model for a system where one uses shrimp as a scaffold organism to produce large amounts of vaccine in response to bio-toxic attacks on populations. Being able to accurately model the dynamics of the size-structured shrimp population is important since the output of the biomass model will serve as input to the vaccine production model.

Variability in size has been observed in experimental data shown in Figures 6.1 and 6.2 for the early growth of shrimp collected from two different raceways at the Shrimp Mariculture Research Facility, Texas Agricultural Experiment Station in Corpus Christi, TX. Although the initial sizes of the shrimp were very similar, a great deal of variability was observed in the aggregate type longitudinal data as time progressed. Therefore, a reasonable model for this population would need to account for the variability in the size distribution data which is perhaps a result of variability in the individual growth rates across the population [33].

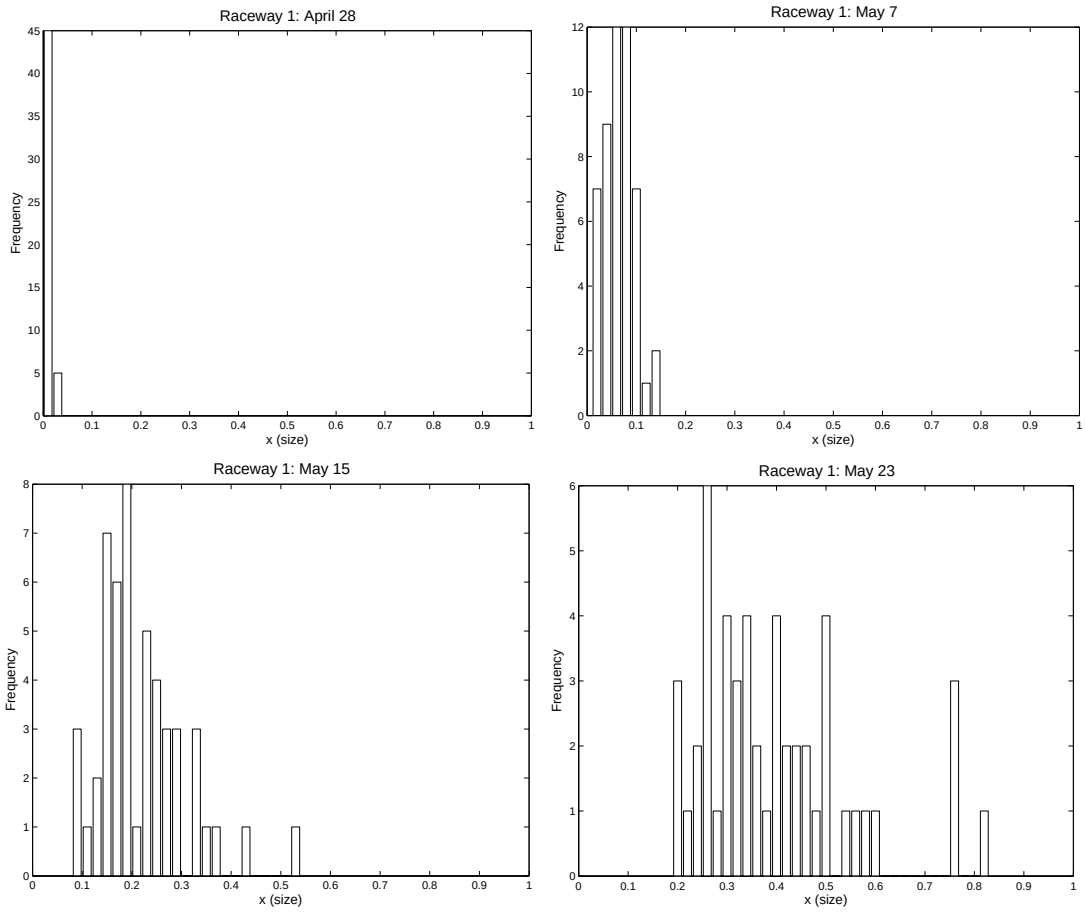

Figure 6.1: Histograms for longitudinal data for Raceway 1.

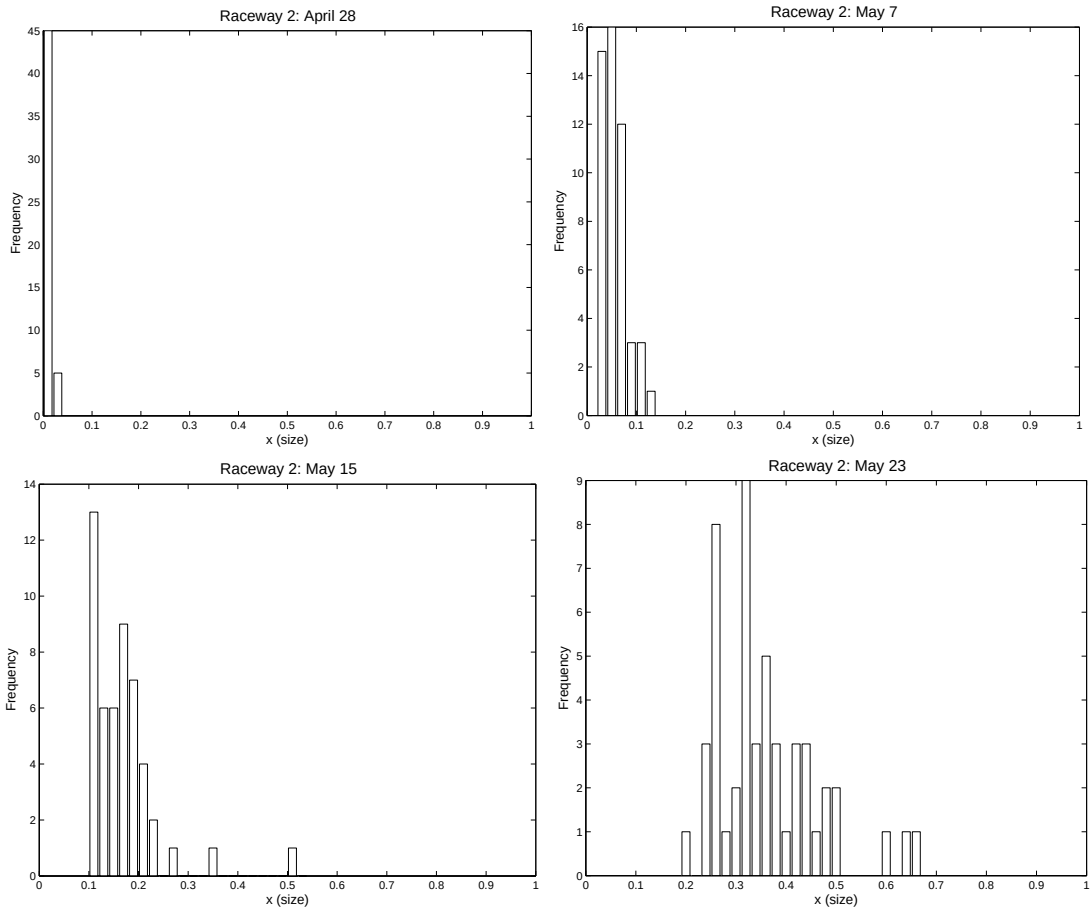

Figure 6.2: Histograms for longitudinal data for Raceway 2.

The GRD model (2.3) represents one approach to accounting for uncertainty in growth rates by imposing a probability distribution on the growth rates in the SS model (2.1). As we noted in Chapter 2, the individuals in the population grow according to a deterministic growth rate function (2.2), but different individuals in the population may have different parameter dependent growth rates in the GRD model. The population is assumed to consist of subpopulations with individuals in the same subpopulation having the same growth rate. The growth uncertainty of individuals in the population is the result of variability in growth rates among the subpopulations. This modeling approach may be most applicable when the growth of individuals is assumed to be the result of genetic variability, such as in [33].

However, a second approach that has been studied as well is based on the assumption that individual growth is a Markov diffusion stochastic process which leads to the Fokker-Planck or forward Kolmogorov model for the shrimp population density [1, 27, 41, 53]. The growth process for each individual is stochastic, and each individual grows according to a stochastic growth model. In the Fokker-Planck model, the uncertainty in growth of individuals is the result of the growth stochasticity of each individual. This modeling approach may be most applicable when the variability in growth rate of individuals is believed to be the result of changes in environmental factors, such as [43, 47, 54]. Theoretical arguments in [10] demonstrate that the population density from the GRD model is the same as the population density obtained from the Fokker-Planck model when equivalent levels of variability are used in both models. Numerical results were also presented in [11] to further validate the theoretical analysis of [10]. Therefore, one can use the computationally "easier" approach to model the population of interest when appropriately chosen forms of variability can be determined. Based on these studies, we chose to use the GRD model (2.3) to incorporate uncertainty in the growth rates in the size-structured population model for the early growth of shrimp.

6.2 Experimental Design

In this section we outline an approach for determining the sampling size and the number of sampling time points one needs to obtain reliable estimates for probabilistic growth rate parameters in the GRD model. We also present some computational results demonstrating the effect of the bin size used in inverse problem calculations on parameter

estimates and model predictions.

Before discussing the results, we first describe the simulated data used in these calculations. The experimental data to be used in the inverse problem calculations will be aggregate type longitudinal data (similar to the data shown in Figures 6.1 and 6.2). Individual shrimp are randomly sampled from the population at each time point; however, there is no guarantee that the same set of individuals are sampled at different time points. The simulated data that we use in the calculations in this section will also be aggregate longitudinal data.

We assume that the mortality rate $\mu(t,x)$ and reproduction rate $K(t,x)$ in (2.1) are both zero because we only consider the early growth dynamics of the shrimp. We also assume that the growth rate function has the form

$$\frac{dx}{dt} = g(x) = b(x+c), \tag{6.1}$$

where b represents the intrinsic growth rate of the individuals and c is a fixed constant. This growth rate function was shown to be reasonable in [10] where the average size (weight in grams) data for fifty randomly sampled shrimp was fit with the exponential function corresponding to the solution of (6.1). To satisfy the assumption of varying growth rates in the GRD model, we assume that the intrinsic growth rate is a random variable taking values in a compact set $\mathfrak{B} = [\underline{b}, \bar{b}]$ with probability density function $p(b)$. Here, we chose it to be a truncated normal (Gaussian) distribution with mean μ_b and standard deviation σ_b, denoted by $\mathcal{N}_{[\underline{b},\bar{b}]}(\mu_b, \sigma_b^2)$. This choice was based on previous analysis that demonstrated that an assumption of a normal distribution on the intrinsic growth rates leads to a lognormal distribution in size, which is typical of data collected on shrimp populations such as those in Figures 6.1 and 6.2. Therefore, the population density in (2.3) is given by

$$u(t,x;\mu_b,\sigma_b) = \int_{\underline{b}}^{\bar{b}} v(t,x;b)p(b;\mu_b,\sigma_b)db = \int_{\underline{b}}^{\bar{b}} v(t,x;b)\frac{\frac{1}{\sigma_b}\phi\left(\frac{b-\mu_b}{\sigma_b}\right)}{\Phi\left(\frac{\bar{b}-\mu_b}{\sigma_b}\right) - \Phi\left(\frac{\underline{b}-\mu_b}{\sigma_b}\right)}db, \tag{6.2}$$

where ϕ is the probability density function of the standard normal distribution and Φ is its corresponding cumulative distribution function.

While individual shrimp are randomly sampled and weighed at each time point, the data used in the inverse problem calculations is the total number of shrimp in each size class. Let Δx be the length of the size class interval that we choose for each size class bin. Then the total number of population $N_{GRD}(t,x;\theta)$ in each size class bin is approximated

by

$$N_{GRD}(t, x; \theta) \approx u(t, x; \theta)\Delta x, \tag{6.3}$$

where $\theta = (\mu_b, \sigma_b)$, and $u(t, x; \theta)$ is obtained from (6.2).

Let the sampling time points be given by $t_k, k = 1, 2, \ldots, N_t$. At each time point t_k, we independently draw N_s samples from $\mathcal{N}_{[\underline{b},\bar{b}]}(\mu_b, \sigma_b^2)$; that is, we obtain N_s samples of intrinsic growth rates b_i^k, $i = 1, 2, \ldots, N_s$. We also independently draw N_s samples of initial sizes $x_{0,i}^k$ from a uniform distribution on the interval $[\underline{x}_0, \bar{x}_0]$, where $\underline{x}_0$ and $\bar{x}_0$ are some constants. Let s_i^k represent the size of shrimp at time t_k with intrinsic growth rate b_i^k and initial size $x_{0,i}^k$. Then by solving (6.1), we have

$$s_i^k = (x_{0,i}^k + c)\exp(b_i^k t_k) - c.$$

For convenience, we reorder $\{s_i^k\}$ in increasing order and reorder $\{b_i^k\}$ in the same order as $\{s_i^k\}$. To avoid introducing more notation, we continue to use the same notation; that is,

$$s_1^k \le s_2^k \le \cdots \le s_{N_s}^k.$$

We then group $s_i^k, i = 1, 2, \ldots, N_s$ into size classes based on Δx. For example, if $|s_1^k - s_2^k| < \Delta x$, then s_1^k and s_2^k are in the same size class. Suppose that we have N_x^k size classes at time t_k after we group $\{s_i^k\}$. We use $[x_j^k, x_j^k + \Delta x)$ to denote size class $j, j = 1, 2, \ldots, N_x^k$ and z_j^k to represent the total number of population in size class j at time t_k (so, $\sum_{j=1}^{N_x^k} z_j^k = N_s, k = 1, 2, \ldots, N_t$). The center point $x_j^k + \Delta x/2$ of size class j is used in the computations to estimate the parameters (μ_b, σ_b) in order to eliminate any left bias by using x_j^k or right bias by using $x_j^k + \Delta x$. Then the estimate $\hat{\theta} = (\hat{\mu}_b, \hat{\sigma}_b)$ of the underlying "true" parameters $\theta_0 = (\mu_b^0, \sigma_b^0)$ (assumed to exist) can be calculated by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} J(\theta) = \arg\min_{\theta \in \Theta} \sum_{k=1}^{N_t} \sum_{j=1}^{N_x^k} |N_{GRD}(t_k, x_j^k + \Delta x/2; \theta) - z_j^k|^2,$$

where Θ is some closed set in $\mathbb{R}_+^2$.

6.2.1 Effect of Sampling Size and Frequency on Parameter Estimation Problem

We carried out a series of inverse problem calculations in order to determine the influence of the sampling size (N_s) and the number of sampling time points (N_t) on the

quality of the parameter estimation. The sampling *frequency* varies from twice a week, once a week to once every two weeks, while sampling *size* N_s varies from 25, 50, 75 to 100. We consider these sampling frequencies and sizes to determine minimal effort experiments to provide data sufficient to accurately estimate variability (i.e., $\theta_0 = (\mu_b^0, \sigma_b^0)$) in the growth rates of the shrimp. A sampling *period* of 6 weeks was used in all of the following simulations. We also considered different σ_b^0 values in order to determine the effect of the amount of variability in the growth rates on the sampling *size* and sampling *frequency* necessary for the experiments. The true standard deviation σ_b^0 used to generate the simulated data was set to $\sigma_{scale} \cdot \mu_b^0$, where $\sigma_{scale} = 0.1$, 0.5, and 0.9. Therefore, we have a total of 36 scenarios. The following parameter values were used in the inverse problem calculations: $\Delta x = 0.01, \quad c = 0.1, \quad \mu_b^0 = 0.045, \quad \sigma_b^0 = \sigma_{scale} \cdot \mu_b^0, \quad \underline{b} = 0.001, \quad \bar{b} = 0.2, \quad \underline{x}_0 = 0, \quad \bar{x}_0 = 0.02$. We used the following function for the initial condition $v_0(x; g)$:

$$v_0(x; g) = \begin{cases} \frac{N_s}{2\varepsilon} & \text{if } x_0 - \varepsilon < x < x_0 + \varepsilon, \\ 0, & \text{otherwise,} \end{cases}$$

where $\varepsilon = 0.01$. The inverse problem calculations were performed 500 times for each scenario with different sets of $\{b_i^k\}_{i=1,k=1}^{N_s,N_t}$ sampled from the truncated normal distribution on $[\underline{b}, \bar{b}]$ with mean μ_b^0 and standard deviation σ_b^0 and $\{x_{0,i}^k\}_{i=1,k=1}^{N_s,N_t}$ sampled from a uniform distribution on $[\underline{x}_0, \bar{x}_0]$. The estimated value for θ_0 in the jth ($j = 1, 2, \ldots, 500$) inverse problem of the lth ($l = 1, 2, \ldots, 36$) scenario is denoted as $\hat{\theta}^{j,l} = (\hat{\mu}_b^{j,l}, \hat{\sigma}_b^{j,l})$.

Figure 6.3 displays the results obtained when $\sigma_{scale} = 0.1$. The upper two plots show the average relative errors of $\hat{\mu}_b$ and $\hat{\sigma}_b$, while the 95% confidence bounds for $\hat{\mu}_b$ and $\hat{\sigma}_b$ are in the lower two plots of Figure 6.3. The average relative errors $RE(\hat{\mu}_b)^l$ and $RE(\hat{\sigma}_b)^l$ of $\hat{\mu}_b$ and $\hat{\sigma}_b$, respectively, in each scenario is given by

$$RE(\hat{\mu}_b)^l = \frac{1}{500} \sum_{j=1}^{500} \frac{|\hat{\mu}_b^{j,l} - \mu_b^0|}{\mu_b^0} \quad \text{and} \quad RE(\hat{\sigma}_b)^l = \frac{1}{500} \sum_{j=1}^{500} \frac{|\hat{\sigma}_b^{j,l} - \sigma_b^0|}{\sigma_b^0}, \quad \text{for } l = 1, 2, \ldots, 36.$$

We obtain the 95% confidence bounds for $\hat{\mu}_b^{j,l}$ and $\hat{\sigma}_b^{j,l}$, $j = 1, \ldots 500$, for each scenario by computing the value where 2.5% of the estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ are below and above based on the results of the 500 inverse problem calculations. While the sampling *size* N_s and sampling *frequency* do not have an effect on the estimation of μ_b in this example, we observe that there is an effect on the estimation of σ_b. The average relative error of $\hat{\sigma}_b$ decreases as we increase either N_s or N_t. Furthermore, the reduction in the average relative

error resulting from an increase in the sampling *size* N_s from 25 to 50 is greater than the reduction caused by an increase in N_s from 50 to 75 and 75 to 100. These results suggest that a sampling *size* of 25 is not large enough to obtain reliable estimates of the variability in the growth rates.

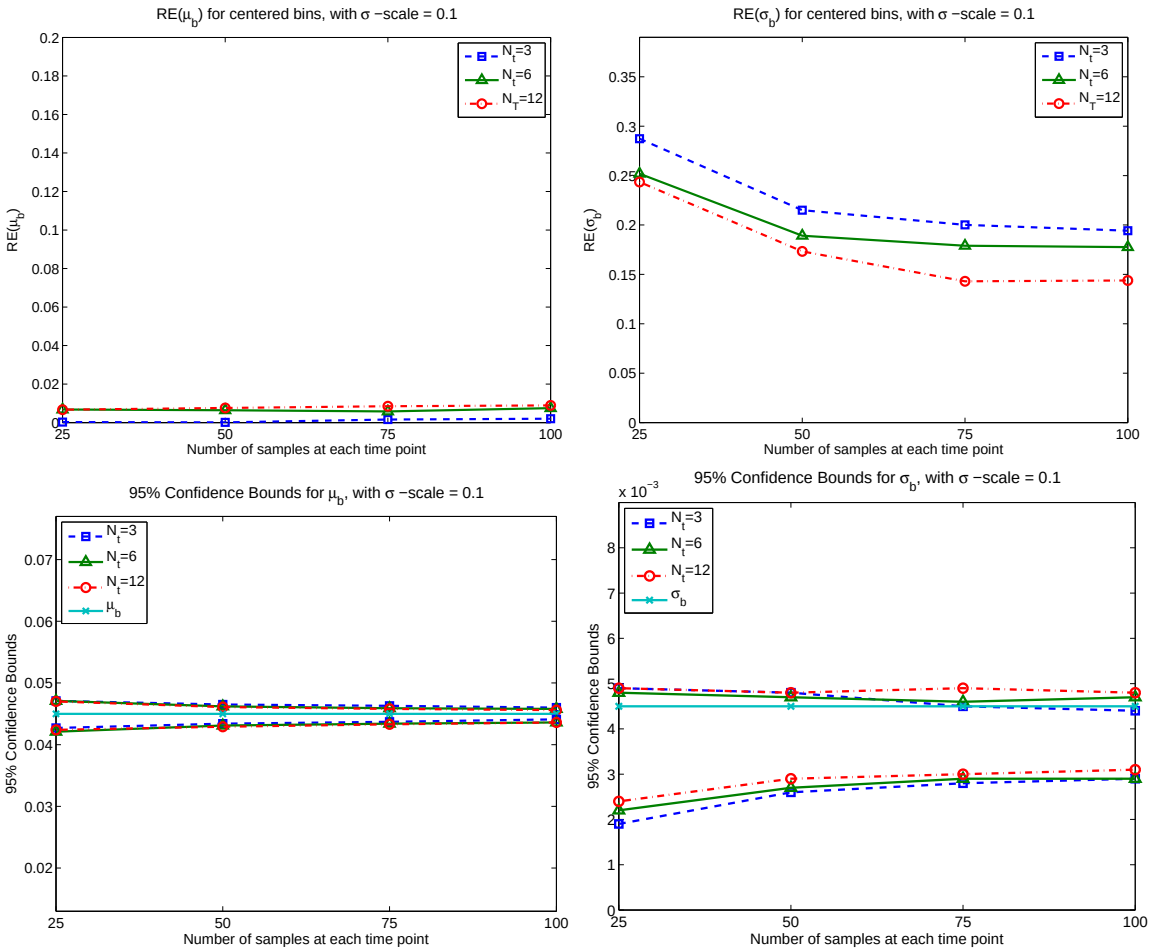

Figure 6.3: (upper): Relative errors of $\hat{\mu}_b$ and $\hat{\sigma}_b$; (lower): 95% confidence bounds for the estimates corresponding to $\hat{\mu}_b$ and $\hat{\sigma}_b$ with $\sigma_{scale} = 0.1$.

The lower plots in Figure 6.3 depicting the 95% confidence bounds for $\hat{\mu}_b$ also show that for this example with $\sigma_{scale} = 0.1$ the sampling *size* and *frequency* have minimal effect on the estimates $\hat{\mu}_b$ based on the very tight bounds around the "true" mean μ_b^0. However, we observe a decrease in the width of the confidence bounds for $\hat{\sigma}_b$ as N_s is increased from 25 to 50 and essentially no change in the width as N_s is increased from 50 to 75 and 75 to 100. This also suggests that $N_s = 25$ is not enough to obtain a reliable estimate of σ_b.

A σ_{scale} value of 0.5 was used in the simulations that produced the results in Figure 6.4. We note in this case that the estimates of μ_b and σ_b are both affected by the sampling *size* and *frequency*. The average relative error of $\hat{\mu}_b$ and $\hat{\sigma}_b$ both decrease as N_s is increased. The decrease in the average relative error is very significant as N_s is increased from 25 to 50 for both parameters; however, the reduction in the average relative error for both parameters is not as dramatic when N_s is increased from 50 to 75 and from 75 to 100. We would again infer from these results that a sampling size of at least 50 is necessary in order to obtain reliable estimates of (μ_b, σ_b).

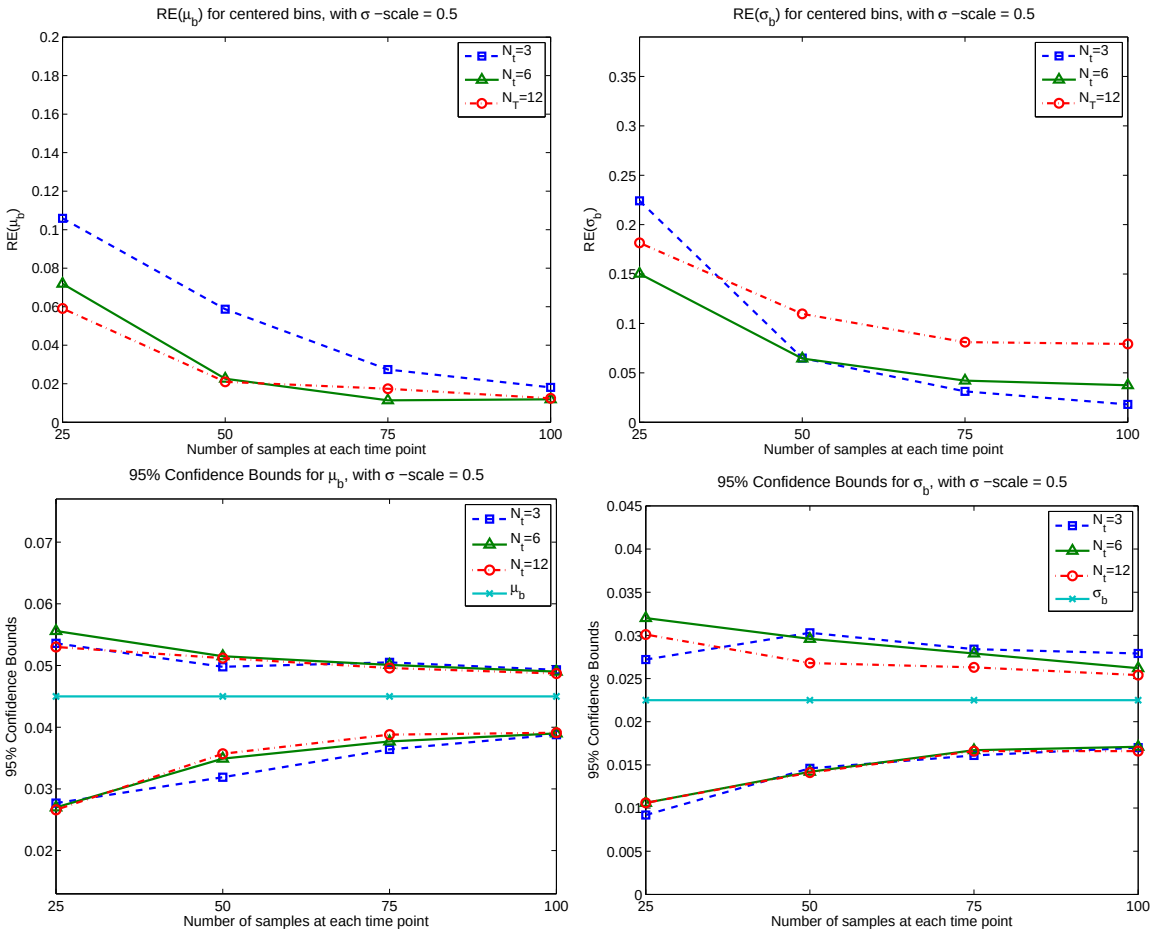

Figure 6.4: (upper): Relative errors of $\hat{\mu}_b$ and $\hat{\sigma}_b$; (lower): 95% confidence bounds for the estimates corresponding to $\hat{\mu}_b$ and $\hat{\sigma}_b$ with $\sigma_{scale} = 0.5$.

The resulting 95% confidence bounds for $\hat{\mu}_b$ and $\hat{\sigma}_b$ when $\sigma_{scale} = 0.5$ are shown in the lower plots of Figure 6.4. As already noted, the estimation of both parameters depends

on N_s and N_t. The decrease in the width of the confidence bounds for $\hat{\mu}_b$ is more significant as N_s is increased from 25 to 50 in comparison to the decrease in the width when N_s is increased from 50 to 75 and 75 to 100. We also observe the same type of behavior in the width of the confidence bounds for $\hat{\sigma}_b$. The reduction in the width of the confidence bounds for $\hat{\sigma}_b$ is much smaller as N_s is increased from 50 to 75 and 75 to 100 versus the decrease in width as N_s is increased from 25 to 50. While a sampling *size* N_s of 25 appears to be too small, a sampling *size* of 50 seems to be sufficient to obtain reasonably accurate estimates of the variability in the growth rates.

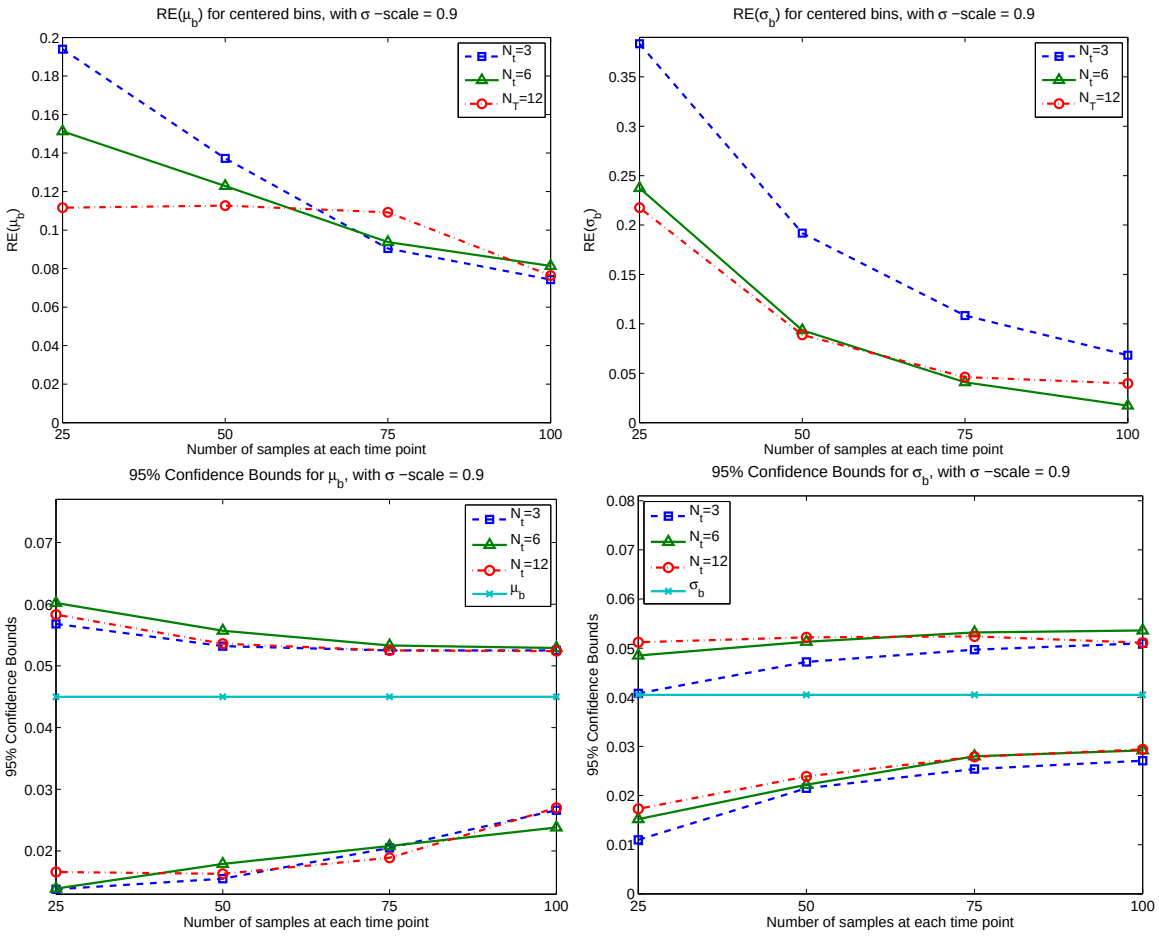

Figure 6.5: (upper): Relative errors of $\hat{\mu}_b$ and $\hat{\sigma}_b$; (lower): 95% confidence bounds for the estimates corresponding to $\hat{\mu}_b$ and $\hat{\sigma}_b$ with $\sigma_{scale} = 0.9$.

The final sets of results in Figure 6.5 were computed with $\sigma_{scale} = 0.9$. The average relative errors for $\hat{\mu}_b$ and $\hat{\sigma}_b$ depicted in the upper plots in Figure 6.5 show that the reliability

of both parameters is influenced by the sampling *size* and *frequency.* As the previous results suggested, a sampling *size* of 25 is not adequate for the parameter estimation problem based on the magnitude of the average relative error for $N_s = 25$ in comparison to the average relative errors obtained for $N_s = 50, 75$, and 100. We also note that the decrease in the average relative error of $\hat{\sigma}_b$ is not as large when N_s is increased from 50 up to 100 in comparison to a much greater decrease when N_s is increased from 25 to 50. These results suggest that there is not a significant improvement in the reliability of the parameter estimates when a sampling *size* larger than 50 is used. We also observe from the lower plots in Figure 6.5 that the effect of N_s on the width of the 95% confidence bounds of $\hat{\mu}_b$ and $\hat{\sigma}_b$ decreases as N_t is increased. As in the previous two cases considered, the width of the confidence bounds for both parameters decreases the most when N_s is increased from 25 to 50 which also seems to suggest that reliable estimates of (μ_b, σ_b) can be obtained with a sampling *size* N_s of 50.

Based on these simulations, we conclude that the sampling *size* N_s has an effect on the reliability of the results of the parameter estimation problem. When the standard deviation σ_b^0 is small relative to the mean μ_b^0, the effect on the accuracy of the estimates $\hat{\mu}_b$ is minimal; however, this effect is more noticeable as the magnitude of σ_b^0 is increased. The influence of the sampling *size* N_s is much more evident in the estimation of σ_b for each of the values considered here. We also conclude that the results when sampling once a week are comparable to those obtained when sampling twice a week. The most desirable experiment involved using a sampling *size* N_s of 100 once a week; however, there appears to be little loss in accuracy in estimating the variability in growth rates if one uses a sample *size* $N_s = 50$. Finally, our computations suggest that experiments with a sampling *size* of only $N_s = 25$ would not be adequate even when one increases the sampling *frequency* to twice a week.

6.2.2 Effect of Sampling Size and Bin Size on Model Predictions and Parameter Estimates

We also carried out numerical simulations with simulated data to investigate the effect of the bin size Δx in relation to the sampling *size* N_s on the parameter estimation results as well as the model predictions. In this section we also compare the results from computations with a fixed bin size versus variable bin sizes (fixed number of bins N_b) for

a fixed sampling *size* in the estimation problem. Since the "true" growth rate distribution of shrimp observed in experiments is unknown, the accuracy of the parameter estimates is determined by how well the model fits the data. The discussion of the results for the inverse problem with simulated data in this section will emphasize the importance in choosing the appropriate bin size Δx with respect to the sampling *size* N_s in the inverse problem calculations.

In the first set of results we used a fixed Δx with two different N_s values to illustrate the significance of using the proper bin size relative to the sample *size* for accurate model predictions. Using simulated data generated with a truncated normal distribution with mean $\mu_b^0 = 0.045$ and standard deviation $\sigma_b^0 = 0.5 \cdot \mu_b^0 = 0.0225$, we obtained the following results from the inverse problem with a fixed bin size Δx of 0.01. A sampling *size* N_s of 50 was used with a sampling *frequency* of once a week for six weeks. Using our inverse problem techniques, we computed good parameter estimates $(\hat{\mu}_b, \hat{\sigma}_b)$ with small relative errors (see Table 6.1) which would suggest that the model does a good job of fitting the data. However, the model does not provide good fits to the data as time progresses based on the plots shown in Figure 6.6. The model appears to do a good job of predicting the population data during the first two weeks but fails to predict the heights seen in the data during the final four weeks. It is clear from Figure 6.6 that the model does a good job qualitatively; however, as a result of the quantitative comparisons of the model predictions to the data, the model fits do not support our initial findings based on the numerical results in Table 6.1.

Table 6.1: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and relative errors for inverse problem with simulated data with $N_s = 50$, $b \sim \mathcal{N}_{[0.001,0.2]}(0.0450, 0.0225^2)$, and $\Delta x = 0.01$.

N_s	$\hat{\mu}_b$	$RE(\hat{\mu}_b)$	$\hat{\sigma}_b$	$RE(\hat{\sigma}_b)$
50	0.04817	0.07044	0.02400	0.06667

We then generated simulated data with a sampling *size* of 5000 and a truncated normal distribution on the intrinsic growth rates with mean $\mu_b^0 = 0.045$ and standard deviation $\sigma_b^0 = 0.0225$. The estimated parameters along with the corresponding relative errors that were computed with the fixed bin size $\Delta x = 0.01$ are in Table 6.2.

Table 6.2: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and relative errors for inverse problem with simulated data with $N_s = 5000$, $b \sim \mathcal{N}_{[0.001,0.2]}(0.0450, 0.0225^2)$, and $\Delta x = 0.01$.

N_s	$\hat{\mu}_b$	$RE(\hat{\mu}_b)$	$\hat{\sigma}_b$	$RE(\hat{\sigma}_b)$
5000	0.04495	0.00111	0.02253	0.00133

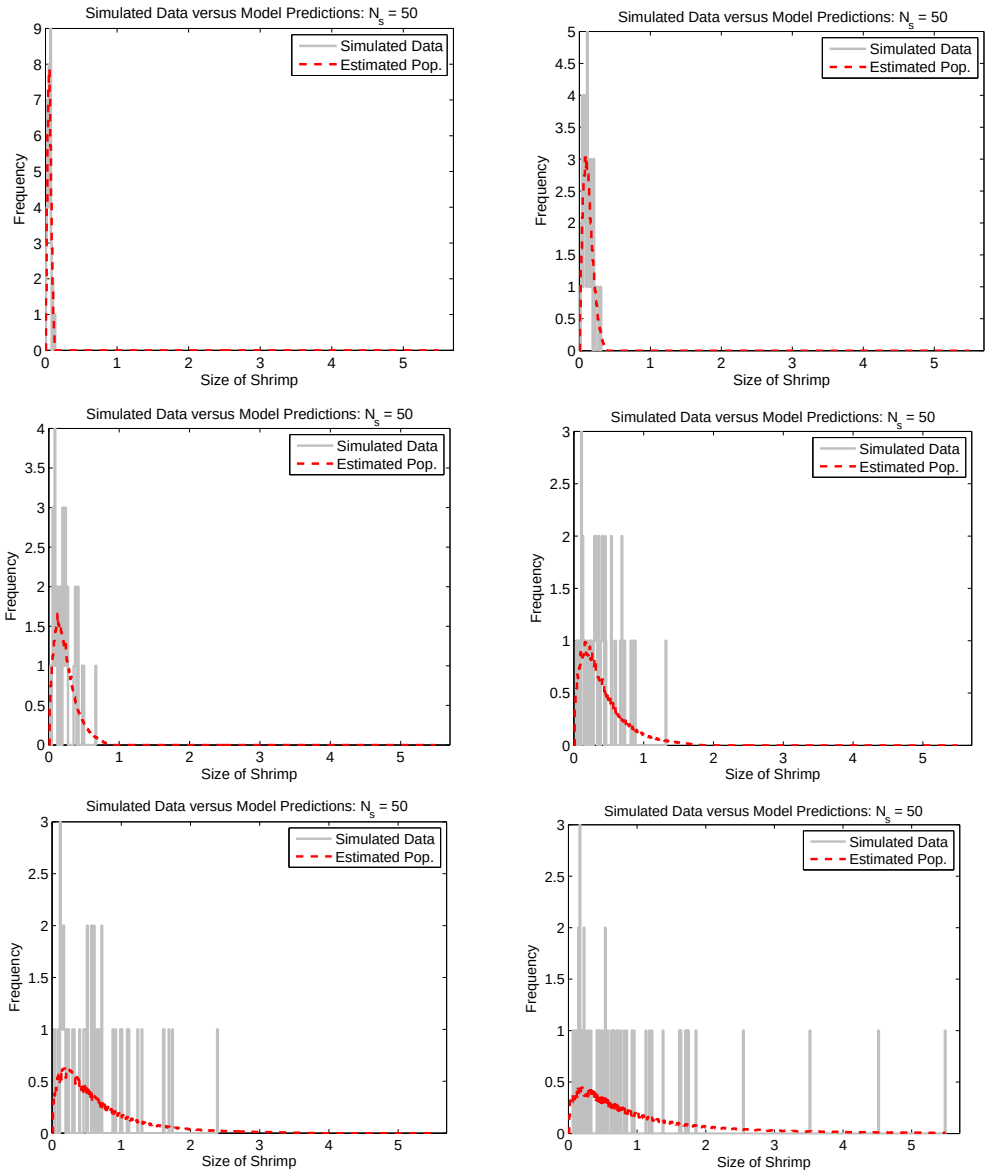

Figure 6.6: Simulated population data, snapshot z_j^k, and model predicted population $N_{GRD}(t, x; \hat{\theta})$ with optimal parameter $\hat{\theta}$ for $t = 7,\ 14,\ 21,\ 28,\ 35,\ 42$ days with sample size $N_s = 50$.

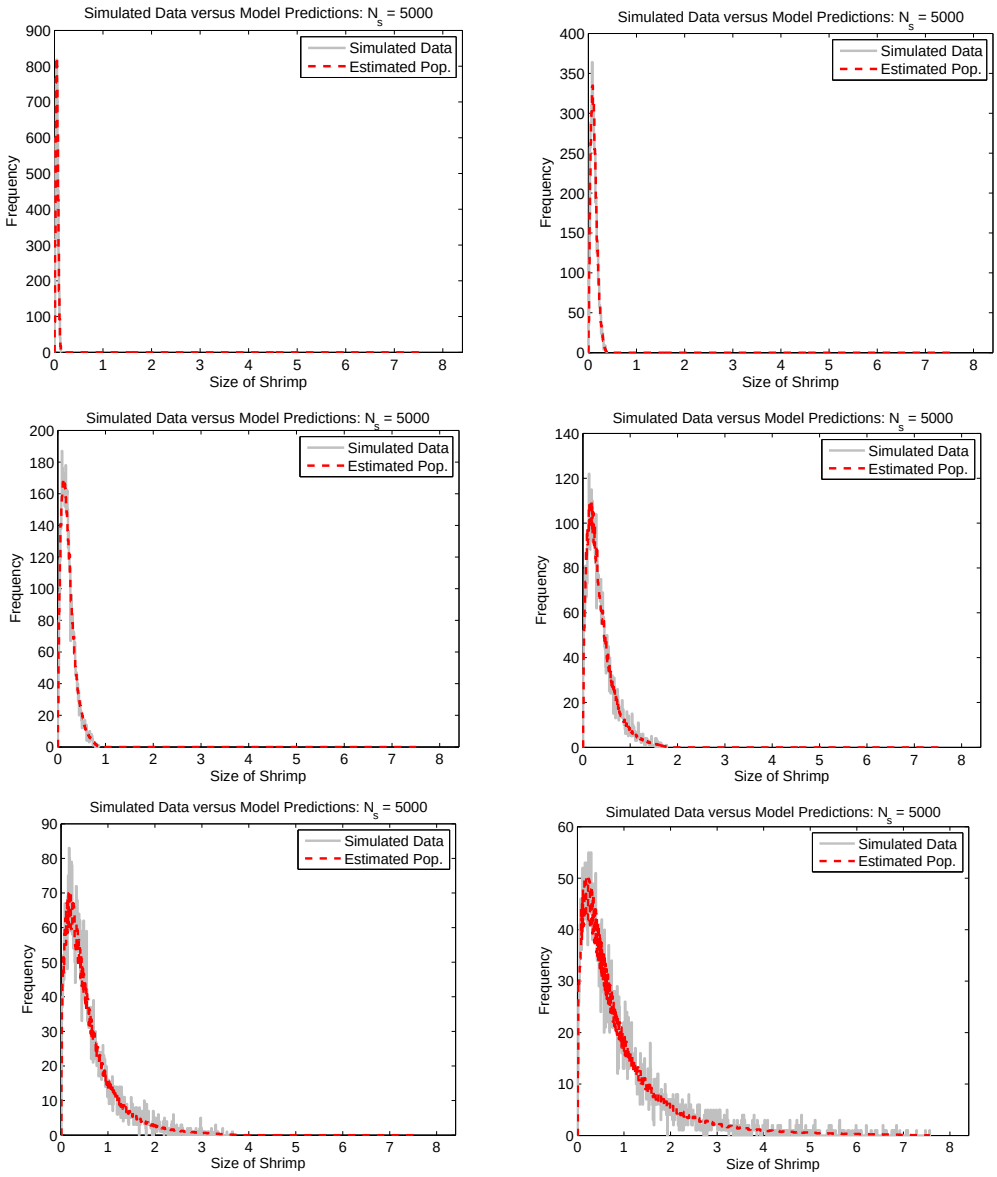

Figure 6.7: Simulated population data, snapshot z_j^k, and model predicted population $N_{GRD}(t, x; \hat{\theta})$ with optimal parameter $\hat{\theta}$ for $t = 7,\ 14,\ 21,\ 28,\ 35,\ 42$ days with sample size $N_s = 5000$.

Based on the very small relative errors, we would also conclude in this example that we have obtained good estimates of the true parameters (μ_b^0, σ_b^0). Moreover, we would assume that the model predictions would fit the data very well. Figure 6.7 shows the model predictions versus the simulated data for six weeks with a sampling *size* of 5000. We see that the model does a very good job both qualitatively and quantitatively. Unlike the previous example with $N_s = 50$, the model provides good fits to the entire data set when $N_s = 5000$.

These results demonstrate the effect of the bin size relative to the sampling *size* N_s on the accuracy of the model predictions. While the parameter estimates obtained for (μ_b^0, σ_b^0) with both sampling sizes were good, the model predictions with a sample *size* of 50 were not as accurate as those obtained with a sample *size* of 5000. As noted earlier, the accuracy of $(\hat{\mu}_b, \hat{\sigma}_b)$ will depend on the model fits to data when experimental data is used since the "true" underlying distribution of the growth rates is unknown and relative errors can not be computed. Therefore, using an appropriately chosen bin size Δx is critical in the parameter estimation problem with experimental data.

To further investigate the effect of Δx on the accuracy of the model predictions and parameter estimates, we considered using a fixed bin size Δx versus a fixed number of bins N_b (variable bin size) in the parameter estimation problem with simulated data for a fixed sampling *size*. A sampling *size* N_s of 50 with a sampling *frequency* of once a week for six weeks is used in the following results. The simulated data was generated with a truncated normal distribution with mean $\mu_b^0 = 0.045$ and standard deviation $\sigma_b^0 = 0.0225$. Using the same data set, we computed results with a fixed bin size $\Delta x = 0.01$ and with a fixed number of bins $N_b = 10$. The bin size for each time point when a fixed number of bins is used is given by

$$\Delta x(t_k) = \frac{\bar{x}(t_k) - \underline{x}(t_k)}{N_b},$$

where $\bar{x}(t_k)$ is the largest size sampled at time t_k and $\underline{x}(t_k)$ is the smallest size sampled at time t_k. The parameter estimates $(\hat{\mu}_b, \hat{\sigma}_b)$ along with the relative errors for both scenarios are given in Table 6.3. While we obtain comparable estimates of the mean μ_b^0, we note that the estimated standard deviation $\hat{\sigma}_b$ is better when using a fixed number of bins versus a fixed bin size. We observe from the model predictions versus the data in Figure 6.8 that as time evolves the model is not able to predict the peaks present in the data when the fixed bin size $\Delta x = 0.01$ is used. However, the model predictions obtained with the fixed number of bins $N_b = 10$ (varying bin size) are more accurate as seen in Figure 6.9. The model does

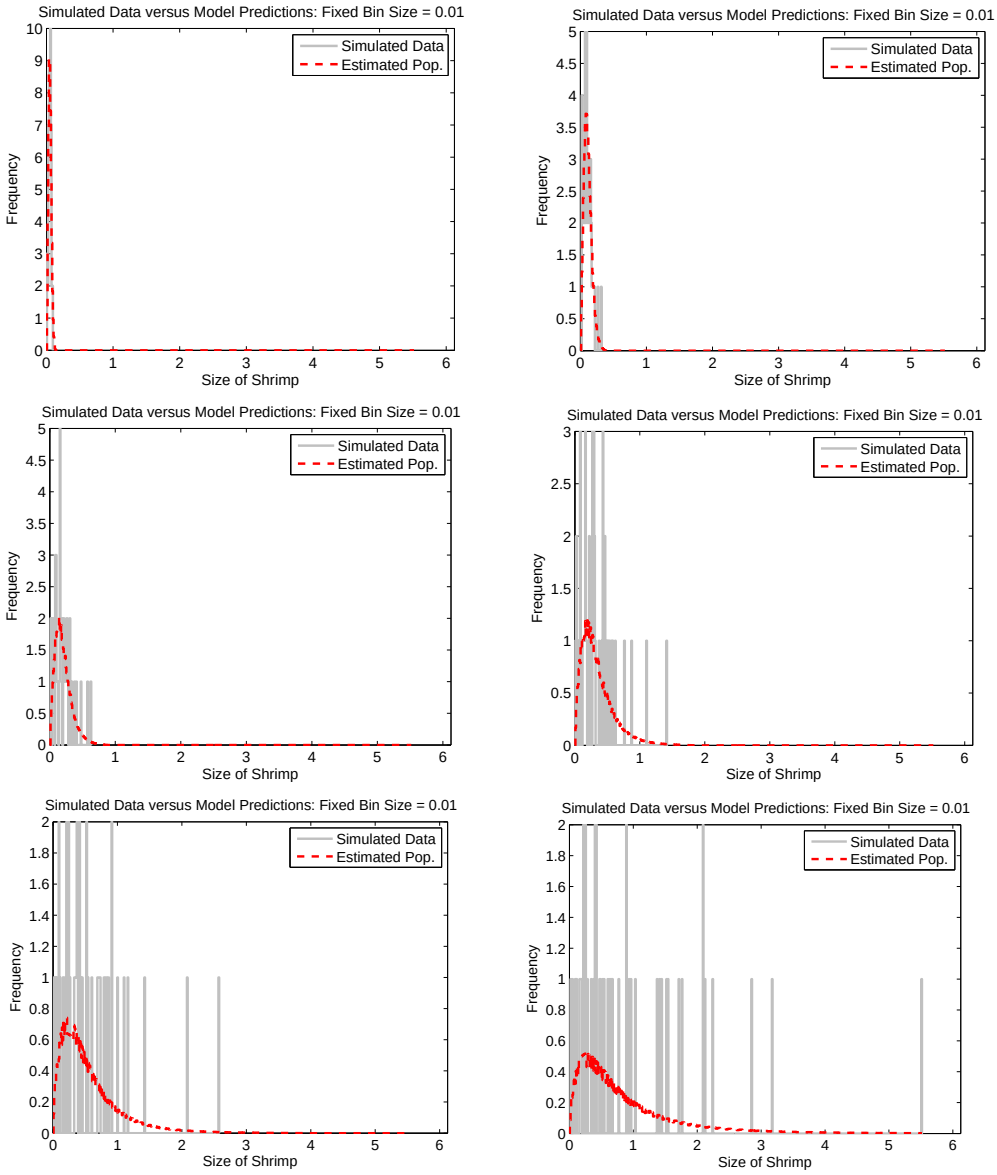

Figure 6.8: Simulated population data, snapshot z_j^k, and model predicted population $N_{GRD}(t, x; \hat{\theta})$ with optimal parameter $\hat{\theta}$ for $t = 7$, 14, 21, 28, 35, 42 days with fixed bin size $\Delta x = 0.01$.

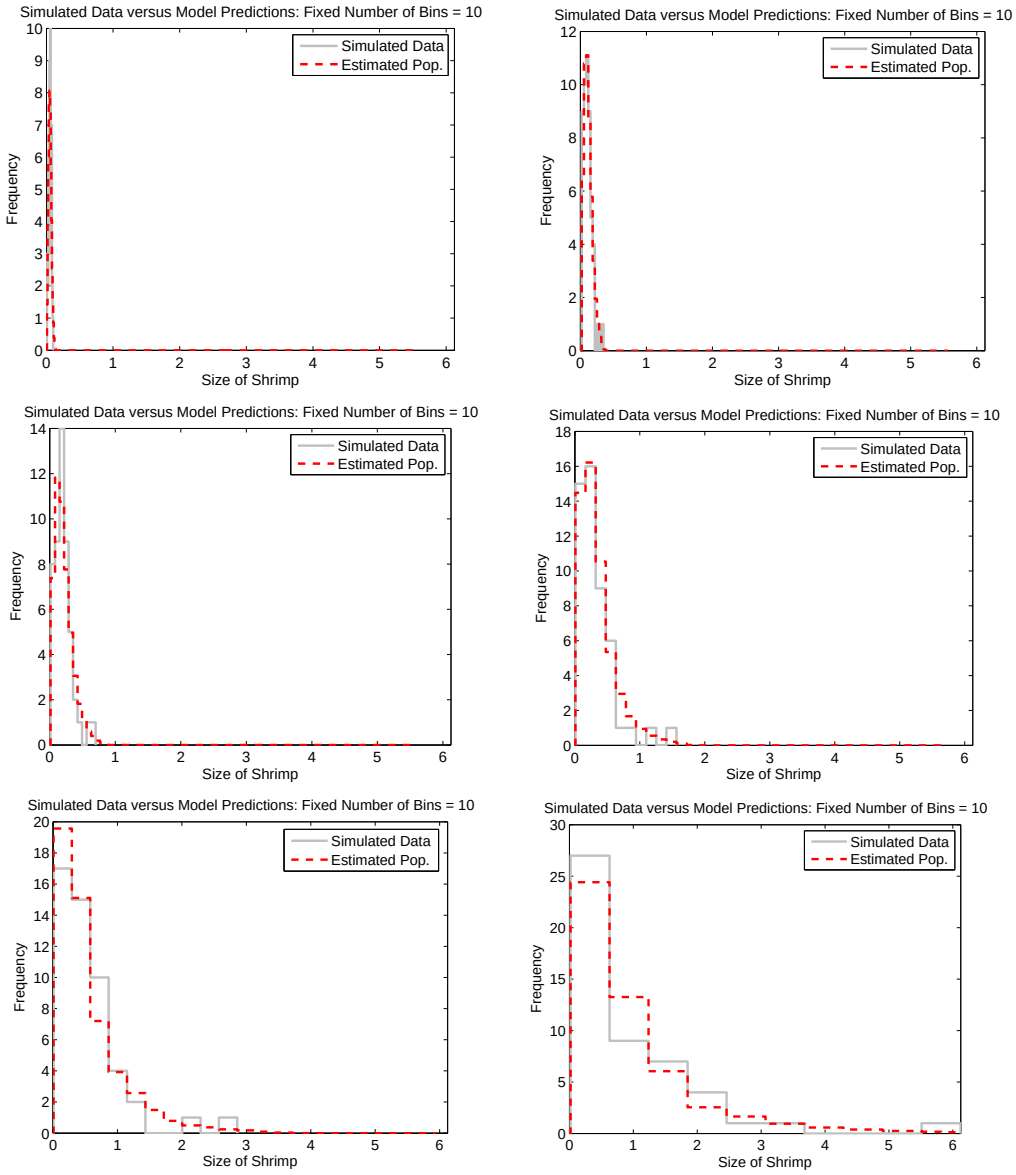

Figure 6.9: Simulated population data, snapshot z_j^k, and model predicted population $N_{GRD}(t, x; \hat{\theta})$ with optimal parameter $\hat{\theta}$ for $t = 7, 14, 21, 28, 35, 42$ days with fixed number of bins $N_b = 10$.

a much better job of fitting the data when a fixed number of bins are used versus a fixed bin size. We infer from the results shown in Figure 6.8 that the fixed bin size $\Delta x = 0.01$ is too small which leads to the increased number of size classes with a frequency equal to one seen in the simulated data that are not present with the fixed number of bins $N_b = 10$.

Table 6.3: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and relative errors for inverse problem with simulated data with $b \sim \mathcal{N}_{[0.001,0.2]}(0.045, 0.0225^2)$.

	$\hat{\mu}_b$	$RE(\hat{\mu}_b)$	$\hat{\sigma}_b$	$RE(\hat{\sigma}_b)$
$\Delta x = 0.01$	0.04471	0.00644	0.01965	0.12667
$N_b = 10$	0.04485	0.00333	0.02146	0.04622

Similar results were obtained when we repeated the inverse problem calculations 500 times. Using a fixed bin size that is too small results in model predictions that are not very accurate during the early growth process of the shrimp when the sampling *size* N_s is relatively small. Although the plots in Figure 6.9 demonstrate the benefits of fixing the number of bins, one must still be careful when selecting the number of bins to use in computations because choosing a value that is too small or too large can bias the data as well and affect the model predictions. Furthermore, we also found that a larger fixed bin size Δx can result in better model predictions. The results shown here demonstrate the importance in using the appropriate number of bins N_b or bin size Δx in ratio to the sampling *size* N_s to obtain good model fits and parameter estimates in inverse problem calculations.

6.3 Parameter Estimation Results with Experimental Data

We present the results of the inverse problem for the estimation of the growth rate distribution of a size-structured shrimp population using data collected from ABN in this section. The experiments were based on the design calculations discussed in Section 6.2. Data was collected during the early growth of shrimp initially in the post-larval stage starting on December 4, 2007. Fifty shrimp were randomly sampled and measured once a week under relatively constant tank conditions (density, temperature, feeding, etc.) through December 31, 2007 resulting in a data set with five longitudinal size distribution observations. As noted earlier, we must compare the experimental data to the estimated population data generated by using the optimal parameter $\hat{\theta}$ in (6.3) since the actual growth rate distribution of the shrimp observed is unknown. The following results were obtained

via computations carried out in MATLAB with

$$\Delta x = 0.02, \quad \underline{b} = 10^{-6}, \quad \bar{b} = 10^{-1}.$$

Before presenting the results, we define the parameters from the asymptotic standard error theory for finite-dimensional parameters [36, 40, 45, 60] which we outlined in detail in Chapter 3. Recall the OLS estimator $\theta_{OLS}(Y)$ used in the inverse problem for the estimation of $\theta = (\mu_b, \sigma_b)$ is given by

$$\theta_{OLS}(Y) \equiv \arg\min_{\theta \in \Theta} \sum_{j=1}^{n} (Y_j - f(\tau_j; \theta))^2.$$

Since the observations Y_j are random variables, then the OLS estimator $\theta_{OLS}(Y)$ is also a random variable, and as $n \to \infty$, the sampling distribution for $\theta_{OLS}(Y)$ is a multivariate normal distribution with mean $\mathcal{E}[\theta_{OLS}(Y)] \approx \theta_0$ and covariance $Var[\theta_{OLS}(Y)] \approx \Sigma_0^n \approx \sigma_0^2[\mathcal{X}^T(\theta_0)\mathcal{X}(\theta_0)]^{-1}$. The $n \times 2$ sensitivity matrix $\mathcal{X}(\theta)$ has entries

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\tau_j; \theta)}{\partial \theta_k}.$$

For the application considered here, we note that we are able to compute the entries of the sensitivity matrix explicitly. Since

$$f(\tau_j; \theta_0) \approx N_{GRD}(\tau_j; \theta),$$

the entries in the sensitivity matrix $\mathcal{X}(\theta)$ are given by

$$\begin{aligned} \mathcal{X}_{jk}(\theta) = \frac{\partial f(\tau_j; \theta)}{\partial \theta_k} &= \frac{\partial}{\partial \theta_k} N_{GRD}(\tau_j; \theta) \\ &\approx \Delta x \int_{\underline{b}}^{\bar{b}} v(\tau_j; g) \frac{\partial p(b; \theta)}{\partial \theta_k} db, \end{aligned}$$

where $\theta = (\mu_b, \sigma_b)$ and $p(b; \theta)$ is the probability density function (pdf) for the truncated normal distribution. We also chose to use $\alpha = 0.05$ which corresponds to a significance level of 95% and a $t_{1-\alpha/2}$-value of approximately 1.96 when n is greater than or equal to 30.

6.3.1 Results with Complete December Data Set

In the first set of results we considered the entire data set collected in December, and we began by fitting the exponential function

$$\bar{x}(t) = a \exp(bt) - c \tag{6.4}$$

to the data $\{(t_k, \bar{x}_d^k)\}$ where a, b, and c are all unknown parameters and $\bar{x}_d^k$ is the average size of shrimp observed at time t_k. This exponential function was determined by solving the deterministic growth model of the form

$$\frac{d\bar{x}(t)}{dt} = b(\bar{x}(t) + c), \tag{6.5}$$

where b and c are assumed to be fixed constants. The optimal parameters corresponding to the fit with (6.4) shown in Figure 6.10 are

$$\hat{a} = 0.049, \quad \hat{b} = 0.043, \quad \hat{c} = 0.036.$$

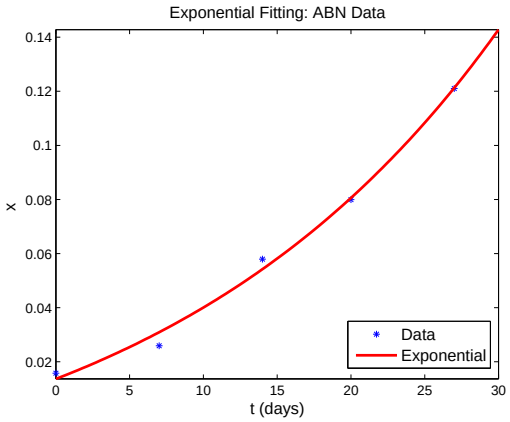

Figure 6.10: Exponential fit of (6.4) to complete ABN average size shrimp data set with $\frac{d\bar{x}}{dt} = 0.043(\bar{x} + 0.036)$.

We then used the optimal value $\hat{c}$ in our inverse problem calculations for the estimation of (μ_b, σ_b) with the entire December 2007 data. The data from December 4, 2007 (see Figure 6.11) is interpolated and used as an approximation for the initial size density $v_0(x; g)$ in the SS model (2.1). Therefore, it can not be used in the estimation of the growth rate distribution, and we are left with four data sets to use in the inverse problem.

The estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ along with the confidence intervals for each of these parameters are in Table 6.4. These results have an optimal cost $J(\hat{\theta})$ of 574.8315, and the estimated variance of the statistical model is $\hat{\sigma}^2 = 17.4191$. We observe from the plots in Figure 6.11 that the model predictions do not fit the data well. The model does a decent job of fitting the data on December 11 but is shifted to the left of the observations for the final three time points. Moreover, the estimates $(\hat{\mu}_b, \hat{\sigma}_b)$ are not very reliable based on the magnitude of the associated confidence intervals relative to the estimated parameter values.

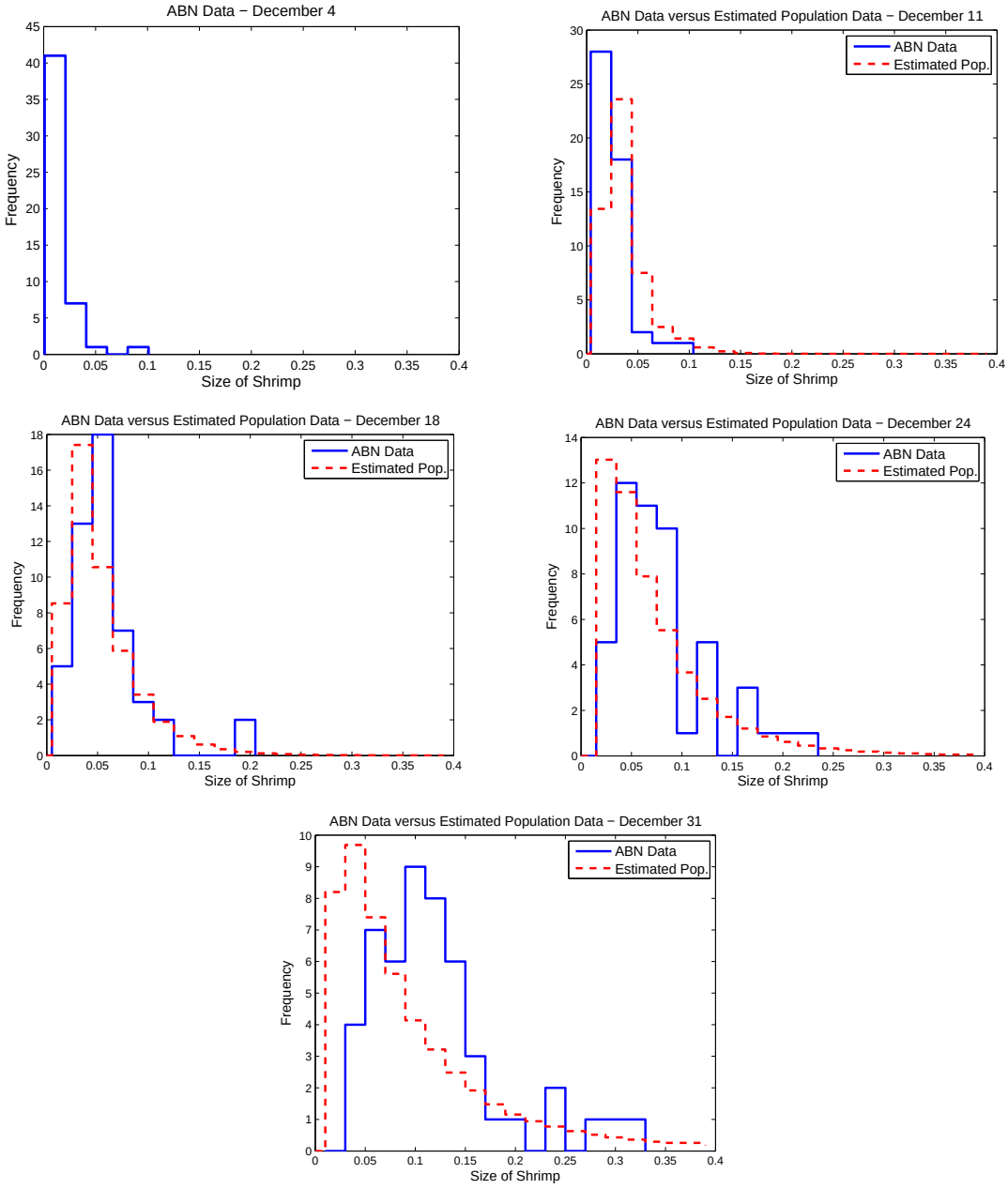

Figure 6.11: Longitudinal data on shrimp weight from ABN on December 4, 11, 18, 24, and 31 versus model predicted population density $N_{GRD}(t, x; \hat{\theta})$ with $\hat{\theta} = (0.0010, 0.0324)$ and $\hat{c} = 0.036$ from exponential fit of (6.4) to average size shrimp data.

Table 6.4: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and confidence intervals for inverse problem with complete December data set using $\hat{c} = 0.036$ from exponential fit of (6.4) to average size shrimp data.

$\hat{\mu}_b \pm 1.96SE(\hat{\mu}_b)$	$\hat{\sigma}_b \pm 1.96SE(\hat{\sigma}_b)$
0.0010 ± 0.0535	0.0324 ± 0.0313

In [10] the authors outline and compare the two previously mentioned approaches (a probabilistic formulation corresponding to the Growth Rate Distribution model versus a stochastic formulation corresponding to the Fokker-Planck model) for modeling variability in the growth process of size-structured populations. As established in [10], the population density from these two approaches is the same when comparable levels of uncertainty are used in both formulations. Hence, we chose to use the computationally "easier" GRD model (2.3) for our purposes. The shortcomings of the deterministic growth model (6.5) in modeling populations exhibiting a great deal of variability was also discussed in [10] with several different approaches for including uncertainty in the growth process also being described. One such approach is based on considering the following probabilistic growth model

$$\frac{dx(t;b)}{dt} = b(x(t;b) + c), \quad b \in \mathfrak{B}, \tag{6.6}$$

where a distribution is placed on the intrinsic growth rates taking values in $\mathfrak{B}$ and c is assumed to be a fixed affine growth term. If a normal distribution with mean μ_b and variance σ_b^2 is used to describe the distribution of the intrinsic growth rates, then the exponential function corresponding to probabilistic growth model (6.6) is given by

$$x(t;b) = (x(0;b) + c)\exp(bt) - c, \tag{6.7}$$

where $b \sim \mathcal{N}(\mu_b, \sigma_b^2)$. Moreover, it was shown that the probabilistic growth model results in the following exponential function for the mean of the stochastic process for the size of the individuals

$$\bar{x}(t) = (\bar{x}(0) + c)\exp\left(\mu_b t + \frac{1}{2}\sigma_b^2 t^2\right) - c \tag{6.8}$$

which corresponds to the time-varying mean growth dynamics

$$\frac{d\bar{x}(t)}{dt} = (\mu_b + \sigma_b^2 t)(\bar{x}(t) + c). \tag{6.9}$$

As an alternative to the deterministic form (6.5), we also chose to fit the average size shrimp data with (6.8) where μ_b, σ_b, and c are the parameters to be estimated and $\bar{x}(0)$ is

the average size shrimp data for the initial start time. The fit, shown in Figure 6.12, was generated with the following estimated parameters:

$$\hat{\mu}_b = 0.045, \quad \hat{\sigma}_b = 0.003, \quad \hat{c} = 0.029.$$

The data from December 4 was used as an approximation for the initial size density $v_0(x; g)$ in these computations as well. We used the estimated values of $\hat{\mu}_b$ and $\hat{\sigma}_b$ from the exponential fit to the average size shrimp data as initial starting points for the optimization calculations to determine the growth rate distribution of the shrimp population. The optimal value of c was also used in our inverse problem calculations for the growth rate distribution parameters.

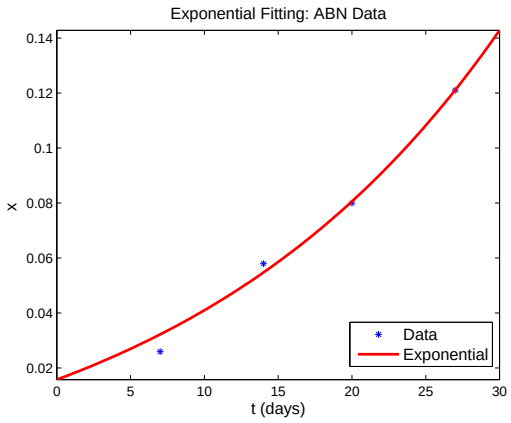

Figure 6.12: Exponential fit of (6.8) to complete ABN average size shrimp data set with $\frac{d\bar{x}}{dt} = (0.045 + 0.003^2 t)(\bar{x} + 0.029)$.

The optimal cost corresponding to the optimized parameters in Table 6.5 is 563.9064. We computed a value of 17.0881 for the estimate of the variance σ_0^2 of the statistical model. The results in this case are very similar to those in the previous case as can be seen by the model fits to data shown in Figure 6.13. We observe the same behavior in the model predictions versus the observations in the final three time points with the model predictions being shifted to the left of the data. The width of the confidence intervals for the estimates $(\hat{\mu}_b, \hat{\sigma}_b)$ also demonstrate how unreliable these parameter values are as well.

It is clear from the results in this section that the model is not able to provide a good fit to the entire data set from December. While the model predictions do exhibit the lognormal behavior characteristic of the observations, the model does a poor job of fitting the data quantitatively. The exponential fits to the average size shrimp data in Figures

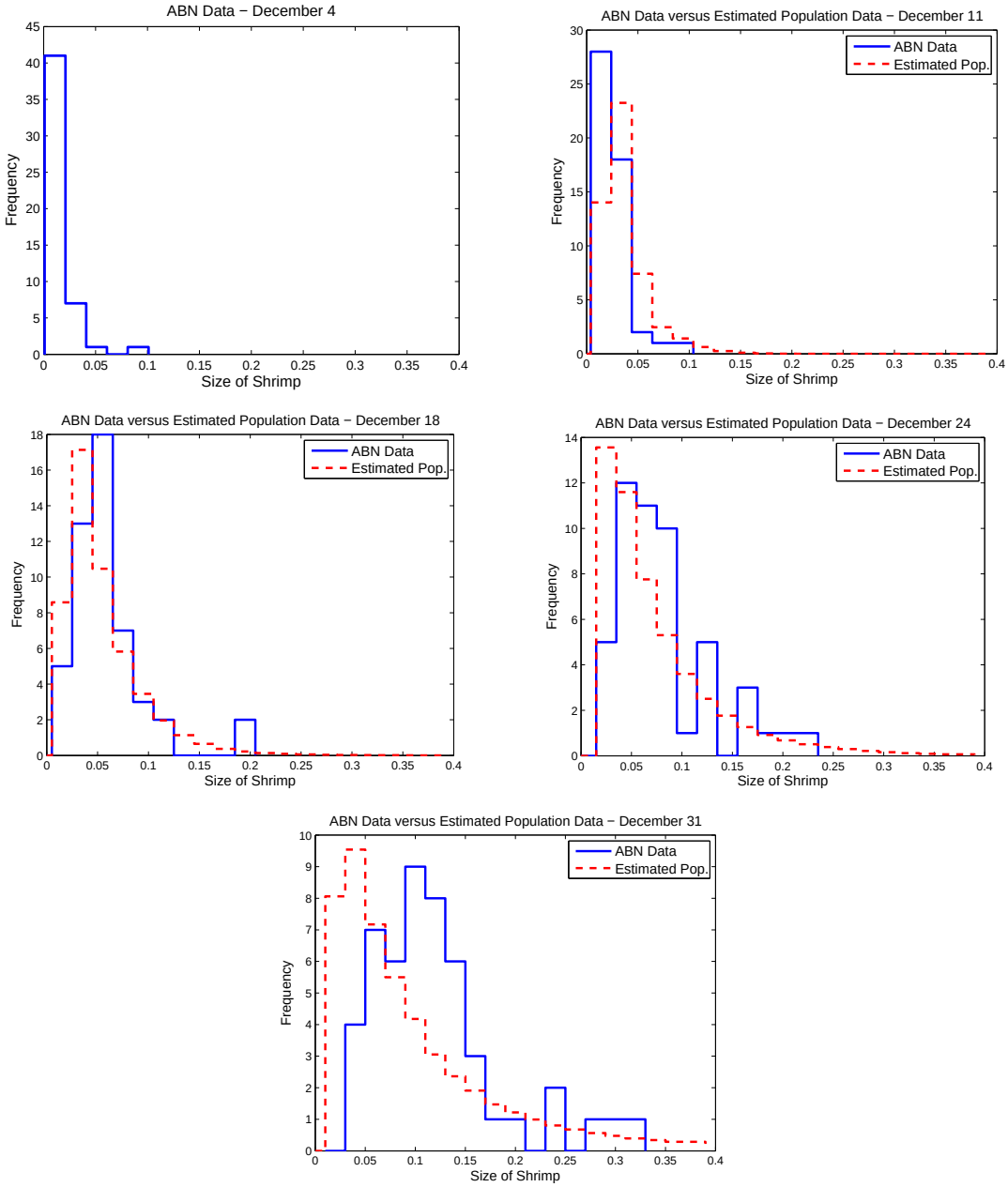

Figure 6.13: Longitudinal data on shrimp weight from ABN on December 4, 11, 18, 24, and 31 versus model predicted population density $N_{GRD}(t, x; \hat{\theta})$ with $\hat{\theta} = (0.0018, 0.0363)$ and $\hat{c} = 0.029$ from exponential fit of (6.8) to average size shrimp data.

Table 6.5: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and confidence intervals for inverse problem with complete December data set using $\hat{c} = 0.029$ from exponential fit of (6.8) to average size shrimp data.

$\hat{\mu}_b \pm 1.96SE(\hat{\mu}_b)$	$\hat{\sigma}_b \pm 1.96SE(\hat{\sigma}_b)$
0.0018 ± 0.0627	0.0363 ± 0.0377

6.10 and 6.12 suggest that the problem in fitting the data could potentially be a result of the data sampled on December 11. The average size shrimp data point for December 11 lies the farthest away from the exponential curves produced by (6.4) and (6.8). Therefore, we reconsidered the inverse problem for the estimation of the growth rate distribution for the shrimp population with the data from December 4, 18, 24 and 31 in an effort to obtain better parameter estimates and model fits to the data.

6.3.2 Results Excluding December 11 Data

In an attempt to provide better fits to the shrimp population data, we excluded the data from December 11 in the parameter estimation problem for the growth rate distribution. The exponential function (6.4) generated by the deterministic growth model was used to fit the average size shrimp data for December 4, 18, 24, and 31. The estimated parameters generating the solid line shown in Figure 6.14 are

$$\hat{a} = 0.043, \quad \hat{b} = 0.046, \quad \hat{c} = 0.026.$$

We note from Figure 6.14 that (6.4) provides a good fit to the average size shrimp data excluding the point for December 11.

Given the optimal value of c from the exponential fitting, we carried out the inverse problem calculations for the mean μ_b and standard deviation σ_b of the truncated normal distribution. We also interpolated the December 4 data and used it as an approximation to the initial size density $v_0(x; g)$ in (2.1). Since we excluded the data from December 11, only the three data sets from December 18, 24, and 31 are used in estimating the growth rate distribution parameters. We obtained an optimal cost of 87.4619 and an estimated variance $\hat{\sigma}^2$ of 3.1236 for these computations, which are much smaller than the values computed in the previous calculations in Section 6.3.1. The estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ are in Table 6.6 along with the corresponding confidence intervals.

The widths of the confidence intervals for both parameters are much smaller in comparison to those computed when the entire data set is considered. We are more confident

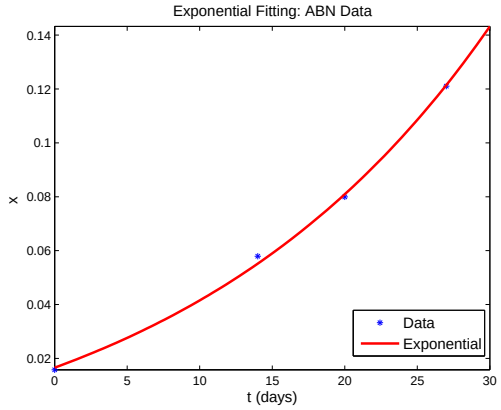

Figure 6.14: Exponential fit of (6.4) to ABN average size shrimp data excluding December 11 with $\frac{d\bar{x}}{dt} = 0.046(\bar{x} + 0.026)$.

Table 6.6: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and confidence intervals for inverse problem with ABN data excluding December 11 using $\hat{c} = 0.026$ from exponential fit of (6.4) to average size shrimp data.

$\hat{\mu}_b \pm 1.96SE(\hat{\mu}_b)$	$\hat{\sigma}_b \pm 1.96SE(\hat{\sigma}_b)$
0.0369 ± 0.0027	0.0159 ± 0.0030

about the estimates obtained in this case based on the relatively small confidence intervals. We were able to use the confidence intervals corresponding to the estimated mean and standard deviation of the truncated normal distribution to construct a confidence band for the estimated probability distribution on the growth rates using techniques discussed in Chapter 5 for the standard parametric approach. These results are shown in Figure 6.15 with the estimated probability density and confidence region on the left and the estimated probability distribution and confidence band on the right. The confidence band gives us a measure of the uncertainty associated with the infinite-dimensional parameter (probability distribution) that is of most interest in this application. Due to the width of this confidence band, we are fairly confident about the reliability of the estimated probability distribution.

Moreover, we note from the model predictions versus the data in Figure 6.16 that the model does a much better job of fitting the data on December 18, 24, and 31 when the data from December 11 is excluded. The model overall does a good job of predicting the key features of the size distribution for these three weeks.

We also used (6.8) corresponding to the probabilistic growth model (6.6) to fit the average size shrimp data with the data from December 11 excluded. The exponential fit

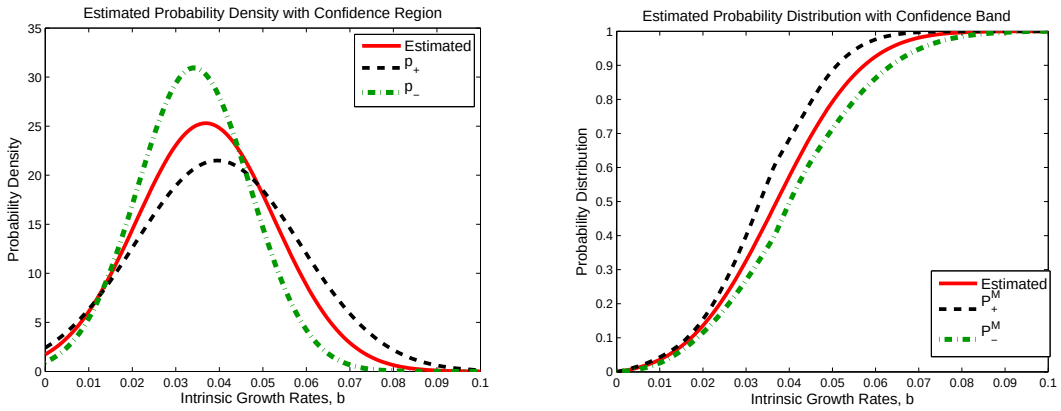

Figure 6.15: (left): Estimated probability density with confidence region; (right): Estimated probability distribution with confidence band for ABN data excluding December 11 using $\hat{c} = 0.026$ from exponential fit of (6.4) to average size shrimp data.

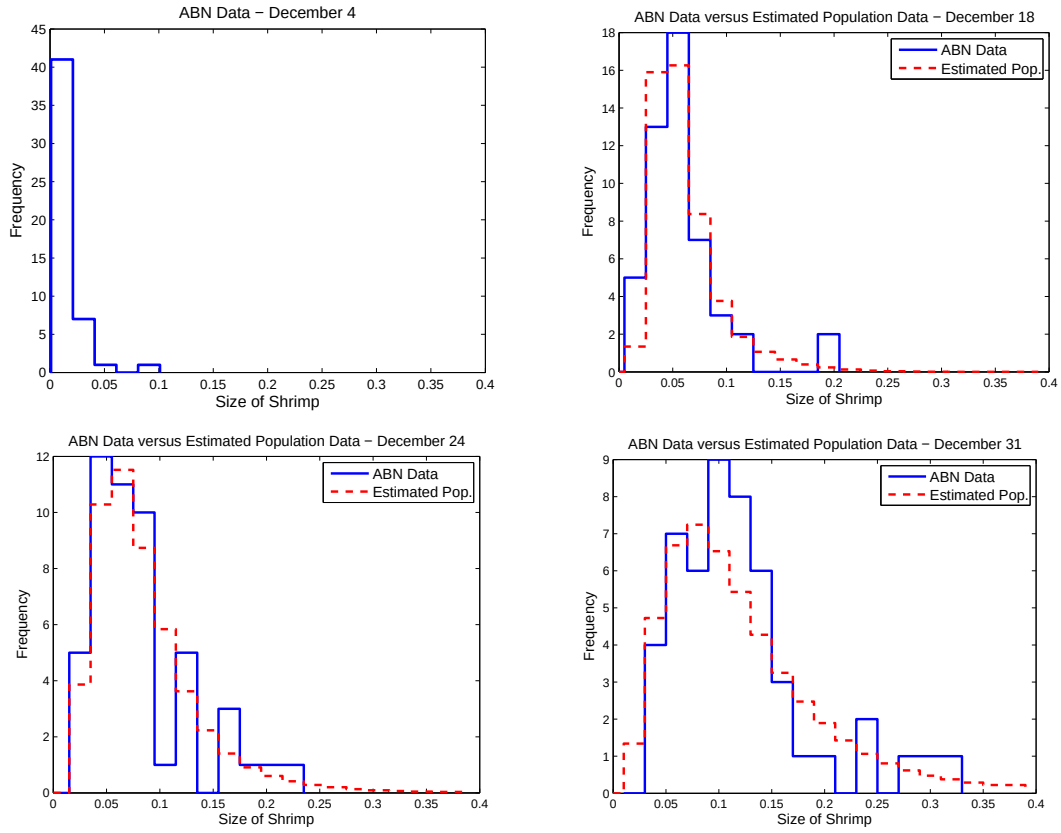

Figure 6.16: Longitudinal data on shrimp weight from ABN on December 4, 18, 24, and 31 versus model predicted population density $N_{GRD}(t, x; \hat{\theta})$ with $\hat{\theta} = (0.0369, 0.0159)$ and $\hat{c} = 0.026$ from exponential fit of (6.4) to average size shrimp data.

shown in Figure 6.17 corresponds to the optimal parameters

$$\hat{\mu}_b = 0.046, \quad \hat{\sigma}_b = 0.001, \quad \hat{c} = 0.027.$$

The estimated parameter values for μ_b and c in this case are very close to the estimated parameter values of b and c when (6.4) was used to fit the data. We see that (6.8) also provides a good fit to the average size shrimp data in the absence of the December 11 data point.

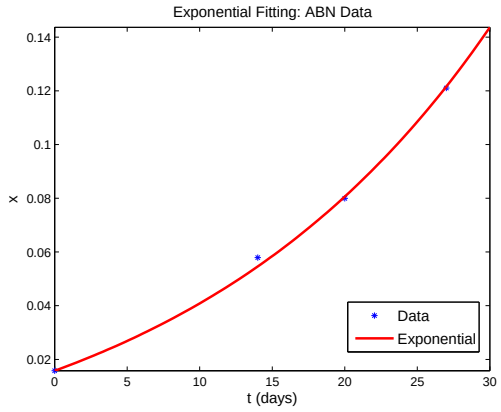

Figure 6.17: Exponential fit of (6.8) to ABN average size shrimp data excluding December 11 with $\frac{d\bar{x}}{dt} = (0.046 + 0.001^2 t)(\bar{x} + 0.027)$.

The estimates for (μ_b, σ_b) for the truncated normal distribution were obtained from the inverse problem calculations with the optimal value of c. The parameter estimates $(\hat{\mu}_b, \hat{\sigma}_b)$ are relatively close to the previously estimated parameters in Table 6.6, and in fact, the confidence intervals in Table 6.7 have the same width (centered differently) as those in Table 6.6.

Table 6.7: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and confidence intervals for inverse problem with ABN data excluding December 11 using $\hat{c} = 0.027$ from exponential fit of (6.8) to average size shrimp data.

$\hat{\mu}_b \pm 1.96SE(\hat{\mu}_b)$	$\hat{\sigma}_b \pm 1.96SE(\hat{\sigma}_b)$
0.0360 ± 0.0027	0.0158 ± 0.0030

We computed an optimal cost for these results of 88.0090 and an estimated variance $\hat{\sigma}^2 = 3.1432$, which are also very close to the values in the previous case in this subsection. These results also support our previous assertions. The estimated parameters obtained by excluding the data from December 11 are much more reliable than those obtained when the

entire data set from December is used in the parameter estimation problem. The small width observed in the confidence band associated with the estimated probability distribution in Figure 6.18 validates our claims as well.

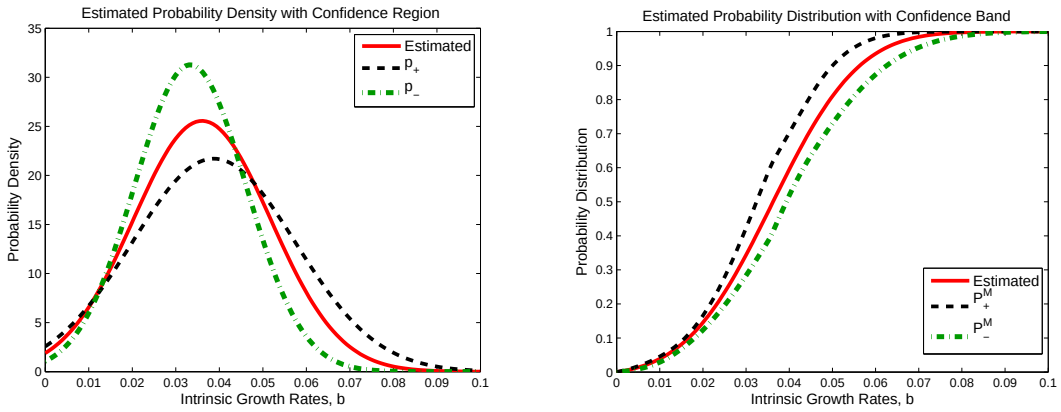

Figure 6.18: (left): Estimated probability density with confidence region; (right): Estimated probability distribution with confidence band for ABN data excluding December 11 using $\hat{c} = 0.027$ from exponential fit of (6.8) to average size shrimp data.

In Figure 6.19 we see similar results to those shown in Figure 6.16 with there being a very small difference in the model predictions on December 18. As previously remarked, the model does a very good job of fitting the data for December 18, 24, and 31 when the data from December 11 is excluded from the inverse problem calculations. The model predictions exhibit the lognormal behavior seen in the observations and provide reasonably accurate quantitative fits to the data.

While we were able to obtain reliable parameter estimates and good model predictions by excluding the data from December 11, we also chose to investigate the effects of the initial data in this problem by excluding the data from December 4 and using the data from December 11 as an approximation to the initial size density $v_0(x; g)$. The results in this section suggest that perhaps there was a change in the environmental conditions, such as feeding or temperature, which affected the growth of the shrimp. However, the poor model fits in Section 6.3.1 where the entire data set from December is considered could also be the result of using the wrong initial conditions. Therefore, we reconsidered the parameter estimation problem for the growth rate distribution with the data collected on December 11, 18, 24, and 31, and the results follow in the next section.

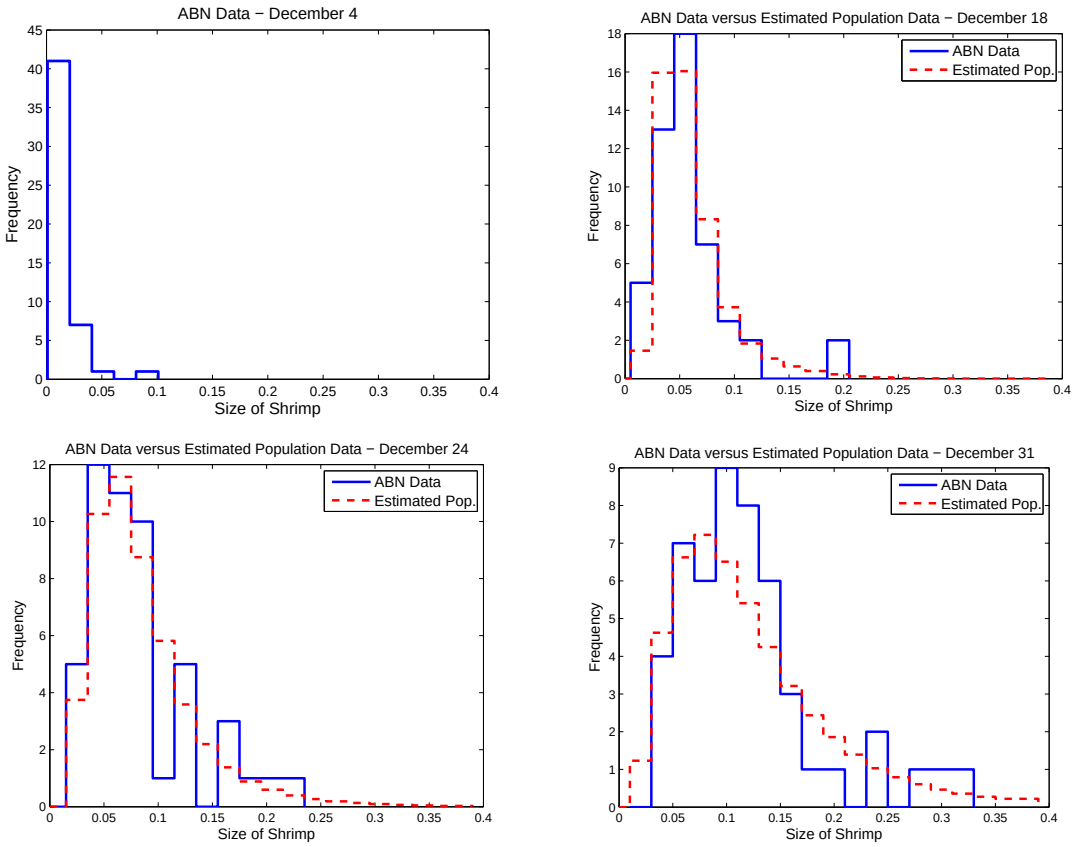

Figure 6.19: Longitudinal data on shrimp weight from ABN on December 4, 18, 24, and 31 versus model predicted population density $N_{GRD}(t, x; \hat{\theta})$ with $\hat{\theta} = (0.0360, 0.0158)$ and $\hat{c} = 0.027$ from exponential fit of (6.8) to average size shrimp data.

6.3.3 Results Excluding December 4 Data

We now present results for the inverse problem for the growth rate distribution of the size-structured shrimp population considering the data from December 11 - 31 in order to determine if using the interpolated December 11 data as an approximation for $v_0(x;g)$ will result in better model predictions. Under the assumption of the deterministic growth model of (6.5), Equation (6.4) was used to fit the average size shrimp data excluding the data point for December 4. The exponential function in (6.4) fits the average size shrimp data well as shown in Figure 6.20 with the optimal parameter values of

$$\hat{a} = 0.058, \quad \hat{b} = 0.047, \quad \hat{c} = 0.028.$$

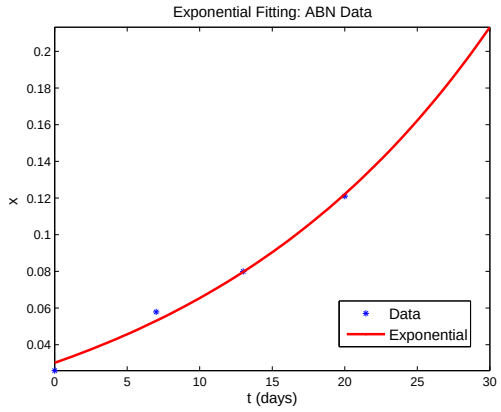

Figure 6.20: Exponential fit of (6.4) to ABN average size shrimp data excluding December 4 with $\frac{d\bar{x}}{dt} = 0.047(\bar{x} + 0.028)$.

The inverse problem for (μ_b, σ_b) was performed with the optimal value of c computed from the exponential fitting to the average size shrimp data. As already noted, we interpolated the data from December 11 and used it as an approximation to the initial size density. We were left with the final three data sets from December 18, 24, and 31 for the parameter estimation problem. The optimal parameter values for the mean and standard deviation of the truncated normal distribution along with the computed confidence intervals are in Table 6.8. We note that the optimal cost $J(\hat{\theta}) = 121.4561$ and the estimated variance of the statistical model $\hat{\sigma}^2 = 4.3377$. Although these values are larger than the values obtained when the data from December 4 was used as an approximation to the initial condition, the results are still significantly better than those obtained with the entire data set.

Table 6.8: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and confidence intervals for inverse problem with ABN data excluding December 4 using $\hat{c} = 0.028$ from exponential fit of (6.4) to average size shrimp data.

$\hat{\mu}_b \pm 1.96SE(\hat{\mu}_b)$	$\hat{\sigma}_b \pm 1.96SE(\hat{\sigma}_b)$
0.0441 ± 0.0077	0.0294 ± 0.0123

We also observe slightly larger confidence intervals corresponding to the estimated parameters $\hat{\mu}_b$ and $\hat{\sigma}_b$. These larger confidence intervals result in the larger confidence band for the estimated probability distribution seen in Figure 6.21. Thus, there is more uncertainty in the parameter estimates obtained in this case in comparison to the estimates obtained when the data from December 11 was excluded in Section 6.3.2. We are more confident about the reliability of the estimates obtained for (μ_b, σ_b) in Section 6.3.2 based on the smaller width of the confidence intervals and bands.

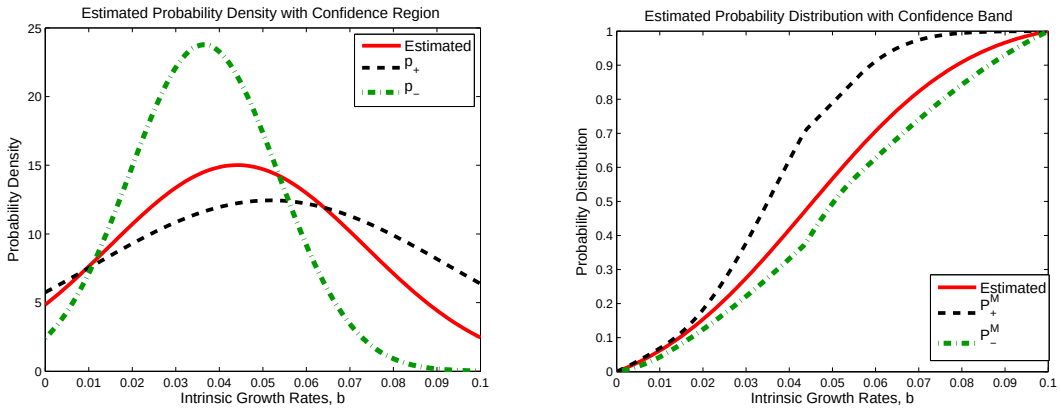

Figure 6.21: (left): Estimated probability density with confidence region; (right): Estimated probability distribution with confidence band for ABN data excluding December 4 using $\hat{c} = 0.028$ from exponential fit of (6.4) to average size shrimp data.

Although the optimal cost was slightly larger in this case, the model predictions shown in Figure 6.22 are similar to those in Section 6.3.2. The model still does a good job of predicting the key features of the shrimp population data when the data from December 11 is used versus the data from December 4. It appears that using the December 11 data instead of the December 4 data does not result in any improvements in the model fits to data.

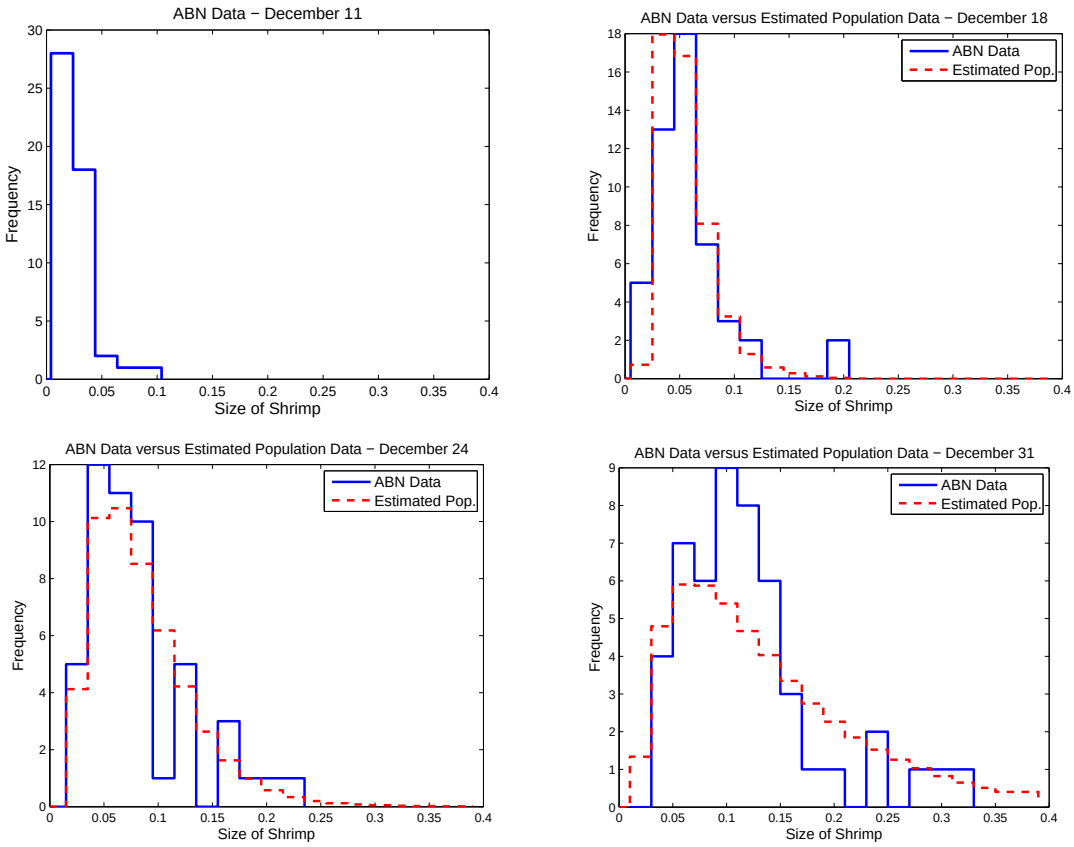

Figure 6.22: Longitudinal data on shrimp weight from ABN on December 11, 18, 24, and 31 versus model predicted population density $N_{GRD}(t, x; \hat{\theta})$ with $\hat{\theta} = (0.0441, 0.0294)$ and $\hat{c} = 0.028$ from exponential fit of (6.4) to average size shrimp data.

In our final set of computations, we fit the average size shrimp data with the data point from December 4 excluded with (6.8) taking into account the probabilistic nature of b. There is some noticeable difference in the estimated parameter values

$$\hat{\mu}_b = 0.038, \quad \hat{\sigma}_b = 0.003, \quad \hat{c} = 0.059$$

determined to give the best fit shown in Figure 6.23. However, we note that the exponential fit in Figure 6.23 is comparable to that shown in Figure 6.20.

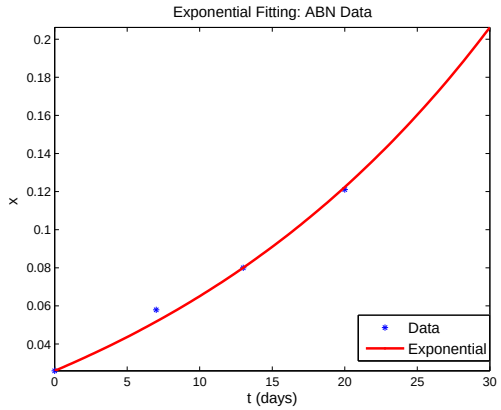

Figure 6.23: Exponential fit of (6.8) to ABN average size shrimp data excluding December 4 with $\frac{d\bar{x}}{dt} = (0.038 + 0.003^2 t)(\bar{x} + 0.059)$.

Table 6.9: Estimated parameters $(\hat{\mu}_b, \hat{\sigma}_b)$ and confidence intervals for inverse problem with ABN data excluding December 4 using $\hat{c} = 0.059$ from exponential fit of (6.8) to average size shrimp data.

$\hat{\mu}_b \pm 1.96SE(\hat{\mu}_b)$	$\hat{\sigma}_b \pm 1.96SE(\hat{\sigma}_b)$
0.0289 ± 0.0051	0.0215 ± 0.0065

The results of the inverse problem calculations for the mean μ_b and standard deviation σ_b of the truncated normal distribution are in Table 6.9 along with the confidence intervals. We computed an optimal cost of 106.6413 and an estimated variance $\hat{\sigma}^2$ of 3.8086 which are slightly better than the previous results in this section. We point out that the confidence intervals for the estimated mean in Table 6.9 do not intersect with those in Table 6.8, but the confidence intervals for $\hat{\sigma}_b$ do intersect. However, we still infer from the results shown here that the parameter estimates obtained in absence of the data from December 4 are not as reliable as those obtained in the previous section where the data from December 11 was omitted. The confidence band associated with the estimated probability

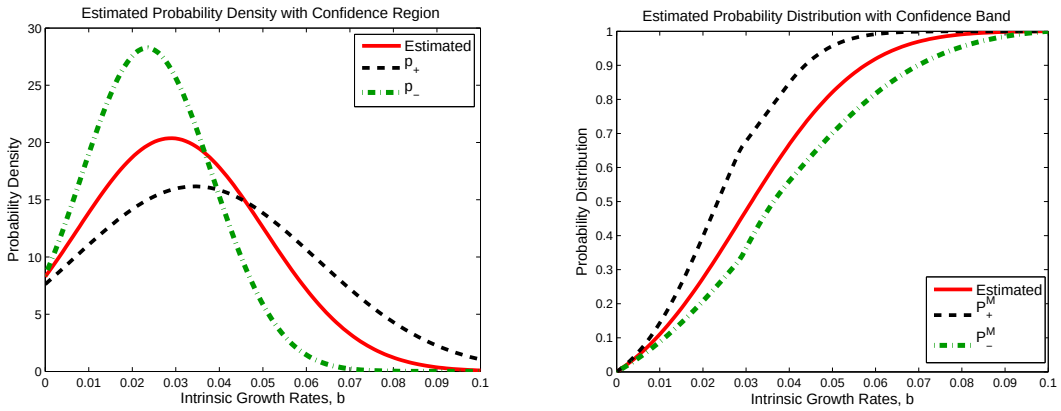

Figure 6.24: (left): Estimated probability density with confidence region; (right): Estimated probability distribution with confidence band for ABN data excluding December 4 using $\hat{c} = 0.059$ from exponential fit of (6.8) to average size shrimp data.

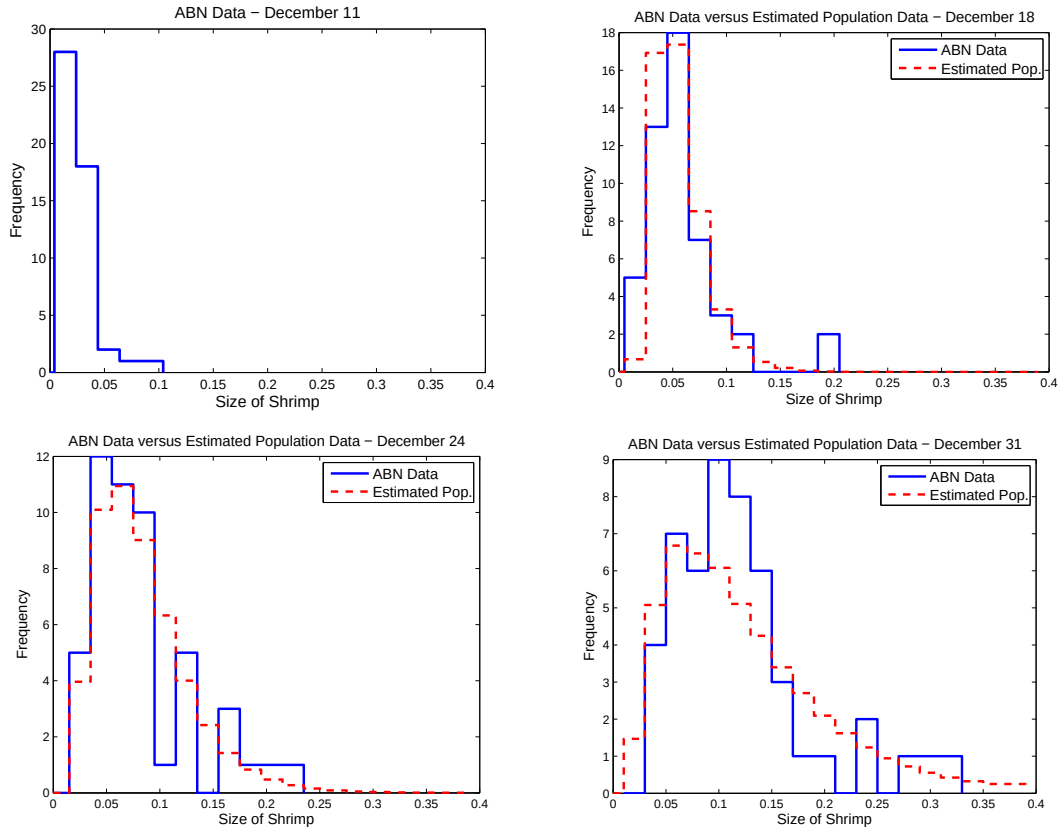

Figure 6.25: Longitudinal data on shrimp weight from ABN on December 11, 18, 24, and 31 versus model predicted population density $N_{GRD}(t, x; \hat{\theta})$ with $\hat{\theta} = (0.0289, 0.0215)$ and $\hat{c} = 0.059$ from exponential fit of (6.8) to average size shrimp data.

distribution in Figure 6.24 also suggests that we are not as confident about the reliability of this estimated probability distribution in comparison to the estimates presented in Section 6.3.2.

The model predictions in Figure 6.25 are very similar to the previous results. We observe nice model fits to the data which validates the model's ability to accurately describe the growth dynamics of the shrimp. The results presented in this section suggest that there was perhaps some change in the environmental conditions between December 4 and December 11 which resulted in very poor model predictions when the entire data set was used in the parameter estimation problem. Although we were able to use the data from December 11 as an approximation to the initial size density to compute good fits to the data, the best numerical and statistical results were obtained when the data from December 11 was excluded from the inverse problem calculations. These results were illustrative of the types of issues that can be addressed with mathematical and statistical tools when using models to describe real world applications.

Chapter 7

Conclusions and Future Directions

We have presented computational and statistical results for both parametric and non-parametric versions of the inverse problem for the estimation of probability distributions arising in problems with aggregate or population level observations. One non-parametric approximation scheme, DEL(M), is at the level of the distribution being sought, while the other non-parametric approximation scheme, SPL(M, N), is at the level of the density (tacitly assumed to exist in the underlying theory [24]) associated with the distribution. Several computational examples with simulated data in the context of functional growth rate estimation in size-structured population models are presented in order to validate the underlying theory that we discussed guaranteeing convergence of distributions. The most recently developed scheme, SPL(M, N), and the standard parametric OLS formulation are used with experimental data for mosquitofish populations and compared to earlier findings with the DEL(M) scheme. While the mathematical and statistical methodologies presented here were used in the context of size-structured population models, the ideas are widely applicable to other areas of applications in biology as well as in viscoelastic materials and in electromagnetic interrogation of dielectric materials.

The results discussed here demonstrate some of the strengths and weaknesses associated with each of the methods considered. When the form of the probability distribution is known a priori, the parametric approach PAR(M, N) is the better method to use because the number of parameters to be estimated is typically small so computations are usually not very expensive. We also observed much tighter confidence bands around the estimated probability distributions in the computational examples considered in this dissertation; hence, we are fairly confident about the reliability of estimates obtained with this

method. However, the accuracy of the parameter estimates relies heavily on one's ability to correctly specify the form of the probability distribution a priori.

In contrast, the non-parametric approaches are a better choice when one cannot (correctly) identify the form of the probability distribution. The delta function approximation method DEL(M) is very easy to implement and computationally inexpensive; however, a large number of elements is usually necessary for convergence of the estimated probability distributions. The underlying theory [4] guarantees convergence of distributions, not densities, in the Prohorov metric. Therefore, estimates of the probability densities are very misleading in terms of accuracy of corresponding estimated probability distributions. While the spline based method SPL(M, N) provides a much smoother approximation of continuous probability distributions in comparison to DEL(M), it is more computationally expensive. With appropriate choices for the weights $\{a_k^M\}$, convergence of distributions in the Prohorov metric (with significantly fewer elements than DEL(M)) is guaranteed as well as convergence of the approximating densities in L^2 [24]. Lastly, over-parametrization of the inverse problem with the non-parametric approaches can not only result in oscillations in the estimated densities but larger confidence bands as well. As a result, we feel less certain about the reliability of the estimated probability distributions obtained with these non-parametric methods.

We demonstrated how to construct "functional" confidence bands for estimated probability distributions in size-structured mosquitofish populations in both parametric and non-parametric settings with several computational examples. However, one would like to fully develop the mathematical and asymptotic statistical theory for OLS problems with functional parameters, such as the probability distributions studied here and the time- and spatial-dependent functional parameters discussed in the Introduction. We derived an OLS estimator for functional parameters in the functional linear regression model and determined its mean and variance. However, the variance was shown to depend on the "true" variance σ_0^2, which is generally unknown. Therefore, future work involves constructing an estimator for σ_0^2 which will require the trace operator in functional space. Moreover, we also need to determine an appropriate form of the Central Limit Theorem for functional space in order to outline in detail the asymptotic sampling distribution for OLS estimators for infinite-dimensional parameters. We are currently working on the development of this fundamental theory in an alternate weak L^2 setting for densities.

In the final computational example presented here, we also used the Growth Rate

Distribution (GRD) model for size-structured shrimp populations whose size distribution data exhibited a great deal of variability. We outlined a computational methodology that proved to be useful in designing experiments that were carried out to validate the GRD model. We were able to determine the sampling *size* and *frequency* sufficient for reliable estimates of the variability in the growth rates of size-structured shrimp populations. Using our inverse problem methodology, we were able to successfully estimate the growth rate distribution from experimental data collected during the early growth of shrimp populations at ABN based on our design calculations. In the future we propose considering different parametric and non-parametric approaches in the GRD model in order to improve the model predictions for the early growth dynamics of shrimp.

Bibliography

[1] L.J.S. Allen. *An Introduction to Stochastic Processes with Applications to Biology.* Pearson Education, Inc., Upper Saddle River, New Jersey, 2003.

[2] H.T. Banks, J.E. Banks, L.K. Dick, and J.D. Stark. Estimation of dynamic rate parameters in insect populations undergoing sublethal exposure to pesticides. *Bulletin of Mathematical Biology*, 69:2139–2180, 2007.

[3] H.T. Banks, J.E. Banks, L.K. Dick, and J.D. Stark. Time-varying vital rates in ecotoxicology: selective pesticides and aphid population dynamics. *Ecological Modelling*, 210:155–160, 2008.

[4] H.T. Banks and K.L. Bihari. Modelling and estimating uncertainty in parameter estimation. *Inverse Problems*, 17:95–111, 2001.

[5] H.T. Banks, V.A. Bokil, S. Hu, F.C.T. Allnutt, R. Bullis, A.K. Dhar, and C.L. Browdy. Shrimp biomass and viral infection for production of biological countermeasures. *Mathematical Biosciences and Engineering*, 3:635–660, 2006.

[6] H.T. Banks, D.M. Bortz, G.A. Pinter, and L.K. Potter. *Modeling and imaging techniques with potential for application in bioterrorism*, chapter 6, pages 129–154. SIAM, Philadelphia, PA, 2003.

[7] H.T. Banks, L.W. Botsford, F. Kappel, and C. Wang. Modeling and estimation in size structured population models. In *Proceedings of the 2nd Autumn Course on Mathematical Ecology*, pages 521–541, Singapore, 1988. World Press.

[8] H.T. Banks and J.L. Davis. A comparison of approximation methods for the estimation of probability distributions on parameters. *Applied Numerical Mathematics*, 57:753–777, 2007.

[9] H.T. Banks and J.L. Davis. Quantifying uncertainty in the estimation of probability distributions with confidence bands. CRSC Technical Report CRSC-TR07-21, North Carolina State University, December 2007. (See also *Mathematical Biosciences and Engineering* (to appear)).

[10] H.T. Banks, J.L. Davis, S.L. Ernstberger, S. Hu, E. Artimovich, A.K. Dhar, and C.L. Browdy. Comparison of probabilistic and stochastic formulations in modeling growth uncertainty and variability. CRSC Technical Report CRSC-TR08-03, North Carolina State University, February 2008. *Journal of Biological Dynamics* (to appear).

[11] H.T. Banks, J.L. Davis, and S. Hu. Comparison of Fokker-Planck model and Growth Rate Distribution model in modeling growth uncertainty. (in progress).

[12] H.T. Banks, S. Dediu, and H.K. Nguyen. Sensitivity of dynamical systems to parameters in a convex subset of a topological vector space. *Mathematical Biosciences and Engineering*, 4:403–430, 2007.

[13] H.T. Banks and B.G. Fitzpatrick. Estimation of growth rate distributions in size-structured population models. *Quarterly of Applied Mathematics*, 49:215–235, 1991.

[14] H.T. Banks, B.G. Fitzpatrick, L.K. Potter, and Y. Zhang. Estimation of probability distributions for individual parameters using aggregate population data. In W. McEneaney, G. Yin, and Q. Zhang, editors, *In Stochastic Analysis, Control, Optimization and Applications*, pages 353–371. Birkhaüser, Boston, 1989.

[15] H.T. Banks and N.L. Gibson. Well-posedness in maxwell systems with distributions of polarization relaxation parameters. *Applied Mathematics Letters*, 18:423–430, 2005.

[16] H.T. Banks and N.L. Gibson. Electromagnetic inverse problems involving distributions of dielectric mechanisms and parameters. *Quarterly of Applied Mathematics*, 64:749–795, 2006.

[17] H.T. Banks and D.W. Iles. A comparison of stability and convergence properties of techniques for inverse problems. Technical Report LCDS-86-3, Brown University, January 1986.

[18] H.T. Banks and D.W. Iles. On compactness of admissible parameter sets: Convergence and stability in inverse problems for distributed parameter systems. In I. Lasiecka and

R. Triggiani, editors, *In Control Problems for Systems Described by Partial Differential Equations and Applications*, volume 97, pages 130–142. Springer - Verlag, Berlin, 1987.

[19] H.T. Banks and P. Kareiva. Parameter estimation techniques for transport equations with application to population dispersal and tissue bulk flow models. *Journal of Mathematical Biology*, 17:253–273, 1983.

[20] H.T. Banks, P. Kareiva, and L. Zia. Analyzing field studies of insect dispersal using two-dimensional transport equations. *Environmental Entomology*, 17:815–820, 1988.

[21] H.T. Banks, P.M. Kareiva, and P.K. Lamm. Modeling insect dispersal and estimating parameters when mark-release techniques may cause initial disturbances. *Journal of Mathematical Biology*, 22:259–277, 1985.

[22] H.T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems*. Birkhaüser, Boston, 1989.

[23] H.T. Banks, Y. Ma, and L.K. Potter. A simulation-based comparison between parametric and nonparametric estimation methods in PBPK models. *Journal of Inverse and Ill-posed Problems*, 13:1–26, 2005.

[24] H.T. Banks and G.A. Pinter. A probabilistic multiscale approach to hysteresis in shear wave propagation in biotissue. *SIAM Journal of Multiscale Modeling and Simulation*, 2:395–412, 2005.

[25] H.T. Banks and L.K. Potter. Probabilistic methods for addressing uncertainty and variability in biological models: Application to a toxicokinetic model. *Mathematical Biosciences*, 192:193–225, 2004.

[26] H.T. Banks and I.G. Rosen. Numerical schemes for the estimation of functional parameters in distributed models for mixing mechanisms in lake and sea sediment cores. *Inverse Problems*, 3:1–23, 1987.

[27] H.T. Banks, H.T. Tran, and D.E. Woodward. Estimation of variable coefficients in Fokker Planck equations using moving finite elements. *SIAM Journal of Numerical Analysis*, 30:1574–1602, 1993.

[28] S.L. Beal and L.B. Sheiner. Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering*, 8:195–222, 1982.

[29] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

[30] L.W. Botsford, B. Vondracek, T.C. Wainwright, A.L. Linden, R.G. Kope, D.E. Reed, and J.J. Cech Jr. Population development of the mosquitofish, *gambusia affinis*, in rice fields. *Environmental Biology of Fishes*, 20(2):143–154, 1987.

[31] R.L. Burden and J.D. Faires. *Numerical Analysis, third ed.* PWS Publishers, Boston, 1985.

[32] G. Casella and R.L. Berger. *Statistical Inference.* Duxbury, California, 2002.

[33] F.L. Castille, T.M. Samocha, A.L. Lawrence, H. He, P. Frelier, and F. Jaenike. Variability in growth and survival of early postlarval shrimp (*Penaeus vannamei* Boone 1931). *Aquaculture*, 113:65–81, 1993.

[34] M. Davidian and A.R. Gallant. Smooth nonparametric maximum likelihood estimation for population pharamcokinetics, with application to quinidine. *Journal of Pharmacokinetics and Pharmacodynamics*, 20:529–556, 1992.

[35] M. Davidian and A.R. Gallant. The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, 80:475–488, 1993.

[36] M. Davidian and D.M. Gilitan. *Nonlinear Models for Repeated Measurement Data.* Chapman & Hall, London, 1995.

[37] M. Davidian and D.M. Gilitan. Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, 8:387–419, 2003.

[38] L.C. Evans. *Partial Differential Equations.* American Mathematical Society, Providence, Rhode Island, 1998.

[39] M.V. Evans, W.K. Boyes, P.J. Bushnell, J.H. Raymer, and J.E. Simmons. A physiologically based pharmacokinetic model for trichloroethylene (TCE) in Long-Evans rats. preprint, 1999.

[40] A.R. Gallant. *Nonlinear Statistical Models.* J. Wiley & Sons, New York, 1987.

[41] T.C. Gard. *Introduction to Stochastic Differential Equations.* Marcel Dekker, New York, 1988.

[42] N.L. Gibson. *Terahertz-based electromagnetic interrogation techniques for damage detection.* PhD thesis, North Carolina State University, 2004.

[43] A. Gross, S. Abutbul, and D. Zilberg. Acute and chronic effects of nitrite on white shrimp, Litopenaeus vannamei, cultured in low-salinity brackish water. *Journal of the World Aquaculture Society*, 35:315–321, 2004.

[44] C. Hawkes. The estimation of the dispersal rate of the adult cabbage root fly (*Erioischia brassicae*) in the presence of (*Brassicae*) crop. *Journal of Applied Ecology*, 9:617–632, 1972.

[45] R.I. Jennrich. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40:633–643, 1969.

[46] M. Kot. *Elements of Mathematical Ecology.* Cambridge University Press, Cambridge, 2001.

[47] S. Laramore, C.R. Laramore, and J. Scarpa. Effect of low salinity on growth and survival of postlarvae and juvenile *litopenaeus vannamei*. *Journal of the World Aquaculture Society*, 32:385–392, 2001.

[48] B.G. Lindsay. The geometry of mixture likelihoods: A general theory. *The Annals of Statistics*, 11:86–94, 1983.

[49] B.G. Lindsay. *Mixture Models: Theory, Geometry and Applications*, volume 5 of *NSF-CBMS Regional Conf. Series in Probability and Statistics.* Institute Mathematical Statistics, Haywood, CA, 1995.

[50] B.G. Lindsay and M.L. Lesperance. A review of semiparametric mixture models. *Journal of Statistical Planning and Inference*, 47:29–39, 1995.

[51] D.G. Luenberger. *Optimization by Vector Space Methods.* John Wiley & Sons, New York, 1969.

[52] A. Mallet. A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, 73:645–656, 1986.

[53] A. Okubo. *Diffusion and Ecological Problems: Mathematical Models*, volume 10 of *Biomathematics.* Springer - Verlag, Berlin, 1980.

[54] J. Ponce-Palafox, C.A. Martinez-Palacios, and L.G. Ross. The effects of salinity and temperature on the growth and survival rates of juvenile white shrimp, *penaeus vannamei boone 1931*. *Aquaculture*, 157:107–115, 1997.

[55] L.K. Potter. *Physiologically based pharmacokinetic models for the systemic transport of trichloroethylene*. PhD thesis, North Carolina State University, 2001.

[56] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Springer, New York, 2000.

[57] A. Schumitzky. Nonparametric EM algorithms for estimating prior distributions. *Applied Mathematics and Computations*, 45:141–157, 1991.

[58] A. Schumitzky. The nonparametric maximum likelihood approach to pharmacokinetic population analysis. In *Proceedings of the Western Simulation Multiconference-Simulation in Health Care*, San Diego, CA, 1993. Society for Computer Simulation.

[59] G.A.F. Seber. *Linear Regression Analysis*. John Wiley & Sons, New York, 1977.

[60] G.A.F. Seber and C.J. Wild. *Nonlinear Regression*. John Wiley & Sons, New York, 1989.

[61] L.B. Sheiner, B. Rosenberg, and K. Melmon. Modeling of individual pharamcokinetics for computer-aided drug dosage. *Computers and Biomedical Research*, 5:441–459, 1972.

[62] J.W. Sinko and W. Streifer. A new model for age-size structure for a population. *Ecology*, 48:910–918, 1967.

[63] G. Wahba. Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society Series B*, 45:133–150, 1983.

[64] G. Wahba. (Smoothing) splines in nonparametric regression. *Encyclopedia of Environmetrics*, 4:2099–2112, 2001.