THREE ESSAYS ON ESTIMATION OF APPLIED ECONOMIC MODELS IN FISHERIES

A Dissertation Presented

by

CHRISTOPHER BURNS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2014

Resource Economics

THREE ESSAYS ON ESTIMATION OF APPLIED ECONOMIC MODELS IN FISHERIES

A Dissertation Presented

by

CHRISTOPHER BURNS

Approved as to style and content by:

____
Daniel Lass, Chair

____
Sylvia Brandt, Member

____
John Staudenmayer, Member

____
Daniel Lass, Department Chair
Resource Economics

This manuscript is dedicated to my patient and loving fiancée,
Katie Fox.

ACKNOWLEDGMENTS

I would like to thank Dan Lass, Sylvia Brandt, John Staudenmayer, and John Buonaccorsi for their expert advice and encouragement. I would also like to thank the UMASS-Amherst Resource Economics department faculty and graduate students for their support throughout the years.

ABSTRACT

THREE ESSAYS ON ESTIMATION OF APPLIED ECONOMIC MODELS IN FISHERIES

MAY 2014

CHRISTOPHER BURNS

B.S., THE COLLEGE OF NEW JERSEY

M.S, THE UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Daniel Lass

Fishery managers face many challenges when setting effective policies. This includes working with fisherman to set the total allowable catch (TAC), preventing overfishing, and monitoring the status of a fishing industry based on imperfect data. This dissertation focuses on the last two issues. In particular we focus on how measurement error and estimation issues can impact fishery policy using two common economic models. The two models we examine are the stochastic frontier model and the Schaefer production model. Both of these models use production data on inputs and outputs to estimate a production function. Chapter 2 looks at the impact of measurement error in one of the inputs of the stochastic frontier model, which is widely used in estimating technical efficiency for an industry. Chapter 3 looks at the impact of measurement of the Schaefer production model, which estimates the biomass using production data only. Chapter 4 revisits

the stochastic frontier model and examines how Bayesian methods can be used to measure production efficiency.

Using panel data from the Mid-Atlantic surfclam fishery from 2001-2009, Chapters 2 and 3 examine measurement error correction methods that empirical researchers and policy makers can use when instrumental variables are not available. We chose this fishery because both logbook data from vessels, and scientific estimates of the biomass are available. We use these data for all three chapters of this study. The Mid-Atlantic surfclam fishery is significant to fishery policy because it was the first U.S. fishery to be regulated by Individual Tradable Quotas (ITQs). These reasons make it an ideal fishery for estimating the two production models and evaluating policy implications.

In Chapters 2 and 3 we make use of estimates of the measurement error variance, and apply a simulation based correction method known as simulation extrapolation (SIMEX). SIMEX is a simulation based method for reducing bias in parameter estimates caused by measurement error. After estimating both models under a naive analysis, SIMEX is used to obtain less biased estimates of technical efficiency, production and biological parameters. In the last chapter we revisit the stochastic frontier model, estimating both a maximum likelihood and Bayesian model. We also seek to understand how the industrial organization of the Mid-Atlantic surfclam fishery is changing, and what this means for the future of the industry and the health of the fishery.

We conclude the dissertation by discussing the results from each chapter, and their implications for fishery policy. The results from Chapter 2 show that not taking measurement error into account would lead fishery managers to miss important relationships between management decisions and vessel technical efficiency. In Chapter 3 we show that the bias from the Schaefer production model can lead to poor estimates of the resource biomass. Given poor information about the biomass regulators may set the total catch too high, the consequences of which would be resource depletion and economic losses to firms. In Chapter 4 we also show how empirical researchers can take advantage of Bayesian methods for measuring production efficiency. Our analysis concludes that both maximum likelihood and Bayesian models of production efficiency reveal changes in the fleet structure are having significant impact on technical efficiency. We also see that the marginal productivities of time fishing and vessel length are decreasing over time, due to spatial changes in the biomass and declining landings per-unit-effort.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

For the first time in human history there is evidence that the world's total fish harvest is declining because of overfishing (Hilborn et al. 2003). Much of this decline is occurring in regulated fisheries, where current management practices have failed to address the economic behavior of fishermen and stop overfishing. Traditional approaches to managing fisheries focused on the fish, and less on the economic incentives fisherman face. Economists have long argued that the best way to prevent overfishing is to create management practices which take into account a fisherman's incentives (Moloney and Pearse 1979, Wilen 1985). More recently, the introduction of management measures designed to internalize the economic costs of overfishing, such as Individual Tradable Quotas (ITQs) (Walden et al. 2012) and landing taxes (Weitzman 2002), have been more successful at aligning fisherman's incentives with goals of maximizing economic efficiency and preserving biological sustainability. To determine whether the goals of preventing overfishing and maximizing economic efficiency are being achieved, fishery managers estimate models using observational data, usually obtained from vessel logbooks. These data can be used to measure economic and biological parameters of fisheries, but are often measured with error (Grafton 2006, Griliches and Hausman 1986).

The theoretical consequences of measurement error are well known in the Statistics and Econometrics literature (Carroll et al. 2012, Hausman 2001), though much of the applied economic literature does not address these issues. The first two chapters of this dissertation will focus on the impact of measurement error in the

Stochastic Frontier Model and the Schaefer Production Model. The last chapter will examine Bayesian methods for estimating production efficiency and compare them to frequentist methods, such as maximum likelihood. Before exploring the theory behind these two models, we first motivate the challenges that fishery managers face, namely aligning the short-run incentives of fishermen with long-run preservation of the resource.

The biggest challenge surrounding effective fishery management is that fisheries are a common property resource. A common property resource is one where harvests are rivalrous, and it is difficult to exclude others from the fishery. A rivalrous fishery can be defined to mean that a fish caught today cannot be caught be someone else tomorrow. Thus, there is an incentive for fishermen to catch as many fish as possible today, because there is no guarantee that the fish will be there tomorrow. The common property aspect of fisheries means that the sea is not owned by any individual or entity, and therefore entry into a fishery is not restricted, making effective management very difficult. This situation often results in too many fishermen chasing too few fish.

There are examples of small scale fisheries where cooperation between fisherman can prevent overfishing, such as the Maine lobster fishery (Acheson 1988), but often times the size of fisheries makes cooperation very difficult. In such as fishery, a fisherman whom decides not harvest until later potentially benefits all other fisherman, but not himself. The typical result in an unregulated common property fishery is that fishermen will continue to enter the fishery until all economic profits are gone. The likelihood that the fishery stock will become significantly depleted in this scenario is very high. The degree to which the fishery becomes depleted depends on a number of factors, including biological parameters such as the intrinsic growth rate, and the mortality rate. Other factors, including the environment and ecosystem conditions, will ultimately determine the fate of the fishery.

For the last half-century, dozens of countries have implemented a wide variety of regulatory measures to control overfishing by regulating effort. The underlying assumption is that if these regulations are enforced vigorously, then further fishing effort will be contained and the fishery will be economically sustainable. In order to regulate effort, management practices have traditionally focused on command and control measures, such as restricting the number of vessels (known as limited access fisheries), limiting the types of gear used for harvesting (input controls), limiting days at sea, and shortening the fishing season. These measures have largely failed to control overfishing because they do not take into account the economic incentives of fishermen. This failure to account for economic incentives is revealed in the continued decline of many fisheries due to ”effort creep”(Grafton 2006). Effort creep refers to the continued increase in fishing effort as fisherman substitute from regulated to unregulated inputs. Fishermen can always find ways to increase effort when regulations do not directly account for economic behavior. An example is when limits are placed on the type of harvesting gear that can be used. Fisherman will then substitute advanced sonar or increase the horsepower of their engines to compensate for this input restriction. The overarching theme is that regulating inputs or controlling fisheries without incorporating the economic incentives fishermen face is a losing proposition for fishery managers.

In an unrestricted fishery, sometimes called an ”open access resource”, each individual fisherman will make harvest decisions based on his own private marginal costs and benefits. While this behavior is individually optimal, it will likely lead to a socially undesirable result, such as the economic and biological collapse of the fishery. The economic reason for this is because each fisherman, while acting rationally, will not internalize the costs of each additional harvest on others. In order to prevent these disastrous results, fishery managers must design policies which ”internalize” the social costs of fishing for each fishermen, and thus change the economic incentives that they face. To accomplish this, fishery policy must account for the ”user cost” or ”resource rent” that accrues to a fisherman. User costs are the additional costs imposed on others from additional harvesting by fishermen. By accounting for the natural provision of the fishery, i.e. the social cost of using it, fishery managers can attempt to achieve a target catch in a way that prevents fisherman from overusing it. Landing taxes and Individual Tradable Quotas (ITQs) are two regulatory mechanisms that attempt to internalize the user cost by forcing fishermen to account for the value of the resource in their decision making.

Landing taxes capture the user cost by assessing a fee for each fish caught. They are price instruments designed to reduce the incentives to over-fish (Weitzman 2002). One advantage to landing taxes is that they are relatively easy to enforce. However, one downside to this management technique is that the regulatory needs information about firm marginal costs and the status of the resource in order to set an optimal tax rate. This is a high standard to meet for most fisheries, where managers commonly operate with incomplete information.

Individual Tradable Quotas are a quantity instrument which have a lower informational burden than landing taxes. The goal of ITQs is to create ”well defined property rights” for the fishery (Moloney and Pearse 1979), which usually take the form of a quota. These quota can then be bought or sold by fishermen who participate in a market. Given that the regulator knows the stock of the fishery, a total allowable catch (TAC) can be set for a given year, allowing fishery managers to set the quota. With this information a fishery manager can allocate ”shares” or quota to active fisherman. Allocation of the TAC to fisherman can be done in several ways. Typically quota are initially allocated based on some predetermined formula, such as history of catch, or number of years each fisherman has been active in a fishery. Fishermen can then buy or sell their quota in a market. After the first year of the ITQ program, fishermen are allocated a share of this TAC based on their current quota holdings. Economic theory states that ITQ management should prevent overharvesting while allowing fishermen to make economic decisions that benefit them, by allowing them to minimize their costs. The economic benefits of ITQs have their root in welfare economics and the idea that a perfectly competitive market will lead to a Pareto efficient outcome (Moloney and Pearse 1979).

An advantage of ITQs is that input controls are not needed. This allows firms to choose their inputs freely and operate more efficiently. Additional benefits include a reduction in effort (i.e. fewer boats), a reduced incentive to "race-to-fish" common in open access fisheries, and an improvement in product quality. Because the "race-to-fish" is no longer incentivized, longer fishing seasons and higher prices for fish are usually observed. Both of these benefits have been observed in the British Columbia Halibut fishery. Before ITQs were implemented the fishing season lasted about two days. This was a result of too many boats operating in the fishery and caused a large portion of the Halibut catch to be frozen before it was sold. After ITQs were implemented, ex-vessel prices for Halibut increased and the fishing season grew to more than 200 days, resulting in a market for fresh Halibut (Casey et al. 1995). Around the world, ITQ management has become widely used to protect fisheries from overfishing and maximize welfare to society. This is particularly the case in countries that rely heavily on fishing for their economies, such as New Zealand and Iceland (Grafton 2006). In 1991, the first U.S. fishery to be regulated by ITQs was the Mid-Atlantic surfclam fishery.

The Mid-Atlantic surfclam fishery spans the U.S. eastern Atlantic coast from the southern Gulf of St. Lawrence to Cape Hatteras, shown in Figure 1.1. Atlantic surfclams are slow-growing bivalve mollusk that live in the water at depths of 20-80m. The U.S. stock is almost entirely in the Economic Exclusive Zone (EEZ), located between 3-200miles offshore. The overall status of the biomass is currently in a period of decline after reaching record levels in the 1990s. About half of the current stock is located on the George Bank, which has not been fished since 1989 due to paralytic shellfish poisoning (PSP) toxins found in the surfclam meat. The highest concentration of fishable biomass is off the northern New Jersey coast, and as a result fishing effort has increased substantially over the last decade (Northeast Fisheries Science Center 2010). The southern end of the fishery, particularly in the Delmarva region, has seen a higher mortality rate in the past decade, leading to a decline in biomass (McCay et al. 2011, Weinberg 2005).

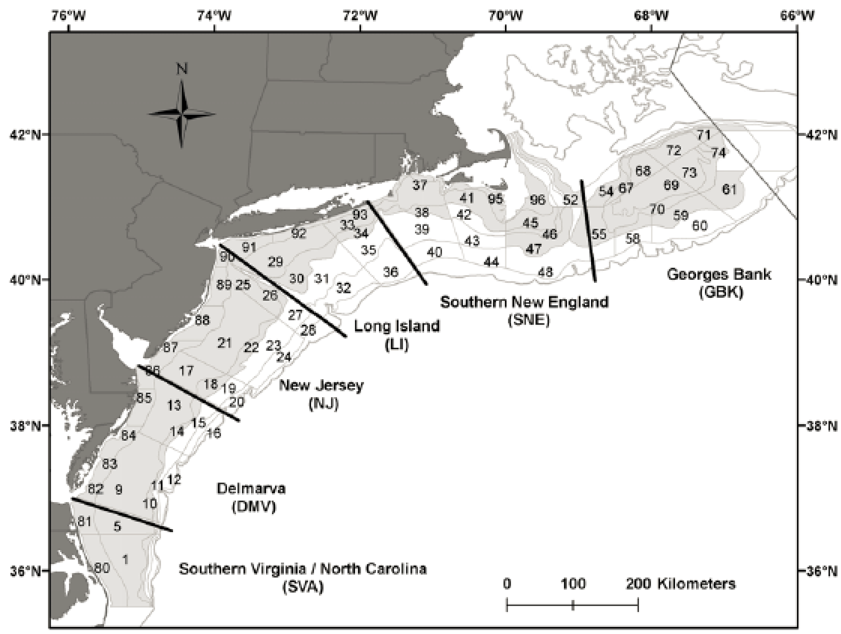

Figure 1.1: Mid-Atlantic Surfclam Fishery

Mid-Atlantic Fishery Management Council (2010)

From 1979-1989 surfclams in the U.S. EEZ were regulated under a command and control system that limited the amount of time a vessel could fish in a calendar quarter (Walden et al. 2012). In 1990 the fishery transitioned from limited entry to ITQs under the direction of the Mid-Atlantic Fishery Management Council. Current management measures include an annual quota for the Exclusive Economic Zone (EEZ) waters and mandatory logbooks that describe each fishing trip. The Mid-Atlantic surfclam industry has consolidated considerably since the introduction of ITQs, going from approximately 120 vessels in 1990 to fewer than 50 vessels in 2005. As detailed in Brandt (2007), many of the vessels that exited just after the regulatory change were very inefficient. Economic theory would predict that the remaining fleet, which is considerably more horizontally and vertically integrated than before, would have significantly higher average levels of production efficiency. Weninger (1998) predicted that the overall surfclam and ocean quahog fleet would eventually settle at between 21-25 vessels, after calculating the optimal level of harvesting capacity. While the active fleet is still much larger than this, continued consolidation is occurring. In 2005 a market crisis, caused by a decline in demand for U.S. surfclams, led to a substantial portion of the surfclam fleet leaving the fishery. The remaining fleet is fewer than 40 vessels, many of which are vertically integrated with processors.

Another factor that is driving consolidation of the industry is declining landings per-unit-effort (LPUE). During the last decade, LPUE for the fishery has been declining at a rate of 10% per year (Mid-Atlantic Fishery Management Council 2010). The result is increasing costs for fuel and vessel maintenance for owners. Because real prices for surfclams have been nearly constant over the last few decades (Mid-Atlantic Fishery Management Council 2010), firms that cannot operate more efficiently are becoming less profitable. In 2008 the second largest processor, located in Mappsville, VA, ceased operations completely. The vertically integrated fleet owned by the Mappsville processor was then sold to another surfclam processor with operations in Maryland and New Jersey. In general, many of the active surfclam vessels are owned by processors located in Maryland, New Jersey and Rhode Island. The overall industry trends appear to favor further consolidation and continued northward movement of the active fishing due to climate change (Weinberg 2005). An additional factor driving the consolidation is that ITQs can act as a barrier to entry for new firms. Without owning quota, there is no way for a new firm to legally harvest in the fishery. Last, it is speculated that market power in the ITQ market may be a problem, as one processor is known to control a substantial portion of the quota. In fact, there is evidence that ”market” for ITQs does not behave like a spot market because very little buying and selling is happening. This development should concern fishery managers as the benefits of ITQs rest on this market being perfectly competitive. Given the institutional background of the fishery, there are many important considerations for future management decisions.

Effectively managing a commercial fishery requires managers to make use economic models to help inform policy decisions. Often times fishery managers work with data that contain measurement error (Griliches and Hausman 1986). Some examples of data that are often measured with error include, time-at-sea, harvest, vessel characteristics and the resource abundance, also known as the biomass or stock. The biomass is an estimated quantity for all fisheries. In the Mid-Atlantic surfclam fishery, the biomass is estimated using data from a stratified random sample (see Figure 1.1 for strata regions) of the fishery with a biological model (Northeast Fisheries Science Center 2010). This biological model, known as the KLAMZ model, gives estimates of the biomass and of the model variability. We make use of this additional data in Chapter 2.

The problem of using data measured with error with a Stochastic Frontier Model is explored in Chapter 2 of this dissertation. In that chapter we introduce a relatively new method for reducing bias caused by measurement error, simulation extrapolation (SIMEX) (Cook and Stefanski 1994). Using data from the Mid-Atlantic surfclam fishery, and building off previous work done by Brandt (2007), we estimate a stochastic frontier model without measurement error correction, and then apply SIMEX. We discuss the results of the naive and SIMEX estimates, noting the the SIMEX estimates are more consistent with economic theory. When looking at factors that affect technical efficiency, we also see that SIMEX estimates reflect the regional differences in resource abundance, as well as the age structure of the surfclam fleet. To measure technical efficiency in the industry we make use of Stochastic Frontier Analysis (SFA).

The two most commonly used methods for estimating production efficiency are stochastic frontier analysis (SFA) and data envelope analysis (DEA). Data envelope analysis (DEA) is widely used in assessing productive efficiency in fisheries (Felthoven 2002, Kirkley et al. 2004, Tingley et al. 2005). DEA estimates a production frontier but uses linear programming methodology to do so, giving a deterministic production frontier, instead of a stochastic one. Kirkley et al. (1995) point out that given the inherent stochastic nature of fisheries production data, including weather and other environmental shocks, stochastic frontier models are more appropriate. First proposed independently by Aigner et al. (1977) and Meeusen and Broeck (1977), the stochastic frontier model combines a deterministic production function with a stochastic component that captures deviations above and below the production frontier. This model can be used to measure technical efficiency of a firm. A firm is defined as technically efficient when it is using the minimal level of inputs to achieve a certain output.

The standard stochastic frontier model is modeled as a neutral shift in the production function, using the observed data to measure technical efficiency. The neutral-shift stochastic frontier model is an econometric model that has a two-part disturbance; the first term captures firm-level inefficiency, while the second term is random noise. The advantage of stochastic frontier analysis is that it combines the idea of a production frontier with a stochastic component, which allows for random shocks. For a panel of firms producing a single output, the stochastic frontier model can be expressed as

$$y_{it} = f(\mathbf{x_{it}}; \beta) - u_i + v_{it} \tag{1.1}$$

where y_{it} is the natural log of output for the i^{th} firm in the t^{th} time period, $\mathbf{x_{it}}$ is a vector of the natural log of inputs for the i^{th} firm in the t^{th} time period, and β is a vector of technology parameters (Kumbhakar 2000). The fixed part of the production frontier can take many forms, with the Cobb-Douglas and Tran-logarithmic forms being the most popular. A Cobb-Douglas production model assumes constant input substitution elasticities, while the translog model is a more flexible functional form based on a second-order Taylor series. The two-part error term in the model captures both random shocks and firm-level inefficiency. The v_{it} term captures random shocks and is assumed to be identically and independent distributed with mean zero and constant variance. The second error component, u_i, is assumed to be a non-negative term which captures time invariant firm-level technical inefficiency. The model allows a firm i to operate on or beneath its production frontier, according to whether $u_i = 0$ or $u_i \geq 0$. Figure 1.2 provides a picture of the standard stochastic frontier model.

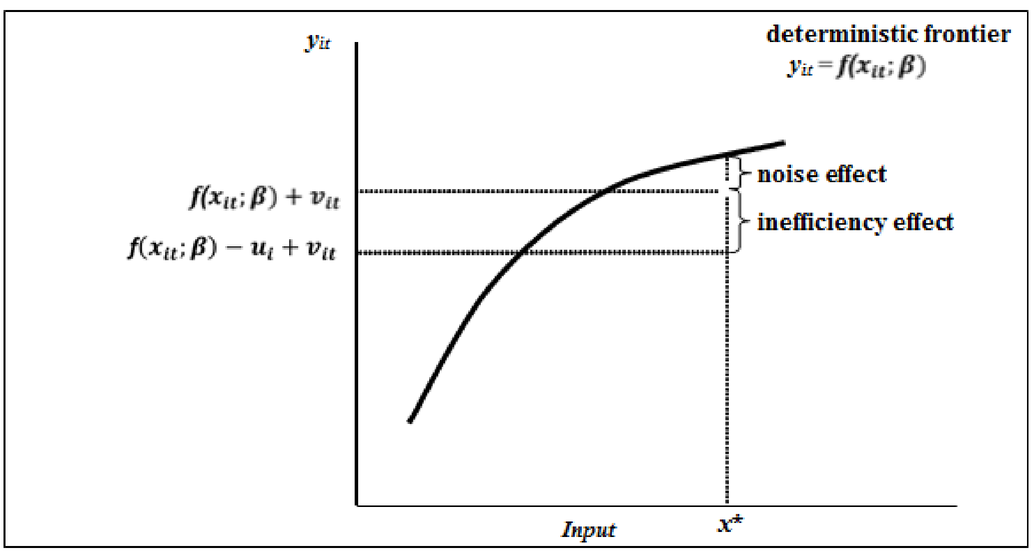

Figure 1.2: Measuring Technical Efficiency

There are several important considerations when estimating a stochastic production frontier, particularly with regard to the distributional assumption on the technical inefficiency term. Common distributional assumptions include the half-normal (Aigner et al. 1977), truncated normal (Battese and Coelli 1995), exponential (Meeusen and Broeck 1977), and gamma distribution (Greene 1990). The literature does not suggest a clear preference for a distribution, but instead is left to the empirical researcher. A fixed and random effects method can be used to estimate the one-sided inefficiency term. Which method is used depends on whether correlation exists between the regressors and the compound disturbance Schmidt and Sickles (1984). Assuming no correlation between the disturbance and regressors, a random effects model can be specified. However, if there is significant correlation, a fixed effects model is appropriate. Typically a Hausman test is performed to conclude which of the two models is appropriate. The objective of these models is not only to produce estimates of industry level efficiency, but also firm specific estimates (Kumbhakar 2000).

Much of the early work in stochastic frontier analysis focused on cross-sectional data, where a group of firms were observed in a single time period. Unfortunately, cross-sectional data do not allow for consistent estimation of technical efficiency, a result of not having repeated observations on firms (Kumbhakar 2000). Panel data provides a richer set of observations, because producers are observed many times, leading to better estimates of technical efficiency and better consistency properties for estimators in the model. It also has significant advantages over cross-sectional data when estimating technical efficiency for individual producers. Schmidt and Sickles (1984), Battese and Coelli (1995), Kumbhakar (2000) examine stochastic frontier models in a panel data setting, looking at both parametric and non-parametric approaches to modeling the inefficiency term.

Research in technical efficiency of renewable resource-based industries was very limited until the mid 1990s. Kirkley et al. (1995) use a stochastic frontier model with a translog parametric form to estimate technical efficiency in the Mid-Atlantic sea scallop industry. They conclude that the stochastic frontier approach is better suited for technical efficiency analysis in fisheries, than either the parametric or nonparametric programming approach, due to the inherent stochastic shocks in the industry, such as weather and captain's ability. Kirkley et al. (1998) look at the effects of managerial skill on technical efficiency in the same fishery. They find that education and experience may be substitutes for "good captains" and attempt to model this in a stochastic production frontier. They use technical efficiency expressions, first proposed by Jondrow et al. (1982) and later generalized by Battese and Coelli (1992, 1995) to estimate skipper skill. In the fisheries literature, established factors that determine the ability of a vessel to produce near the production frontier include, skipper skill (Kirkley et al. 1998), and vessel age (Pascoe and Coglan 2002).

With data on input and output prices, the stochastic frontier model can be used to estimate allocative and cost efficiency. Unfortunately, reliable data on input and output prices are not available for the Mid-Atlantic surfclam industry (Walden et al. 2012), so we make use of production data only in this study. It is the case that good data on input and output prices is hard to come by in the majority of fisheries worldwide. Using these same production data, we estimate a generalized Schaefer production model in Chapter 3.

The Schaefer production model is commonly used to estimate biological parameters for a fishery using catch and effort data (Zhang and Smith 2011, Punt 1992, Uhler 1980). Chapter two explores the issue of measurement error in the Schaefer production model, caused by the estimating the model using proxy variables. Because these biased estimated parameters are then used to estimate the biomass, fishery managers will have poor data to work when determining total catch. Measurement error in the Schaefer production model has been explored previously (Uhler 1980, Zhang and Smith 2011). Uhler (1980) shows that under certain circumstances the biomass estimates can be biased upwards of 40%. This is important because a fishery manager using the naive estimates from this model could make poor policy decisions that could lead to overfishing and loss of economic efficiency. In Uhler's example, the manager would set the total allowable catch (TAC) too high, and possibly cause the fishery to collapse in the future. Not only would this be a biological disaster, but also an economic disaster as all profits would be lost. While this scenario may appear extreme, there are many examples of regulated fisheries that have experienced this result (Hilborn et al. 2003).

There are several reasons why investigating the effects of measurement error in the Schaefer production model is important. First, National Standard One the Magnuson Stevens Conservation Act of 2007 (U.S. Department of Commerce 2007) states that any new fishery management policies must prevent overfishing, while achieving the optimum yield. National Standard Two states that new fishery management measures should use the latest scientific methods. Second, the use of production data in estimation of biological parameters is common in fishery management (Zhang and Smith 2011). Because fishery managers do not always possess scientific estimates of the fishery biomass, they must infer it using production data and the Schaefer production model.

Two quantities of interest for fishery managers are maximum sustainable yield (MSY) and maximum economic yield (MEY). MEY is attained where the difference between the total revenue curve and total cost curve is the largest, shown in Figure 1.3. The bioeconomic equilibirum (BE) shown in the figure, represents the typical common pool resource problem where fishermen increase effort until profits are zero.

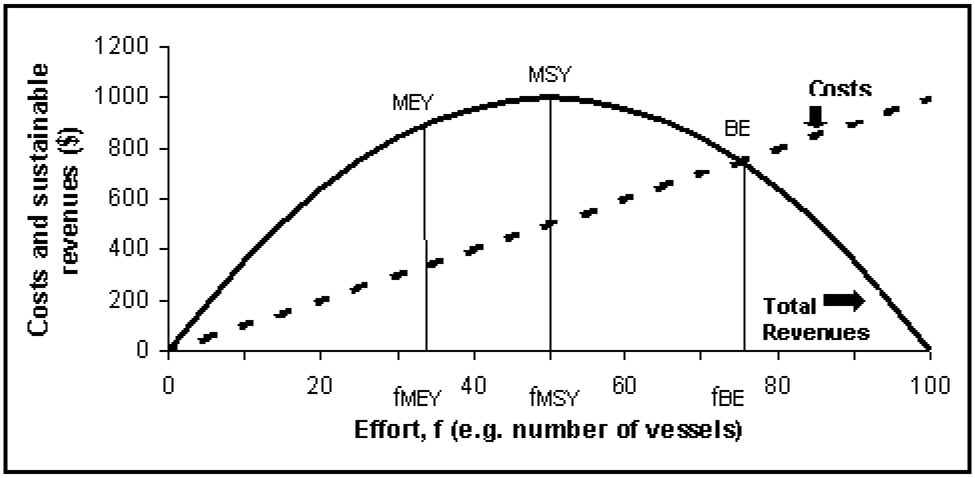

Figure 1.3: Effort-Yield Curve

Cochrane (2002)

MSY refers to the point where the maximum harvest is obtained in a given year, and the fishery biomass will remain constant. This occurs because the growth rate of the fishery is equal to the harvest. While the MSY may seem like an optimal target for a fishery manager, it does not take into account the costs of fishing, nor the stochastic dynamics which can quickly send a fishery towards collapse. Both the MSY and MEY are concepts which make sense in a fishery which is in steady state, but this is often not the case. The steady state assumption is not realistic in many fisheries due to the natural shocks that can occur from environmental factors. It is important to note that fishery managers and resource economists are typically concerned with managing a resource in such as way as to maximize the net present value of benefits to society. This is a dynamic concept and requires significantly more information about the discount rate and the value of resource in the future. This can result in a different optimal harvest than either MSY or MEY would suggest.

In order to estimate the MEY or MSY, a fishery manager needs information about the biological parameters of a fishery. The Schaefer production model is a two-stage model which is widely used to estimate these parameters. In the first stage a production function is specified. In the second stage a growth model for the biomass is specified. Let H_t be total harvest, E_t total effort, and X_t be the unobserved biomass or stock, and q a parameter which measures the "catchability" of the stock. All variables are indexed by time t. The classic Schaefer production model is

$$H_t = qE_tX_t \tag{1.2}$$

and

$$X_{t+1} = X_t + rX_t(1 - \frac{X_t}{K}) - H_t \tag{1.3}$$

where the intrinsic growth rate r and carrying capacity K are biological parameters. If biomass, harvest and effort data are all available, this model can be estimated under regularity conditions (Zhang and Smith 2011). However, the biomass of the fishery is almost never observed, and is therefore a latent variable to the fishery manager. To estimate the above model when biomass is unknown a proxy is used, such as catch-per-unit-effort (CPUE). Let y_t denote CPUE. Then CPUE is $y_t = \frac{H_t}{E_t}$. From equation 1.2, CPUE is proportional to the unobserved biomass, such that $X_t = \frac{y_t}{q}$. This last equation allows fishery managers to estimate the biomass.

The main issue with using the CPUE proxy is the measurement error brought into the estimation of the production model. This leads to biased and inconsistent estimates of biological parameters in the second stage growth model, and thus biased estimates of the biomass. In their recent paper, Zhang and Smith (2011) propose a two-step method that makes use of the panel data to correct for the bias. Chapter 3 extends that paper by proposing an additional step by using simulation extrapolation (SIMEX) (Cook and Stefanski 1994) to reduce the bias caused by measurement error.

Chapter 4 specifies a time-varying stochastic frontier model and uses both Bayesian and maximum likelihood estimation (MLE) methods to examine changes in production efficiency of the Mid-Atlantic surfclam fleet between 2001 and 2009. In particular, we examine the impact of a market crisis in 2005, which resulted in much of surfclam fleet exiting the fishery. Our results show that marginal productivities of effort and capital are declining occurring over time, resulting in higher harvesting costs for firms. We also see a significant decrease in mean technical efficiency after the 2005 market crisis, which brings more questions about whether increased consolidation will return greater efficiency gains. This chapter concludes by reflecting on changes in production efficiency observed in the data, and how continued consolidation of the surfclam fleet is being affected by climate change, and the industrial organization of the industry

Previous studies of Stochastic Frontier Analysis using Bayesian methods include (Osiewalski and Steel 1998, Ehlers 2011, Griffin and Steel 2007, Fernandez et al. 1997, Tsionas 2005, Van den Broeck et al. 1994). (Van den Broeck et al. 1994) first introduced Bayesian methods for Stochastic Frontier models, showing the advantages of exact, small-sample inference for efficiencies, as well as how to incorporate prior information into the model. We use the Bayesian methodology in this Chapter because the data effectively represent the population, capturing all vessel trips in the federally regulated areas of the Mid-Atlantic surfclam fishery. Because the data are population data and not sample data, a repeated sampling design, which used in the frequentist statistical framework, does not make sense. We discuss more advantages of Bayesian estimation methods for social science research in the motivation of Chapter 4.

We follow Ehlers (2011) , who illustrated estimation of stochastic frontier models using JAGS (Just Another Gibbs Sampler) (Plummer et al. 2003) in R (R Core Team 2013). Griffin and Steel (2007) outlined the estimation of a Stochastic Frontier Model using another popular Bayesian software platform for social sciences, WinBUGS. One particular advantage of using Bayesian methods is that posterior inferences for technical efficiency are easy to produce using readily available software, and often have smaller variances than frequentist methods. Following (Ehlers 2011), we specify the conditional posterior distributions for the parameters in the model. Important considerations for Bayesian Frontier models include assumptions for prior distributions and model functional form. We place normal priors on all the parameters in the model, and a half-normal prior on the technical inefficiency term. We also specify a Cobb-Douglas production model, similar to Chapter 2.

CHAPTER 2

MEASUREMENT ERROR IN A STOCHASTIC FRONTIER MODEL

2.1 Introduction

Technical efficiency is defined as the ability of a firm to produce the maximum output given a level of inputs (Kumbhakar 2000). In natural resource-based industries such as fishing, the measurement of technical efficiency is important to policy decisions regarding fishery management and preventing overfishing. Fishery managers use technical efficiency data to determine whether current management practices allow firms to operate efficiently, meaning close to the production frontier. Managers can also use this data to compare alternative policies. For example, in fisheries where excess capacity is an issue, managers would expect technically inefficient firms to leave the industry if proper incentives are put in place. This could happen when a fishery transitions from limited entry or some other command and control management, to a market-based approach such as Individual Tradable Quotas (ITQs).

To know whether fishery policies are working as expected, managers can use a stochastic frontier model to estimate technical efficiency for the industry. These models are widely used in fisheries to understand which factors significantly impact the ability of firms to operate efficiently. Panel data that contain unknown amounts of measurement error are commonly used in this type of modeling (Griliches and Hausman 1986). Logbook, survey, landings and biomass data all potentially have measurement error, meaning that naive estimation using one or more of these data could result in biased and inconsistent parameter estimates. This article applies a Monte Carlo method for reducing bias caused by measurement error called Simulation Extrapolation (SIMEX) (Cook and Stefanski 1994). We make use of two unique datasets, one containing logbook data from the Mid-Atlantic surfclam industry from 2001-2009, and the other scientific biomass estimates for the fishery (Mid-Atlantic Fishery Management Council 2010). We make use of the biomass data to get estimates of the measurement error variance, and use this with the SIMEX estimator.

The two most popular methods for estimating technical efficiency are data envelope analysis (DEA) and the stochastic frontier model (Aigner et al. 1977, Meeusen and Broeck 1977). The stochastic frontier model is a production function that has a two-part error term. The first term captures firm-level inefficiency and the second term captures random noise. Recent advances in econometric modeling have looked at stochastic frontier models with time-varying technical efficiency (Kumbhakar 1990) , random coefficients (Tsionas 2002) and Bayesian approaches (Kim and Schmidt 2000). Data envelope analysis (DEA) is another popular technique for assessing productive efficiency in fisheries (Felthoven 2002, Kirkley et al. 2004, Tingley et al. 2005). DEA also estimates a production frontier but uses linear programming methodology to do so, giving a deterministic production frontier, instead of a stochastic one.

Research in technical efficiency of renewable resource-based industries was very limited until the mid 1990s. Kirkley et al. (1995) use a stochastic frontier model with a translog parametric form to estimate technical efficiency in the Mid-Atlantic sea scallop industry. They conclude that the stochastic frontier approach is better suited for technical efficiency analysis in fisheries, as opposed to other methods such as the parametric or nonparametric programming approach. The advantage of the stochastic frontier model is the two part error term captures inherent stochastic shocks in the industry, such as weather, and separates these effects from the inefficiency term. Past research has established important factors that make vessels more technically efficient, such as captain's skill (Kirkley et al. 1998, Squires and Kirkley 1999), and vessel age (Pascoe and Coglan 2002).

Another use of measuring technical efficiency is in determining whether changes in management lead to overall efficiency gains (Grafton et al. 2000). Previous work in estimation of technical efficiency in the Mid-Atlantic surfclam fishery by Brandt (2007) found gains from a change in management from limited access to Individual Transferable Quotas (ITQs) in 1990. Recently, Walden et al. (2012) found that the Mid-Atlantic surfclam industry has not seen long term productivity gains from the switch to ITQs, possibly due to spatial changes in the biomass. Since the introduction of ITQs the fishery has seen considerable consolidation, with many of the smaller, independent vessels exiting the fishery. An important question this article will attempt to answer is what factors are important in explaining the behavior currently observed, and how technical efficiency plays a role. All else equal, theory would predict that vessels that are more technically efficient will be more profitable and less likely to exit (Grafton 2006).

The data for this study are logbook data from the Mid-Atlantic surfclam fishery. The data describes input and output relationships for seventy separate vessels that harvested surfclams during the years 2001-2009. In addition to the logbook data, variables on vessel characteristics such as age, hull material, and port location of the vessels are obtained. The factors such as vessel age and the hull material should be important in explaining technical efficiency. A measure of the resource availability, biomass, is also obtained from the Northeast Fisheries Science Center (NEFSC) (Northeast Fisheries Science Center 2010). We note that the biomass is only an estimated quantity, obtained from a method that uses data from a stratified random sample of the fishery with a scientific growth model. As we show later in the article, the estimated biomass contains significant amounts of noise, a potential measurement error problem. Because measurement error can lead to loss of statistical power and can mask important relationships in the data, important boat characteristics that determine technical efficiency may not appear significant.

Measurement error, also called error-in-variables, in regression models is a vexing problem for empirical researchers. When measurement error is present in the covariate(s) of a regression model there are three effects a researcher should be worried about; 1) biased parameter estimates; 2) loss of power for detecting interesting relationships among variables; and 3) the measurement error masking features of the data, making graphical analysis difficult (Carroll et al. 2012). Without instrumental variables or some knowledge of the measurement error variance, empirical researchers are left with no other alternative than to proceed cautiously with naive estimation of the model, knowing there is some level of bias in the parameter estimates. Survey data, logbooks, landing files and estimates of resource abundance all potentially contain measurement error in fisheries data (Grafton 2006). More generally, when using observational data measured with error in a linear model, the correlation between variables with measurement error will result in parameter estimates that will be biased not only for the variable measured with error, but all variables correlated with it in the model (Hausman 2001, Greene 2003). This article will show how another measurement error correction technique called Simulation Extrapolation (SIMEX) to obtain bias-reduced parameter and technical efficiency estimates.

We proceed with estimation of technical efficiency using a stochastic frontier approach. Both the naive and SIMEX estimates are obtained for the same specified model, and technical efficiency estimates are derived using a random effects model (Kumbhakar 2000). After some exploratory data analysis, the functional form for the stochastic frontier model is specified. We follow Brandt (2007) and specify a Cobb-Douglas production technology, which says that the natural log of output is linear in the natural log of inputs. The measurement error problem arises in the estimation of the model because the covariate $log(biomass)$ is assumed measured with additive error. To reduce the bias induced by the measurement error additional data is brought into the model. The Northeast Fisheries Science Center (NEFSC) biomass survey contains both an estimate of the biomass and the sampling variability. We use this estimate of sampling variability as the measurement error variance, and this additional information is used to reduce the measurement error bias using simulation extrapolation (SIMEX). SIMEX has been shown to provide approximately consistent parameter estimates under a variety of measurement error models. The SIMEX method is also used to obtain estimates of the standard errors using the sandwich estimator (Carroll et al. 2012).

After obtaining both the naive and SIMEX estimates of the fixed parameters and technical efficiency measures, the two sets of estimates are compared. We find bias in the all parameters in the model. More importantly, significant factors that explain technical efficiency are found using the SIMEX estimates, whereas they are not found to be significant in the naive model. In particular, we find evidence that vessel age, hull material and region are important factors in explaining technical efficiency. Technical efficiency is found to be lower in the southern regions of the fishery. One possible explanation for this difference is because vessels are older in the southern regions because firms are not investing in new capital. This is likely due to climate change and negative impacts on the resource abundance (McCay et al. 2011, Weinberg 2005).The regional differences in technical efficiency are also likely related to the industrial organization of the fishery, with vessels owned by processors operating closer to the production frontier than independently owned vessels.

The remainder of the paper proceeds as follows: In Section 2.2, the Mid-Atlantic surfclam fishery is described in more detail. In Section 2.3, the logbook and biomass data are described, and the measurement error problem is motivated. In Section 2.4, the theoretical framework for the stochastic frontier model, measurement error model, and SIMEX are presented. In Section 2.6, we present results from the naive and measurement error corrected models. Section 2.7 describes a Monte Carlo study to determine the large sample properties of the SIMEX estimator. Section 2.8 summarizes the findings of the paper and presents topics for future research.

2.2 Mid-Atlantic Surfclam Fishery

The Mid-Atlantic surfclam fishery spans the U.S. eastern Atlantic coast from the southern Gulf of St. Lawrence to Cape Hatteras, shown in Figure 2.1. Atlantic surfclams are a bivalve mollusk distributed along the coast of North America. Overall the biomass is currently in a period of decline after reaching record levels in the 1990s. The southern end of the fishery, particularly in the Delmarva region, has seen a higher mortality rate in the past decade, leading declining biomass due to climate change (McCay et al. 2011, Weinberg 2005). The highest concentration of biomass is off the northern New Jersey coast, which has resulted in most of the fishing effort being concentrated in this region. Overall, landings per-unit-effort (LPUE) has declined substantially for the fishery as a whole, largely due to concentrated fishing effort (Northeast Fisheries Science Center 2010).

In 1990 the fishery transitioned from limited entry to ITQs under the direction of the Mid-Atlantic Fishery Management Council. The current management measures include an annual quota for Exclusive Economic Zone (EEZ) waters and mandatory logbook entries for each vessel trip. The Mid-Atlantic surfclam industry has consolidated considerably since the introduction of ITQs, going from ap-

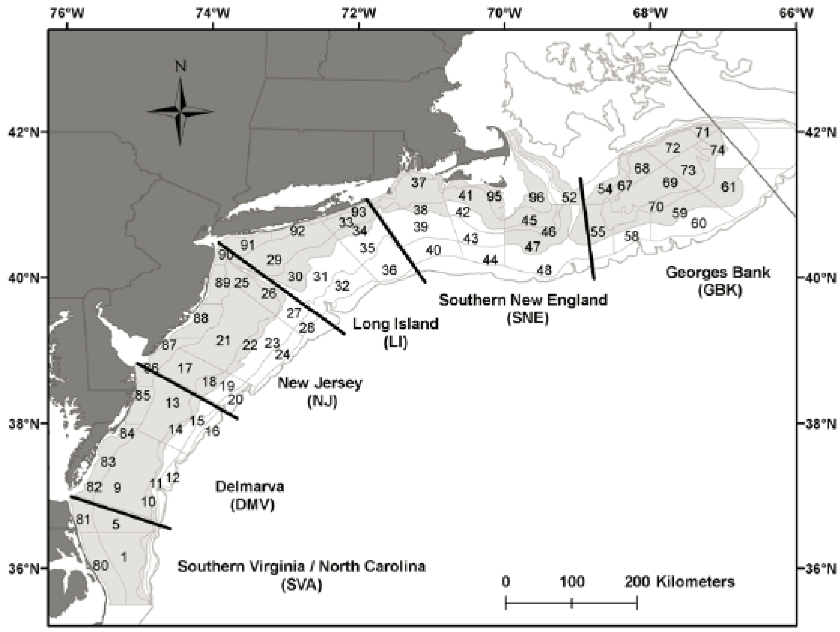

Figure 2.1: Mid-Atlantic Surfclam Fishery

Northeast Fisheries Science Center (2010)

proximately 120 vessels in 1990 to fewer than 50 vessels in 2005. As detailed in Brandt (2007), many of the vessels that exited just after the regulatory change were very inefficient. The remaining fleet consists of a small number of horizontally and vertically integrated firms, with a few independent vessel owners. Nominal revenues for the fleet in 2011 were approximately $29 million (Northeast Fisheries Science Center 2010).

2.3 Data

The data for the empirical analysis come from the National Marine Fishery Service logbook reporting system, which documents every harvesting trip taken by every vessel in the Mid-Atlantic surfclam fishery in the U.S. EEZ (3-200miles offshore). The logbook data are a panel data set containing approximately 24,000 vessel-trip observations, for years 2001-2009. The trip-level data set includes variables such as bushels harvested, time fishing, time-at-sea, and vessel characteristics such as vessel length, gross-tons and horsepower. There are a total of eighty-eight different vessels observed over the nine year period.

To simplify the correlation structure within each vessel and because biomass is observed annually, data are aggregated by vessel-year. The new data set has one observation for each vessel in a year. One trade off of using aggregated data is that trip-level variability is not observed. Using the aggregated data also means making certain assumptions about the measurement error model structure, which is discussed in section 2.4. Before aggregating the data, the same linear model was estimated using both sets of data. The estimation results did not change substantially, further suggesting that the aggregated data are more appropriate. The resulting data are reduced to 70 vessels and 285 vessel-year observations. Summary statistics for the data can be found in Table 2.1.

	Obs	Mean	Std.Dev	Min	Max
Harvest (bushels)	285	93749	82007.6	864	442496
Time Fishing (hours)	285	1209.2	951.9	58	3959.4
Fuel (gallons)	285	65896	65979.3	876	388204
Length (feet)	70	85.7	18.4	28	162
Biomass (1000 metric tons)	9	1037	171.9	750	1294

Table 2.1: Summary statistics 2001-2009

In order to estimate the stochastic frontier for the fishery, we need a measure of the resource abundance, similar to a measure of land quality in agriculture. This is a non-traditional input, but important to the model since resource abundance or biomass will affect harvest. We use estimates of biomass from the National Oceanic and Atmospheric Association's (NOAA) Northeast Fisheries Science Center (NEFSC)-Resource Evaluation and Assessment Division. These estimates are obtained by a biological model called the KLAMZ model.

2.3.1 KLAMZ model

The KLAMZ model is the primary model used to determine the status of the biomass for the Mid-Atlantic surfclam fishery and used in fishery management decisions for setting the quota (Northeast Fisheries Science Center 2010). The KLAMZ assessment model is based on the Deriso-Schnute delay-difference equation. This model accounts for growth rates and mortality rates by clams from different age groups. These growth patterns are also allowed to vary over time. To calibrate the model, survey tows are conducted. A stratified random sample of the fishery is conducted using a survey dredge, which harvests the surfclams from the ocean floor. For each survey strata, denoted by the numbered areas in figure 2.1, estimates of the mean (kg/tow) of surfclams can be calculated. These estimates are used to inform trends in the growth of certain age groups of clams.

Based on the variability of the survey tows, known as survey dredge efficiency, an estimate of the coefficient of variation (CV) for the sampling process can be obtained. More specifically, the CV was estimated by bootstrapping the median of all survey dredge efficiency estimates. The CV is the inverse of the single-to-noise ratio, also known as the unitized risk. According to the Northeast Fisheries Science Center (2010) the CV for the sampling data is about 0.14. Because the survey data is assumed to be log-normally distributed we estimate the standard deviation for the data on the natural log scale s using the formula

$$CV = (e^{s^2} - 1)^{1/2} \tag{2.1}$$

We use this estimate of the sampling variability later in the study to construct the measurement error variance for the biomass, described in section 2.4. A boxplot of the bootstrapped biomass estimates from the KLAMZ model for years 1980-2008 are shown in figure 2.2. Inspection of the data shows that sampling variability for the biomass changes by year, which is later incorporated into the measurement error model. Because the biomass is measured using stratified random sampling, we assume that the sampling variability is independent from year to year.

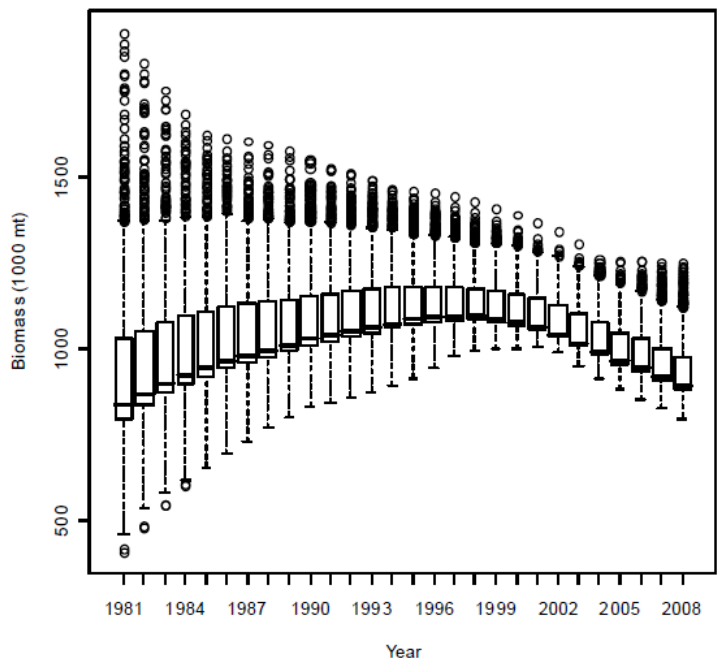

Figure 2.2: Boxplots with bootstrap biomass estimates for KLAMZ model

Northeast Fisheries Science Center (2010)

2.4 Methodology

Following previous work in this fishery by Brandt (2007), we specify a Cobb-Douglas model for production. This assumes $log(output)$ is linear in the sum of the $log(inputs)$. This functional form allows the coefficients to be interpreted as input elasticities, meaning each β_k represents the percentage change in output due to a 1% increase in input k. Although it is a convenient functional form, it does impose constant elasticity of substitution on inputs, something a more flexible form, such as translog model, does not. With these limitations in mind we proceed with estimation of the model. We first specify a random effects model, or linear mixed model for production following Kumbhakar (2000). The random effects model is then transformed into a stochastic frontier model.

Using the Cobb-Douglas model of production, the following variables are included in the model. Let $i = 1, \ldots, 70$ denote vessel and $t = t_1, \ldots, t_{n_i}$ denote the n_i years in which vessel i is observed.

y_{it} = log_e(total bushels harvested by vessel i in year t),

x_{it1} = log_e(total hours fished by vessel i in year t),

x_{it2} = log_e(total gallons consumed by vessel i in year t),

x_{it3} = log_e(length of vessel i in year t)

w_{t4} = log_e(biomass in year t).

Note that the observed $log(biomass)$ is identified as w_{t4}, to denote that it is measured with error. Later we specify a measurement error model in which the observed $log(biomass)$ is a function of true $log(biomass)$, x_{t4}, plus error. The inclusion on biomass in the model is based in economic theory of production, which says that a measure of resource abundance, such as land in agriculture, is important to the model. Time fishing is included as a proxy for fishing effort and fuel is included as another input in production. The length of the vessel is a standard proxy as a measure of capital in the fisheries literature.

2.4.1 Linear Mixed Model of Production

The standard stochastic frontier model is a linear model with a two-part disturbance term. The first part is a one sided error term for measuring technical inefficiency, while the second is a two-sided white noise term. The production parameters are fixed in this specification, while the two-part error term is random, meaning we can write the stochastic frontier model as a linear mixed model. Let

b_i be the vessel-level random effect, and let e_{it} be within vessel errors. The vessel-level effect is specified as a random effect, following assumptions of the stochastic frontier model Kumbhakar (2000). A linear mixed model for production can be written as

$$y_{it}|b_i, e_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it3} + \beta_4 w_{t4} + b_i + e_{it} \tag{2.2}$$

with $b_i \overset{\text{ind.}}{\sim} N(0, \sigma_b^2)$ and $e_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_e^2)$, $i = 1, \ldots, 70$, $t = t_1, \ldots, t_{n_i}$. The normality assumption is not crucial in this specification. This is a time-invariant specification for the random effect b_i, which will then be transformed into the one-sided inefficiency term. We assume a time-invariant inefficiency term because vessel captains stay with their boats over long periods of time, and the characteristics of the vessels themselves do not change significantly over time. Technical efficiency is defined as the ratio of a firm's realized output to its potential output. In the model above, the random effect b_i should capture vessel specific characteristics that are not part of the production process, such as captain's ability or experience, age of the vessel or crew motivation. In theory, these unobserved random effects will determine how close the vessel operates to its potential output. Before specifying the stochastic frontier we build the model for each vessel, and then the entire fleet.

Next, notation for a version of the model for vessel i is defined. Let

$$\mathbf{y_i} = (y_{it_1}, \ldots, y_{it_{n_i}})^T, \mathbf{X_i} = \begin{pmatrix} 1 & x_{it_1 1} & x_{it_1 2} & x_{it_1 3} & w_{t_1 4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{it_{n_i} 1} & x_{it_{n_i} 2} & x_{it_{n_i} 3} & w_{t_{n_i} 4} \end{pmatrix}, \tag{2.3}$$

$$\beta^T = (\beta_0, \ldots, \beta_4), \mathbf{e_i}^T = (e_{it_1}, \ldots, e_{it_{n_i}}), \text{ and } \mathbf{1}_{n_i} = \text{ a vector of length } n_i \text{ of all 1s.} \tag{2.4}$$

Using that notation, a vessel level model is

$$\mathbf{y}_i|b_i, \mathbf{e}_i = \mathbf{X_i}\beta + \mathbf{1}_{n_i}b_i + \mathbf{e_i} \tag{2.5}$$

with $b_i \overset{\text{ind.}}{\sim} N(0, \sigma_b^2)$ and $\mathbf{e_i} \overset{\text{i.i.d.}}{\sim} MVN(0, \sigma_e^2\mathbf{I_{n_i}})$.

Finally, with

$$\mathbf{y} = \begin{pmatrix} \mathbf{y_1} \\ \vdots \\ \mathbf{y_{70}} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X_1} \\ \vdots \\ \mathbf{X_{70}} \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} \mathbf{1_{n_1}} & \mathbf{0_{n_1}} & \dots & \dots & \mathbf{0_{n_1}} \\ \mathbf{0_{n_2}} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots\, \mathbf{0_{n_{69}}} & \\ \mathbf{0_{n_{70}}} & \dots & \dots & \mathbf{0_{n_{70}}} & \mathbf{1_{n_{70}}} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_{70} \end{pmatrix}, \tag{2.6}$$

$$\text{and } \mathbf{e} = \begin{pmatrix} \mathbf{e_1} \\ \vdots \\ \mathbf{e_{70}} \end{pmatrix}, \tag{2.7}$$

Finally, we specify a linear mixed model for the entire fleet. Using matrix notation, a linear mixed model for the fleet is

$$\mathbf{y}|\mathbf{b}, \mathbf{e} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e} \tag{2.8}$$

with $\mathbf{b_i} \overset{\text{ind.}}{\sim} N(\mathbf{0_{70}}, \sigma_b^2\mathbf{I_{70}})$ and $\mathbf{e} \overset{\text{i.i.d.}}{\sim} MVN(\mathbf{0_{285}}, \sigma_e^2\mathbf{I_{285}})$.

2.4.2 Stochastic Production Frontier

In the productivity literature (Kumbhakar 2000) there are several methods for transforming the random effects in order to calculate technical efficiency. We follow the standard transformation for a random effects model. Before calculating technical efficiency the b_i's, which are the empirical Best Linear Unbiased Predictors (eBLUPs), they need to be normalized. Let $\hat{b}_{i^*} = max_j[\hat{b}_j] - \hat{b}_i$ be the normalized random effect term.

This normalization makes the (eBLUPs) a non-negative random variable. In order to calculate vessel-level technical efficiency a distributional assumption must be placed on the b_{i^*}'s. There are no *a priori* reasons to choose one distribution over another for the technical inefficiency term. The productivity literature typically uses the half-normal, truncated normal and exponential. Following Kirkley et al. (1995) the technical inefficiency term b_{i^*} is assumed to be distributed as a half-normal, $|N(0, \sigma_b^2)|$. The stochastic production frontier model is then specified as

$$y_{it}|b_{i^*}, e_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it3} + \beta_4 w_{t4} - b_{i^*} + e_{it} \tag{2.9}$$

with $b_{i^*} \overset{\text{ind.}}{\sim} |N(0, \sigma_b^2)|$ and $e_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_e^2)$, $i = 1, \ldots, 70$, $t = t_1, \ldots, t_{n_i}$. [1]

2.4.3 Calculating Technical Efficiency

Following Jondrow et al. (1982), technical inefficiency for each observation is calculated as the expected value of $\hat{b}_{i^*}$, conditional on $\epsilon_i = e_i - b_{i^*}$, where $e_i = \sum_{t=t_1}^{t_{n_i}} e_{it}$. Technical inefficiency for vessel i can be calculated as

[1]Estimation of the model is performed using the linear mixed effects models package "nlme" (Pinheiro et al. 2013) in the R statistical software (R Core Team 2013). The estimates for σ_b^2 and σ_e^2 are found using Restricted Maximum Likelihood or REML. The fixed effects, $\beta's$, are computed using maximum likelihood under the assumption of normality. REML is a form of maximum likelihood estimation that uses a transformed version of the data so that nuisance parameters have no effect on the estimates. It has been shown to provide less biased estimates of the variance-covariance parameters than maximum likelihood McCulloch and Searle (2000).

$$TI_i = \frac{\sigma_b \sigma_e}{\sigma}\left[\frac{\frac{\phi(\epsilon_i \lambda)}{\sigma}}{1 - \Phi(\frac{\epsilon_i \lambda}{\sigma})} - (\frac{\epsilon_i \lambda}{\sigma})\right] \tag{2.10}$$

where $\phi(.)$ is the standard normal density, $\Phi(.)$ is the cumulative normal distribution, $\sigma = (\sigma_b^2 + \sigma_e^2)^{1/2}$ and $\lambda = \frac{\sigma_b}{\sigma_e}$. The vessel-specific technical efficiency estimate is given as $TE_i = exp(-TI_i)$. After calculating technical efficiency for each vessel we conduct post-estimation to determine which factors determine a vessel's ability to operate on its production frontier. We also plot the distribution of technical efficiency under both the naive and SIMEX estimates, to see if significant differences can be seen.

2.4.4 Measurement Error Model

The stochastic frontier model specified above reflects an ideal world where the data obtained contains no measurement error. In reality, the biomass is an estimated quantity, with significant noise for a given year. The presence of this measurement error has consequences under the classical regression model. The classic regression model assumption of strict exogeneity is crucial for unbiased parameter estimates, i.e. $E[\mathbf{e}|\mathbf{x}] = 0$. This means that on average the disturbance term should be equal to zero for a given level of the covariates. Assuming biomass is measured with error, and this error is correlated with the error in the regression, the strict exogeneity condition will not hold. The result is that our classic regression model assumptions for unbiased parameters no longer apply (Greene 2003). Additionally, since the biomass is correlated with other variables in the model, all parameter estimates will be biased under naive estimation.

Using the fact that the biomass is estimated through a stratified random sampling method, we assume it to be independent from year to year. Because the biomass is only available annually, we further assume that vessels face a constant biomass in each year. The additive measurement error model for observed $log(biomass)$ is specified as

$$w_{t4} = x_{t4} + v_t \tag{2.11}$$

where x_{t4} is the true biomass in year t and v_t is the error. The distribution for measurement error is specified as $v_t \overset{\text{ind.}}{\sim} N(0, \sigma^2_{vt})$. The measurement error model above requires $E[v_t|x_{t4}] = 0$ but does not require $v_t \perp\!\!\!\perp x_{t4}$, a much stronger assumption. This means that the variance in the stock index can vary over time, but the errors are not serially correlated, meaning the measurement error variance is heteroscedastic, in congruence with the sampling data. This also means that the observed biomass is unbiased for the true biomass, $E[w_{t4}|x_{t4}] = x_{t4}$. Finally, because the true σ^2_{vt} is not observed, we estimate it $\hat{\sigma}^2_{vt}$, which is taken from the biomass sampling data.

2.4.5 Measurement Error in a Linear Mixed Model

Using the additive measurement model specified in the previous section it is possible to motivate the measurement error problem in the estimation of the naive stochastic frontier model. Recall that the original linear mixed model for the fleet is

$$\mathbf{y}|\mathbf{b}, \mathbf{e} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e} \tag{2.12}$$

with $Var(\mathbf{e})$ = $\mathbf{R}$, and $Var(\mathbf{b}) = \mathbf{G}$. Let $\hat{\beta}$ be the restricted maximum likelihood or REML estimator for β. The estimating equations for $\hat{\beta}$ and $\hat{\mathbf{b}}$ are

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{2.13}$$

Let $\mathbf{W}$ be the $\mathbf{X}$ matrix as defined above, but with observed $log(biomass)$ replacing the true variable. Inserting the $\mathbf{W}$ matrix into the estimating equations in place of

$\mathbf{X}$ will result in biased an inconsistent parameter estimates. Furthermore, all variables correlated with observed $log(biomass)$ will also be biased and inconsistent. Even in large samples the naive estimator of this model with additive measurement error in $log(biomass)$ has the result $\text{plim}_{n\to+\infty}\hat{\beta} \neq \beta$.

Another concern with measurement error in the linear mixed model is that estimated standard errors, and estimates of technical efficiency will be biased. It can be shown in the case of a linear mixed model, the inconsistency in the parameter estimates will also lead to inconsistent estimated random effects, $\hat{\mathbf{b}}$, or eBLUPS. The consequences for the stochastic frontier model will be that technical efficiency measures are biased. The direction of bias for parameter estimates is hard to know, and will depend on the correlation structure of the covariates and parameter estimates. A comprehensive explanation of measurement error in linear mixed models can be found in Carroll et al. (2012) or Buonaccorsi (2010).

Using methods from Wang et al. (1998), it can be shown under the assumption of a normally distributed measurement error, a linear mixed model with additive error in one predictor can be written as another linear mixed model, a generalized linear mixed measurement error model or GLMMeM. Bringing in outside information can lead to identification of the parameters in this GLMMeM. SIMEX has been shown to give approximately consistent parameter estimates when estimating a linear mixed model with additive measurement error in a covariate. Another problem with measurement error is identification of the parameters in the model. We make use of outside information by estimating the measurement error variance. In the next section we show how the measurement error variance estimate is used with SIMEX to reduce bias in the parameter estimates.

2.4.6 SIMEX

SIMEX is a two-step simulation-based method of estimating and reducing bias due to measurement error. First, simulated data are obtained by adding additional measurement error to the data in a resampling-like process, establishing a trend of measurement error-induced bias versus the variance of the added measurement error. After that, the extrapolation step follows the fitted trend line back to a point where the measurement error variance is zero. The key underlying SIMEX is the fact that the effect of measurement error on an estimator can be determined experimentally through simulation (Carroll et al. 2012). It can be shown that under a number of different measurement error specifications that SIMEX provides approximately consistent parameter estimates. SIMEX is very general in the sense that the bias due to measurement error in almost any estimator of almost any parameter can be estimated and corrected, at least approximately. SIMEX is described below for the case of additive measurement error in the predictor in four steps, as explained in Buonaccorsi (2010).

Assume an additive error in the predictor $w_{t4} = x_{t4} + v_t$ and $Var(v_t)$ = σ^2_{vt}. Begin by defining $\theta_j(\lambda)$ as the expected (or limiting) value of the naive estimator of θ_j if $Var(v_t) = (1+\lambda)\sigma^2_{vt}$. Then true value of the jth coefficient is $\theta_j = \theta_j(-1)$.

1. For each λ_m, generate: $w_{t4b}(\lambda_m) = w_{t4} + \lambda_m^{1/2} U_{bt}$ for b = 1,...B, where B is a large number and the U_{bt} are independent with mean 0 and variance $\hat{\sigma}^2_{vt}$. Since $w_{t4}|x_{t4}$ already has variance σ^2_{vt}, the generated w_{t4b} would have exactly the variance $(1+\lambda_m)\hat{\sigma}^2_{vt}$ assuming $\hat{\sigma}^2_{vt} = \sigma^2_{vt}$. In practice we usually only have an estimate of σ^2_{vt}.

2. Find $\theta(\lambda_m, b)$, which is the naive estimator for θ_j based on $(\mathbf{y}, \mathbf{X})$. Then define: $\bar{\theta}(\lambda_m)$= $\sum_b \hat{\theta}(\lambda_m, b)/B$. So, $\bar{\theta}_j(\lambda_m)$ is the average of the B estimated $\hat{\theta}_j$'s at a particular λ_m .

3. For each j, fit a model $g_j(\lambda)$ for $\bar{\theta}_j$ (λ_m), the jth component of $\bar{\theta}_j$ (λ_m), as a function of λ_m.

4. Get the SIMEX estimate of θ_j using: $\hat{\theta}_j(j, SIMEX) = g_j(-1)$. Because the variance of w_{t4b} is exactly $(1+\lambda_m)\hat{\sigma}^2_{vt}$, at the point where $\lambda = -1$ the measurement error variance collapse to zero. This gives an approximately consistent estimate of the true parameter θ_j.

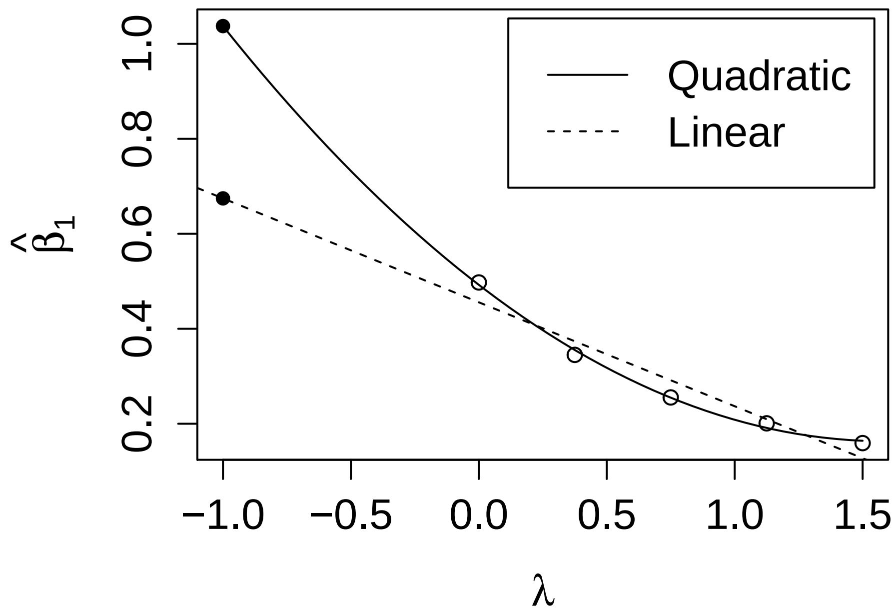

Figure 2.3: An example of the SIMEX Extrapolation Step

The last step of the SIMEX method is the extrapolation step. There are several functional forms which can be chosen, including the linear, quadratic and rational extrapolant functions. Figure 2.3 shows an example of both the linear and quadratic extrapolation functional forms. From the figure above, it should be noted that the choice of the extrapolation function can affect the SIMEX estimates.

In order to justify the functional form of our SIMEX extrapolation step, we perform a Monte Carlo study to test which extrapolation function has the least bias and smallest mean square error (MSE) later in the article. Based on this Monte Carlo simulation, we choose a quadratic extrapolation function for this model. We also apply the SIMEX algorithm to obtain the estimates of technical efficiency, estimated residuals, and standard errors. To obtain the standard errors, a method known as the SIMEX sandwich estimator is used. We describe this method in the next section.

2.5 SIMEX Sandwich Estimator

Inference for the SIMEX estimators can be performed either via the bootstrap or the theory of M-estimators (Carroll et al. 2012). The bootstrap method can be very computationally intensive, so we chose to apply the M-estimator method. M-estimators are broadly defined as any estimator which minimizes sums of functions of the data. Classic examples of an M-estimator are Ordinary Least Squares estimators (OLS), Maximum Likelihood Estimators. The sandwich variance estimator is also a member of this family of M-estimators. In the OLS framework, the best linear unbiased estimator of β is $\hat{\beta} = (X'X)^{-1}(X'Y)$. The corresponding estimator for the variance is $Var(\hat{\beta}) = (X'X)^{-1}X'(Var(Y))X(X'X)^{-1}$, which resembles a sandwich of the variance of Y around two inverse matrices of X (White 1980).

The sandwich estimator method exploits the fact that $\hat{\Theta}_{SIMEX}$ is asymptotically equivalent to an M-estimator and thus makes use of the sandwich formula to construct the variance-covariance matrix. We explain how SIMEX can be applied to obtain the approximate variance-covariance matrix, as outlined in Carroll et al. (2012).

Applying the results of Cook and Stefanski (1994) it can be shown that the the SIMEX estimator is approximately consistent, given a large-sample and the appropriate extrapolant function. It can further be shown that

$$Var(\hat{\Theta}_{SIMEX}) \approx Var(\hat{\Theta}_{True}) + Var(\hat{\Theta}_{SIMEX} - \hat{\Theta}_{True}) \tag{2.14}$$

This equation decomposes the variance of $\hat{\Theta}_{SIMEX}$ into two components. The first component contains the sampling variability, $Var(\hat{\Theta}_{True}) = \tau^2$. The second component contains the measurement error variability, $Var(\hat{\Theta}_{SIMEX} - \hat{\Theta}_{True})$.

Let $\hat{\tau}_b^2(\lambda)$ be the sandwich estimator for the SIMEX estimator $\hat{\Theta}_b(\lambda)$, and $\hat{\tau}^2(\lambda)$ be the average for $b = 1, \ldots, B$. Then $\hat{\tau}^2(\lambda)$ can be plotted as a function of λ, with an extrapolant model fit to $\lambda = -1$, the true variance τ^2 is obtained asympotically. To estimate the second component of the variance,$Var(\hat{\Theta}_{SIMEX} - \hat{\Theta}_{True})$, we need the difference between each SIMEX estimate and its average for $b = 1, \ldots, B$, at each λ.

$$\Delta_b(\lambda) = \hat{\Theta}_b(\lambda) - \hat{\Theta}(\lambda), b = 1, \ldots, B. \tag{2.15}$$

In addition, we need sample variance-covariance matrix for $[\hat{\Theta}_b(\lambda)]_{b=1}^{B}$.

$$s_\Delta^2(\lambda) = (B-1)^{-1} \sum_{b=1}^{B} \Delta_b(\lambda)\Delta_b^t(\lambda) \tag{2.16}$$

The significance of these two quantities comes from the fact that

$$Var(\hat{\Theta}_{SIMEX} - \hat{\Theta}_{True}) = -\lim_{\lambda \to -1} Var(\hat{\Theta}_b(\lambda) - \hat{\Theta}(\lambda)) \tag{2.17}$$

Since $E[\hat{\Theta}_b(\lambda) - \hat{\Theta}(\lambda)|data] = 0$, it follows that unconditionally $E[s_\Delta^2(\lambda) \; = Var(\hat{\Theta}_{SIMEX}(\lambda) - \hat{\Theta}_{True}(\lambda)]$. So the component we really want to estimate is

$$Var(\hat{\Theta}_{SIMEX} - \hat{\Theta}_{True}) = -\lim_{\lambda \to -1} E[s_\Delta^2(\lambda)] \tag{2.18}$$

Given this result, the estimator of $Var(\hat{\Theta}_{SIMEX})$ is computed as the difference $\hat{\tau}^2_{SIMEX} - s^2_{\Delta}$. For this model we wrote the SIMEX sandwich estimator routine in R, and the components $\hat{\tau}^2_{SIMEX}(\lambda) - s^2_{\Delta}(\lambda)$ are modeled and extrapolated to $\lambda = -1$. There were no issues with the covariance matrix not being positive definite, though Carroll et al. (2012) mentions this can be an issue.

2.6 Results

This section presents results from both the naive and SIMEX estimation of the production parameters, standard errors and technical efficiency. Model estimates are compared and contrasted, and the significance of results are discussed later in this section. Table 2.2 shows SIMEX estimates based on the quadratic extrapolation from 1000 simulations at each λ_m.

	Model 1		Model 2	
Variable	Naive	SIMEX[1]	Naive	SIMEX[1]
Intercept	-6.634***	-20.434***	-9.248***	-22.537***
	(0.956)	(1.910)	(0.923)	(1.173)
$timefish$	0.498***	1.029***	1.044***	1.124***
	(0.086)	(0.095)	(0.022)	(0.020)
$fuel$	0.551***	0.106		
	(0.085)	(0.089)		
$length$	-0.599***	0.047	0.180	0.138
	(0.171)	(0.167)	(0.127)	(0.119)
$biomass$	1.598***	3.344***	1.793***	3.659***
	(0.095)	(0.158)	(0.098)	(0.139)
$\hat{\sigma}^2_b$	0.070	0.036	0.067	0.018
$\hat{\sigma}^2_e$	0.054	0.020	0.064	0.087
note:[1] SIMEX	sandwich standard	errors reported		
*p $<$ 0.1,	**p<0.05,	***p<0.01		

Table 2.2: Model Estimates

Comparing the naive and SIMEX parameter estimates for Model 1 we see distinct differences in the coefficients of $biomass$, $timefish$ and $length$. The coefficients for all three are biased downwards in the naive model. The coefficient on $timefish$ approximately doubles in the SIMEX estimates, suggesting it has a larger impact on harvest than in the naive model. The direction of the biases differ for the intercept and $gallons$, which are biased upwards. A comparison of the standard errors shows that there is a tradeoff in bias versus variance. The standard errors are typically larger in the SIMEX model. However, an important point to consider is that researchers typically only get one sample, so bias in the point estimates may be more of a concern than the variability.

We also estimate a simpler Cobb-Douglas production model, without including the $gallons$ variable. The naive and SIMEX estimates for Model 2 show similar magnitudes and standard errors, except for $biomass$. Again the naive estimate has a much smaller coefficient than the SIMEX estimate. Another interesting difference in the two models is the variability in the estimates of technical inefficiency, σ_b^2. This estimated variance is consistently smaller in the SIMEX estimates, suggesting that the measurement error is inflating the variance of inefficiency term.

Another important finding related to economic theory is the coefficient on $length$ changes signs between the models. It appears negative and significant in the naive model, and changes to positive and not significant in the SIMEX model. Because vessels are owned and operated for many years, there is not a lot of variation in $length$, possibly explaining the lack of significance in the model. The variable is still included in the model because length of the vessel is a typical measure of capital in the fisheries literature. The positive coefficient on $length$ is what would be expected from economic theory. A simpler model with $timefishing$, $length$ and $biomass$ as covariates was also specified, giving almost identical parameter estimates and standard errors.

Another quantity of interest is the returns-to-scale of the industry. This is measured as a linear combination of the coefficients on $timefishing$, $gallons$ and $length$.

Under the naive model a 95% confidence interval for $\beta_1 + \beta_2 + \beta_3$ is $[0.115, 0.785]$. Under the SIMEX corrected model it is $[0.843, 1.521]$. The SIMEX estimates suggests the industry is operating under constant returns-to-scale, while the naive estimates suggest it is operating under decreasing returns-to-scale. A profit maximizing firm should be operating in the decreasing returns-to-scale portion of the production function, but given the nature of industry, it is more likely that firms operate in a cost minimization framework. In the context of the ITQ market, vessels must minimize costs subject to their quota and the contracted deliveries to the surfclam processors. In this scenario it is very possible that the industry could be operating under constant or even increasing returns-to-scale.

2.6.1 Technical Efficiency

Figure 2.4 shows a density plot of the technical efficiency estimates for Model 1. Median technical efficiency is 64% under the naive estimation and 61% under SIMEX. A paired Wilcoxin Signed-Rank Test confirms that these two distributions are significantly different at the 1% level of significance. The test for a difference in medians reveals that the naive estimate is significantly greater than the SIMEX estimates. This result suggests the naive model tends to overstate the mean technical efficiency of the surfclam industry. Fishery managers using the naive model to assess the impact of ITQs on overall technical efficiency would be using biased estimates, and possibly conclude the industry is operating more efficiently that in reality. The distribution of the SIMEX estimates also has considerably less spread than the naive estimates.

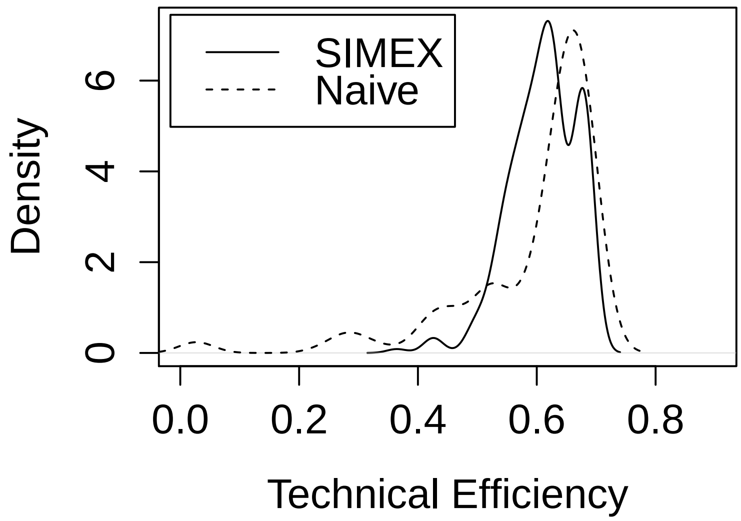

Figure 2.4: Technical Efficiency Density Plot

Next, we look examine factors that may effect vessel-level technical efficiency using data on vessel characteristics such as age of the vessel, hull material and home port state. The results in Figure 2.5, Figure 2.6, and Figure 2.7 show median technical efficiency by vessel age, hull material, and home port state. There are distinct differences in the between the naive model and SIMEX model, demonstrating how measurement error can mask important relationships in the data.

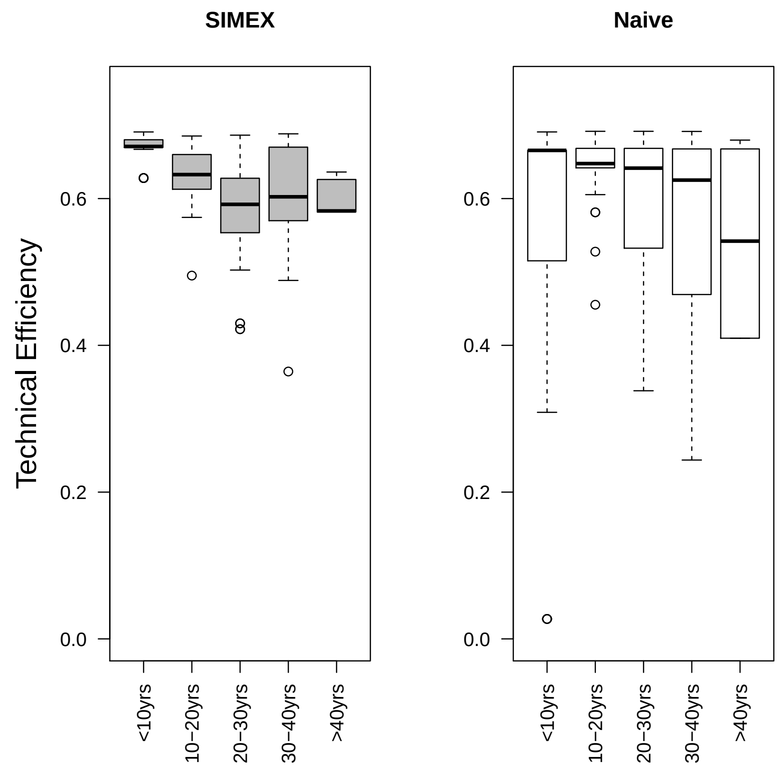

Figure 2.5: Median Technical Efficiency by Vessel Age

The results for the SIMEX model in Figure 2.5 are consistent with economic theory in that older vessels are less technically efficient. A Tukey pairwise comparison of means reveals that vessels less than ten years old are significantly more technically efficient than vessels older than ten years. This result suggests that firms with larger amounts of capital would have an advantage in the fishery. These firms would be able to increase technical efficiency because they can purchase newer vessels, made of more advanced materials. This conclusion is supported by the Figure 2.6 seen below. A Tukey pairwise comparison of means shows that vessels with fiberglass and steel hulls are significantly more technically efficient than vessels with wooden hulls.

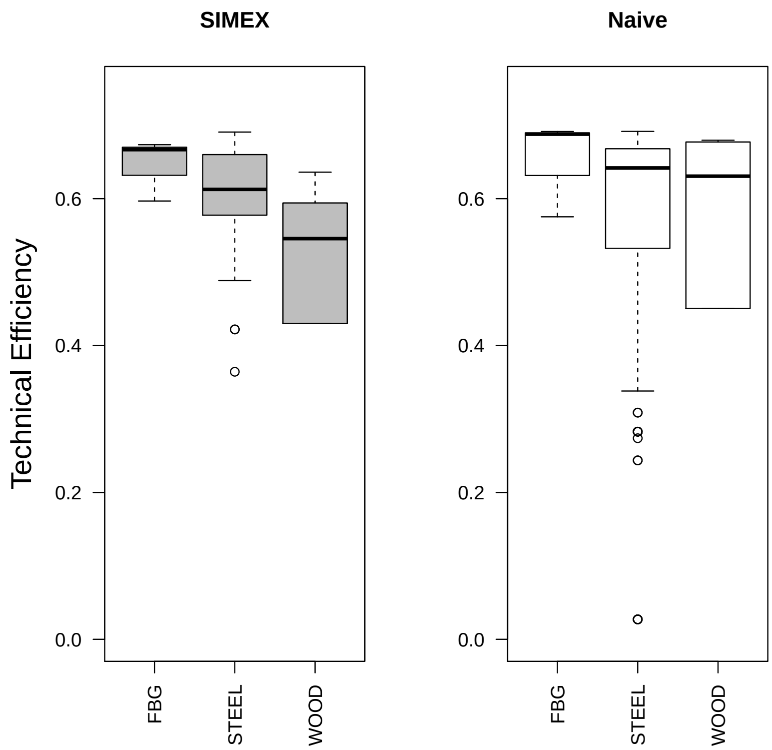

Figure 2.6: Median Technical Efficiency by Hull Material

An additional factor in explaining variation in technical efficiency is the home port of the vessel. Figure 2.7 shows technical efficiency for vessels operating out of ports in New Jersey, Maryland and Rhode Island, and the remaining vessels operating in the northern and southern parts of the fishery. For boxplots of the estimated technical efficiency by vessels in each home port state, see Appendix A.

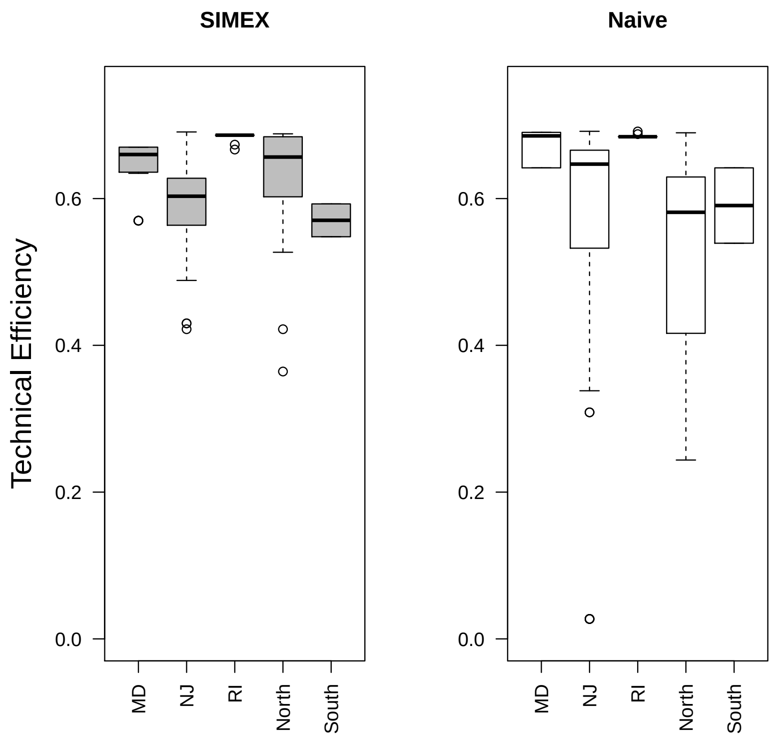

Figure 2.7: Median Technical Efficiency by Principal Port State

Comparing the two sets of estimates we can see distinct differences in how technical efficiency varies by region. Vessels operating near the largest surfclam processors are found in New Jersey, Maryland and Rhode Island. The boxplots above clearly show that many vessels in these states are operating closer to the production frontier. Because processors likely have access to larger amounts of capital they will be able to hire the best captains and crews. Additionally, these results suggest vessels that are vertically integrated with processors are more profitable, and less likely to exit the fishery.

The SIMEX results also show that vessels in ports located in the southern region of the fishery, such as North Carolina and Virginia, have significantly lower technical efficiency than vessels in the northern part, such as Massachusetts and New Hampshire. One possible explanation for this finding is that climate change is having adverse effects on the resource abundance in the southern part of the fishery (McCay et al. 2011, Weinberg 2005). The data confirms that firms the southern part of the fishery are not investing in newer vessels. The results is lower technical efficiency and lower profitability for these firms.

2.7 Monte Carlo Study of SIMEX

In this section we examine the large sample properties of the SIMEX estimator and show that it leads to bias reduction. The purpose of this Monte Carlo study is to show that SIMEX will result in bias reduction for a stochastic frontier model with additive measurement error in a single covariate. First, a data set is generated from known parameters according to equation 2.2, and measurement error is generated using equation 2.11. The simulated data includes nine time periods and 50 vessels, for a total of 450 vessel-year observations. We replicate the data using the same correlation matrix found in the original data set. For each replication of the Monte Carlo experiment, the vessel-level data is generated using equations 2.2 and 2.11, and a naive model is estimated. This process is repeated 5,000 times for this Monte Carlo study.

In each replication the stochastic frontier model is first estimated naively. The SIMEX algorithm is then applied to the simulated data, using the known measurement error variance to add additional amounts of measurement error to data. Finally, both a linear and quadratic extrapolation function were used to create the SIMEX parameter estimates for each replication. We report measures of bias and root-mean-square-error (RMSE) for the naive estimation and both extrapolant function forms. Bias is calculated as $E(\hat{\beta} - \beta)$ and root-mean-square-error (RMSE) $\sqrt{E(\hat{\beta} - \beta)^2}$. These two statistics are displayed in tables 2.3 and 2.4 below.

The gains in bias reduction can be seen clearly in table 2.3. The results show that the quadratic extrapolant function does better in bias and MSE when com-

	Naive	Linear	Quadratic
$\hat{\beta}_0$	0	0	0
$\hat{\beta}_1$	-0.091	-0.070	-0.033
$\hat{\beta}_2$	0.066	0.050	0.024
$\hat{\beta}_3$	-0.149	-0.114	-0.050
$\hat{\beta}_4$	-1.950	-1.500	-0.710

Table 2.3: Bias measures for Monte Carlo Study

	Naive	Linear	Quadratic
$\hat{\beta}_0$	0.095	0.122	0.170
$\hat{\beta}_1$	0.105	0.063	0.084
$\hat{\beta}_2$	0.084	0.077	0.077
$\hat{\beta}_3$	0.184	0.161	0.158
$\hat{\beta}_4$	1.996	1.609	1.242

Table 2.4: RMSE measures for Monte Carlo Study

pared with either the naive model or the linear extrapolant. The variance tradeoff can also be seen in table 2.4 where the SIMEX estimator with a quadratic extrapolation shows smaller RMSE than the naive or linear extrapolation function in β_4,β_3, and β_2, but slightly larger RMSE in β_1 and β_0. The simulation study confirms that if the measurement error model is truly additive, then the SIMEX algorithm with a quadratic extrapolant will give less biased results. The resulting sampling distributions for each parameter in the simulation study can be found in Appendix A.

2.8 Discussion

This paper examines the consequences of measurement error in a stochastic production frontier model using a panel of logbook and biomass survey data from the Mid-Atlantic surfclam fishery. A stochastic frontier model is estimated using a Cobb-Douglas functional form specified by economic theory. Measurement error is assumed to be an additive component for the variable $biomass$, leading to a problem of inconsistent parameter estimates for the stochastic frontier model. The results show that naive estimation of the model leads to inconsistent estimates of the parameters and technical efficiency. The measurement error problem is addressed using a Monte Carlo method called SIMEX. SIMEX is a bias-reducing estimation method that establishes a relationship between the measurement error variance and the estimated parameters in the mixed model. The SIMEX sandwich estimation method is used to get corrected standard errors.

The results show that the estimated parameters are significantly biased, giving fishery managers poor information about the returns-to-scale for the fishery. The results also show that not accounting for the measurement error in the data leads to overstating technical efficiency for the fishery. The SIMEX estimates also show that there are significant relationships between technical efficiency and vessel characteristics. In particular, older vessels are less technically efficient, and vessels with hulls made of wood are less technically efficient when compared to vessels with fiberglass and steel hulls. An important finding is that vessels whose home ports are located in the southern end of the fishery have lower technical efficiency on average. The effect of declining biomass in the southern end of the fishery with the combination of older, less technically efficient vessels suggests firms are not replacing these vessels because fishing is not profitable. Additionally, returns-to-scale estimates are consistent with consolidation in the ITQ market. We show that not accounting for the measurement error in the data would mask these important relationships, leading a researcher to incorrect conclusions about factors affecting technical efficiency.

CHAPTER 3

MEASUREMENT ERROR IN THE SCHAEFER PRODUCTION MODEL

3.1 Introduction

Since the reauthorization of the Magnuson-Stevens Fishery Conservation Act in 2007 (M.S. Act), fishery managers have been given a high standard to meet for determining future management plans. National Standard One of the M.S. Act states that management plans must prevent overfishing while achieving the optimum yield for the fishery. National Standard Two stipulates that new management measures should be based on the best scientific information available. Additionally, National Standard Eight states that fishery managers should take into account the effect that new regulations have on fishing communities and minimize the adverse economic impacts on these communities (U.S. Department of Commerce 2007). In order to achieve the objectives outlined in the M.S. Act, fishery managers must know what constitutes overfishing. This requires information on the biological parameters of fishery, which can be estimated using catch and effort data, obtained from vessel logbooks.

The motivation for this paper comes from the fact that fish stocks are not directly observed, and it is costly to collect biological data. Because biological information may not be available, fishery managers make use of catch and effort data to make inference about the status of the biomass. The Schaefer production model, a widely used model developed by Schaefer (1954), is a bioeconomic model that links an economic model of production to a biological growth model for a fishery using catch and effort data. Fishery managers make use of the data obtained from vessels active in the fishery to estimate the biological parameters. The data contained in logbook records include such variables as, time-at-sea, crew size, gear type, latitude and longitude of fishing, and vessel characteristics. Using these estimated parameters and the logbook data it is possible to estimate the biomass itself. This paper will examine how to mitigate the effects of measurement error, which is brought in through the classical estimation of the Schaefer production model.

The Schaefer production model is a two-stage model. In the first stage production function, harvest is modeled as a function of effort and the biomass. Typically a Cobb-Douglas functional form is assumed, though more recent work has looked at more flexible functional forms (Zhang and Smith 2011). This first stage model cannot be estimated with the available data, since biomass is not observed by the econometrician. Because biomass is not directly observed, a proxy variable is typically used in its place. The most common common proxy variable is catch-per-unit effort (CPUE), which is the ratio of harvest over the time fishing. This proxy variable is then substituted into the second stage of the model. This second stage is a growth model for the biomass, usually assumed to be logistic. The growth model relates the biomass in year t+1 to the biomass in year t, biomass squared in year t, and total catch in year t. Using this growth model, parameters for the intrinsic growth rate (r), carrying capacity (K), catchability coefficient (q) can be estimated. After obtaining these parameters a fishery manager can solve for the unobserved biomass by rearranging the first stage production function. The biomass information can be used to make management decisions about the fishery. The problem with this method is that using the CPUE proxy in the second stage growth model, without correcting for measurement error, will lead to inconsistent estimates of biological parameters in the second stage of the model.

Cochrane (2002) notes that production models, like the Schaefer production model, are used only for single-species stock assessment. Multi-species fisheries are inherently more complex and require additional information. The main advantage of single-species models is they only require information on annual catch and an index of stock abundance, such as CPUE. However, there are several drawbacks to using these models. Because they ignore information about environmental factors, age structure, and populations of other important species, such as predators and prey, these simplifying assumptions mean the dynamics of the biomass depend only on the abundance and the harvest. The tradeoff in simplicity means complex relationships between the biomass and environmental/biological factors are not present in the model. Cochrane (2002) further notes that another limitation of production models is that they require good data contrast in effort and biomass. Better contrast in these two variables will give more precise estimates.

According to Punt (1992), when estimating the biomass from catch and effort data, the method used to fit the model has been shown to be much more important than the actual parametric form. Important questions about the methods used to obtain the biological parameters revolve around the classical regression model assumptions, and the error terms. First, in the first stage equation of the Schaefer production model an observation error is appended when using OLS estimation. Second, the second-stage growth model will also contain error in the estimation, called the process error. Third, the use of a proxy variable such as catch-per-unit effort (CPUE), which is often a misleading measure of the resource abundance, can lead to correlation between the errors and the proxy variable. These three issues represent the greatest challenge in estimating biological parameters with the Schaefer production model.

Polacheck et al. (1993) note that there are three widely used models for fitting a dynamic biomass to observed data; the effort-averaging estimation method (Gulland 1961, Fox 1975), process-error estimators (Walters and Hilborn 1976, Schnute 1977) and observation-error estimators (Pella and Tomlinson 1969, Butterworth and Andrew 1984). Observation error estimators refer to appending error in the first stage production model, while process-error estimators refer to appending error to the second stage growth model. The effort-averaging estimation method refers to averaging effort from the vessel data. Each of these methods has potential drawbacks. Polacheck et al. (1993) examine each of these estimators, concluding that the effort-averaging estimator is biased and process-error estimators have high variability. They further conclude that the observation error method should provide the most precise parameter estimates, but that any method should be assessed with a Monte Carlo study before implementation. Zhang and Smith (2011) use a generalized Schaefer Production Model, which attempts to correct simultaneously for process and observation error. This paper will further improve upon their model by using a Monte Carlo method for reducing bias in the parameter estimates.

As Zhang and Smith (2011) point out, there are three empirical problems with estimating this model: 1) the biological dynamics have natural variation or process error; 2) the production function has stochastic shocks, which makes the inferred stock noisy; and 3) the fishing production function, usually Cobb-Douglas, has an extremely restrictive form. Without taking these three factors into account simultaneously, the resulting estimation of the classic Schaefer production model will lead to biased and inconsistent parameter estimates. This means that estimates of the growth parameters and the biomass estimates are not reliable. Particularly in the second stage of the estimation process, which is non-linear in the parameters, there is great potential for biased results (Uhler 1980).

Measurement error in a covariate of linear model can lead to biased and inconsistent parameter estimates (Carroll et al. 2012). This is particularly important in the Schaefer production model because the measurement error enters into the second stage equation in a non-linear form. Thus, it is possible that small amounts of measurement error in a single covariate could lead to very large biases in the parameters.

This paper contributes to the existing literature on the Schaefer production model by applying a relatively new method for reducing the bias induced by measurement error. This method, known as simulation extrapolation (SIMEX), was first developed by Cook and Stefanski (1994). In keeping with the goal of the M.S. Act, we make use of the "best scientific information available" to improve upon a generalized version of the classic two-stage Schaefer production model recently proposed by Zhang and Smith (2011). This more generalized model is a significant improvement over the classic model, using panel data methods by creating a time-varying stock index to reduce bias in the parameter estimates. One drawback of these methods is they rely on large panel datasets for consistency. This means that the parameters of the model will converge to their true values as the number of cross-sections (e.g. firms) and time periods increase towards infinity (Hsiao 2003). We argue that this method can further be improved using SIMEX (Cook and Stefanski 1994), particularly when a manager has more limited dataset. It could be argued that realistically managers may only have data on a few time periods, and small number of cross sections of catch and effort data. Our data, which comes from the Mid-Atlantic surfclam fishery for 2001-2009, contains about thirty to forty vessels (cross sections) in each time period, and these vessels are observed over nine time periods.

Using logbook data from the Mid-Atlantic surfclam fishery, we estimate the generalized Schaefer Production model. In order to apply the measurement error correction method known as SIMEX, we also need to estimate the measurement error variance. Using the theoretical two-stage panel data method explained in

Zhang and Smith (2011), an estimate of the measurement error variance can be obtained.

In order to assess the bias of the SIMEX generalized Schaefer production model, we use the biomass estimates for the Mid-Atlantic surfclam fishery from the Northeast Fisheries Science Center. Although these are estimates have are known to contian measurement error, we use these as the truth when evaluating the models presented in this study. We then compare the various biomass estimates by looking at measures of bias. This has important policy implications since many fishery managers often base the total allowable catch (TAC) on estimates on limited datasets. If fishery managers make poor decisions about setting the TAC based on biased estimates of the biomass, it is certainly likely that goals set in the reauthorized M.S. Act will not be met (U.S. Department of Commerce 2007).

The rest of the paper will proceed as follows: Section 3.2 explains the methodology of the Schaefer and generalized Schaefer Production Model, Section 3.3 explains a measurement error model and describes the SIMEX estimator, Section 3.4 details the functional form of the generalized Schaefer Model, Section 3.5 describes the Mid-Atlantic surfclam logbook data, Section 3.6 reports results from the naive, generalized Schaefer Production model and SIMEX model, and Section 3.7 summarizes the findings of this paper and discusses future work.

3.2 Methodology

We start by describing the classic Schaefer production model. The first stage is estimation of a production function, followed by the second stage estimation of a logistic growth model for the biomass.

3.2.1 Classic Schaefer Model

The classic Schaefer production model is

$$H_t = qE_tX_t \tag{3.1}$$

where X_t is the biomass in time t, H_t is harvest in time t, and E_t is fishing effort in time t. The state equation or growth model is

$$X_{t+1} = X_t + rX_t(1 - \frac{X_t}{K}) - H_t \tag{3.2}$$

where the intrinsic growth rate r, carrying capacity K, and catchability coefficient q are the biological parameters to be estimated. If biomass, harvest and effort information are all observed, this model can be estimated under regularity conditions (Zhang and Smith 2011). However, the fishery biomass is typically not observed, and is therefore a latent variable to the econometrician.

To estimate the above model when biomass is unknown a proxy is used, such as catch-per-unit-effort (CPUE). Let y_t denote CPUE, defined as $y_t = \dfrac{H_t}{E_t}$. From equation 3.1, CPUE is proportional to the unobserved biomass, such that $X_t = \dfrac{y_t}{q}$. The standard approach to the Schaefer production model substitutes this proxy into equation 3.1 and appends an additive error term. The resulting estimating equation is

$$y_{t+1} = (1 + r)y_t - \frac{r}{qK}y_t^2 - qH_t + \epsilon_t \tag{3.3}$$

Under certain exogeneity conditions, $E(\epsilon_t|y_t) = 0$, this equation can be estimated by least squares. With a distributional assumption placed on the error term, such as $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$, then maximum likelihood estimation can be used. This model is referred to as the CPUE estimator.

There are several drawbacks to this particular estimator. First, the production function assumes a very restrictive form, with constant returns to scale imposed. Second, the production function is usually assumed to have an error term that captures variability in the harvest from random events, such as weather. Thus when y_t is substituted in equation 3.3, the right-hand side of the equation has y_t and y_t^2, both of which are measured with error.

Another restrictive assumption is the functional form of the growth model in equation 3.3. This assumes a logistic growth model for the fishery. Many other functional forms for the growth model could be fitted to the data, and we note that this is not explored in this study.

3.2.2 CPUE-like estimator

A more flexible model, known as the CPUE-like estimator (Zhang and Smith 2011), allows for a more flexible production function. The new harvest function is

$$H_t = qE_t^\alpha X_t^\gamma \tag{3.4}$$

where α represents the input elasticity of effort, and γ represents the elasticity of the stock. This more flexible form allows the production function to have decreasing, constant, or increasing returns-to-scale, depending on whether α is greater than, equal, or less than one. Rearranging equation 3.4, the unobserved stock is now estimated by $X_t = (\frac{H_t}{qE_t^\alpha})^{1/\gamma}$. Using this as a proxy for the stock, the second stage state equation becomes

$$(\frac{H_{t+1}}{qE_{t+1}^\alpha})^{1/\gamma} = (1+r)(\frac{H_t}{qE_t^\alpha})^{1/\gamma} - \frac{r}{K}(\frac{H_t}{qE_t^\alpha})^{2/\gamma} - H_t + \epsilon_t \tag{3.5}$$

While this more flexible model deals with the restrictive homogeneity assumptions of the original Schaefer production model, it does not address the issue of measurement error in the second stage of the equation. Uhler (1980) points out that if the stock is measured with significant noise due to the error in the production function, then the more flexible CPUE proxy for stock will lead to biased estimators.

3.2.3 Generalized Schaefer Production Model

To deal with the measurement error issue Zhang and Smith (2011) propose a two-stage panel data estimator that incorporates more information from the log-book data, which they call the generalized Schaefer production model. A more generalized production function is specified, accounting for the fishing vessels (i), gear deployed (g), visited areas (j) and periods over time (t). The generalized Schaefer production function is

$$H_{ijgt} = q_{ijgt} E_{ijgt}^{\alpha_g} X_t^{\gamma} exp(\epsilon_{ijgt}) \tag{3.6}$$

The stock is assumed to be homogeneous across space for the sake of simplification of the model. This model allows for the catchability coefficient q_{ijgt} to vary by vessel, gear, area and time. This means $q_{ijgt} = exp(\phi_g + a_j + \psi_{ijgt})$, where ϕ_g is a gear-specific constant and a_j is an area specific constant, and ψ_{ijgt} is a random error term to capture unobserved heterogeneity. This more flexible production function also allows the catch-effort elasticity α_g to vary by gear type. The error term, ϵ_{ijgt}, captures unobserved shocks in harvesting, such as weather.

Taking the natural log of both sides of equation 3.6 the production model becomes a linear model. Letting lowercase letters represent the natural log of the uppercase variables, i.e. for the biomass, x_t = $log(X_t)$, the new generalized production model is

$$h_{ijgt} = \phi_g + a_j + \alpha_g e_{ijgt} + \gamma x_t + \eta_{ijgt} \tag{3.7}$$

where $\eta_{ijgt} = \psi_{ijgt} + \epsilon_{ijgt}$. Zhang and Smith (2011) show that if the period is short enough, the unobserved stock x_t can be assumed to be constant in each time period t, and treated as a fixed-effect. To avoid perfect multicollinearity (i.e. dummy variable trap), let $q = exp(\phi_1 + a_1)$. The advantage to using this fixed-effects method is that the stock or biomass can be canceled out through demeaning among fishing trips in the same period. This model will be estimated in the first stage of the generalized Schaefer model. Later in this study we specify a slightly different production function, based on the Mid-Atlantic surfclam data that is available.

From equation 3.7 above, the stock index is defined as

$$c_t = \gamma x_t + lnq \tag{3.8}$$

The stock index is essentially a time-varying fixed-effect. It is a linear function of the stock, with slope γ and intercept lnq. As the underlying stock changes, so too will the stock index. Rewriting equation 3.7 so that z_{ijgt} denotes the vector of variables except the stock index (i.e. gear, area, and vessel constants, plus effort), and β a vector of parameters, the new production function can be written

$$h_{ijgt} = z_{ijgt}^{'}\beta + c_t + \eta_{ijgt} \tag{3.9}$$

This form highlights the panel data methodology of this model. The advantage of this model is that under regularity conditions, it can be estimated consistently via a within-period estimator. Using the asymptotic properties of panel data this estimator can be shown to be consistent. Let n be the number of cross-sections and T be the number of periods, then if n or $T \to \infty$, than this model can be consistently estimated (Hsiao 2003).

All parameters in equation 3.9 are identified, except for the stock (biomass) index, c_t. This can be estimated as a time-varying fixed effect

$$\hat{c}_t = \bar{h}_t - \bar{z}_t^{'}\hat{\beta} \tag{3.10}$$

where $\bar{h}_t = \sum_{i,j,g} h_{ijgt}$ and $\bar{z}_t = \sum_{i,j,g} z_{ijgt}$. From this equation it is apparent that $\hat{c}_t$ is unbiased if the production function is specified correctly, but is only consistent as $n_t \to \infty$. If the data has a large number of cross sections (i.e. a large number of vessels in each time period), then this requirement will be satisfied. If not, then $\hat{c}_t$ will contain significant amounts of noise, which can lead to bias in the second-stage estimation.

Combining equations 3.8 and 3.10, the stock can be expressed as a function of $\hat{c}_t$ if parameters γ and q are known. Let $\hat{Y}_t = exp(\hat{c}_t) = exp(\bar{h}_t - \bar{z}_t'\hat{\beta})$. Then the stock can be estimated as

$$\hat{X}_t = \left[\frac{exp(\bar{h}_t - \bar{z}_t'\hat{\beta})}{q}\right]^{1/\gamma} = \left(\frac{\hat{Y}_t}{q}\right)^{1/\gamma} \tag{3.11}$$

In order to estimate the stock, we require parameter estimates for γ and q. The second-stage estimation uses a logistic growth model to estimate these parameters.

Substituting $\hat{X}_t$ for X_t into the logistic growth model, the second stage equation is written as

$$\left(\frac{\hat{Y}_{t+1}}{q}\right)^{1/\gamma} = (1+r)\left(\frac{\hat{Y}_t}{q}\right)^{1/\gamma} - \frac{r}{K}\left(\frac{\hat{Y}_t}{q}\right)^{2/\gamma} - H_t + \epsilon_t \tag{3.12}$$

This equation is nonlinear in the parameters and can be estimated through maximum likelihood or nonlinear least squares. Because of high multicollinearity in the variables, to be identify the model, Zhang and Smith (2011) propose restricting the catch-stock elasticity γ to one. This dramatically simplifies that estimation procedure by making the model linear in the parameters. Allowing the catch-stock elasticity to equal one assumes that a 1% increase in the stock causes a 1% increase in the harvest, all else equal. If this assumption does not hold, then the estimated biological parameters and biomass will be biased. The alternative is to estimate the model using non-linear least squares or some another numerical method. We proceed estimation of the model with the assumption of $\gamma = 1$.

Let $s = \frac{r}{qK}$ and $\Delta\hat{Y}_{t+1} = \hat{Y}_{t+1} - \hat{Y}_t$, then equation 3.12 can be written

$$\Delta\hat{Y}_{t+1} = r\hat{Y}_t - s\hat{Y}_t^2 - qH_t + \epsilon_t^* \tag{3.13}$$

where $\epsilon_t^* = q\epsilon_t$. This form for the growth model allows us to estimate the biological parameters of interest, r,s and q. Assuming the classical regression model assumptions hold, this model can be estimated using feasible generalized least squares (FGLS) or maximum likelihood (MLE). The generalized Schaefer production model uses this form to estimate the growth model. However, there are still unresolved issues with estimation of equation 3.13. We turn our attention to unresolved issues with measurement error in the stock index, and adapt the generalized Schaefer production model using SIMEX.

The measurement error in equation 3.13 comes from the estimation of $\hat{Y}_{t+1}$, $\hat{Y}_t$, and $\hat{Y}_t^2$, which are a function of the estimated stock index, $\hat{c}_t$. The remaining error in the estimated stock index will be correlated with the disturbance in the model, which violates an important assumption of the classical regression model. The consequences are that the estimated biological parameters in the second-stage growth model will be biased an inconsistent. We propose another way to correct for this measurement error, using a Monte Carlo method known as Simulation Extrapolation (SIMEX).

3.3 Measurement Error Model

In this section we specify a measurement error model for the estimated stock index, $\hat{c}_t$. With variance for $\hat{c}_t$ known, or at least approximately known, it will then be possible to reduce the bias by implementing SIMEX. The error is assumed to be additive, with the estimated stock index equal to the true index plus random error. The measurement error for the stock index is specified as

$$\hat{c}_t - c_t = u_t = \bar{z}_t{}'\beta - \bar{z}_t{}'\hat{\beta} + \bar{\eta}_t \tag{3.14}$$

where $u_t|\bar{z}_t \sim N(0, \sigma^2_{ut})$. The measurement error model requires $E[u_t|\bar{z}_t] = 0$ but does not require the error to be independent of the variables conditioned on. The assumption $u_t \perp\!\!\!\perp \bar{z}_t$, is a much stronger assumption than required here.

From the measurement error model above, we can write the measurement error variance as

$$\hat{\sigma}^2_{ut} = \bar{z}_t{}' \left[\frac{\hat{Cov}(\hat{\beta})}{n_t} \right] \bar{z}_t + \frac{\hat{\sigma}^2_\eta}{n_t} \tag{3.15}$$

where $\hat{Cov}(\beta)$ is the variance-covariance matrix for the estimated parameters in the first stage production function. Equation 3.15 defines the measurement error variance for the stock index. The variance is allowed to vary over time. We note that because of the off-diagonal elements in the variance-covariance matrix, the measurement error is not independent between years. We also allow the measurement error variance to be heteroscedastic over the nine years of our observed data. Given this estimated of the measurement error variance, we can apply SIMEX as a method to reduce bias in the parameter estimates. Before addressing the SIMEX algorithm, we first describe the measurement error problem in the second stage growth model in more detail.

3.3.1 Linear Models with Nonadditive Measurement Error

In this section we discuss the implications of nonadditive error in the estimation of the generalized Schaefer Production Model. We start by examining how the second stage growth model is estimated. This model estimates the biological parameters using the three covariates; $\hat{Y}_t = exp(\hat{c}_t)$, $\hat{Y}_t^2 = exp(\hat{c}_t)^2$, and H_t, where $\hat{c}_t$ is the stock index in time period t and H_t is the harvest or catch in time period t. $\hat{c}_t$ has approximately known additive measurement error. While the measurement error for $\hat{c}_t$ is additive, the model is estimated by using $\hat{Y}_t$, $\hat{Y}_t^2$ and H_t. Since the first two covariates are nonlinear function of the mismeasured $\hat{c}_t$, the second stage growth model parameter estimates will not be unbiased.

We motivate the measurement error problem by describing the properties of a naive estimator for the model. Let $\tilde{\mathbf{x}}_t$ be a (1 x 3) vector containing the true variables [c_t c_t H_t]. Then let $\tilde{\mathbf{w}}_t$ be a (1 x 3) vector containing the estimated or mismeasured variables [$\hat{c}_t$ $\hat{c}_t$ H_t], such that

$$\tilde{\mathbf{w}}_t = \tilde{\mathbf{x}}_t + \tilde{\mathbf{u}}_t \tag{3.16}$$

Defining the vector $\tilde{\mathbf{u}}_\mathbf{t}$ as a (1 x 3) vector, $\tilde{\mathbf{u}}_t$=[u_t u_t 0] , then $E[\tilde{\mathbf{u}}_t|\tilde{\mathbf{x}}] = 0$. However, because $\tilde{\mathbf{x}}_t$ enter into the model through a nonlinear function, call it $g(\tilde{\mathbf{x}}_t)$, this means $E[g(\tilde{\mathbf{w}}_t)] \neq g(\tilde{\mathbf{x}}_t)$.

To show how this will impact estimation of the model, we specify the design matrix by stacking the vector of $\tilde{\mathbf{w}}_t$ for all nine time periods, $t = 1, \ldots, 9$

$$\tilde{\mathbf{W}} = \begin{pmatrix} \hat{Y_1} & \hat{Y_1}^2 & H_1 \\ \vdots & \vdots & \vdots \\ \hat{Y_9} & \hat{Y_9}^2 & H_9 \end{pmatrix} \tag{3.17}$$

In the traditional setup of the generalized Schaefer production model, the second stage growth model is specified as linear in the parameters. It is also assumed to follow the classical regression model assumptions. Thus, the biological parameters can be estimated using ordinary least squares (OLS). Let $\mathbf{y}$ be the response vector of the difference $\hat{Y}_{t+1} - \hat{Y}_t$ for $t = 1, \ldots, 9$, such that

$$\mathbf{y} = \begin{pmatrix} \mathbf{y_1} \\ \vdots \\ \mathbf{y_9} \end{pmatrix} \tag{3.18}$$

The naive growth model can be written

$$\mathbf{y} = \tilde{\mathbf{W}}\beta + \epsilon \tag{3.19}$$

Then the OLS estimator for the naive growth model parameters is

$$\hat{\beta} = (\tilde{\mathbf{W}}'\tilde{\mathbf{W}})^{-1}(\tilde{\mathbf{W}}'\mathbf{y}) \tag{3.20}$$

The classical regression model assumptions define this estimator to be unbiased and consistent when $\tilde{\mathbf{W}}$ is not measured with error. Using our naive estimator above, the asymptotic properties can be evaluated by making a few more assumptions. Following, Greene (2003) let

$$\text{plim}(\frac{\tilde{\mathbf{W}}'\tilde{\mathbf{W}}}{n}) = \mathbf{Q}^* + \mathbf{\Sigma}_u \tag{3.21}$$

where $\mathbf{Q}^* = \text{plim}\,\frac{\tilde{\mathbf{x}}'\tilde{\mathbf{x}}}{n}$, and $\mathbf{\Sigma}_u$ is the Variance-Covariance matrix of $\tilde{\mathbf{u}}$, i.e. the measurement error variance-covariance matrix. The probability limit of $\hat{\beta}$ is then

$$[\mathbf{Q}^* + \mathbf{\Sigma}_u]^{-1}\mathbf{Q}^*\beta \neq \beta \tag{3.22}$$

Thus, the estimator is not consistent for the true parameters. The result of the non-additive measurement error in the linear model is that the estimates of biomass, which are a function of the biased parameter estimates, will be biased and inconsistent. In a situation where measurement error is present, identification of the parameters in the model is often an associated issue. One option is to bring in outside information to help identify the model (Greene 2003). We accomplish this by estimating the measurement error variance, and then using this information to identify the parameters. In the next section we show how the estimated measurement error variance to reduce bias in the naive estimates using SIMEX.

3.3.2 SIMEX Parameter Estimates

Since we can estimate the variance for $\hat{c}_t$ it is possible to apply SIMEX and attempt to reduce bias in the parameter estimates. SIMEX is a two-step simulation-based method of estimating and reducing bias due to measurement error. First, simulated data are obtained by adding additional measurement error to the data in a resampling-like process, establishing a trend of measurement error-induced bias versus the variance of the added measurement error. After that, the extrapolation step follows the fitted trend line back to a point where the measurement error variance is zero. The key underlying SIMEX is the fact that the effect of measurement error on an estimator can be determined experimentally through simulation (Carroll et al. 2012). It can be shown that under a number of different measurement error specifications that SIMEX provides approximately consistent parameter estimates. SIMEX is very general in the sense that the bias due to measurement error in almost any estimator of almost any parameter can be estimated and corrected, at least approximately. SIMEX is described below for the case of additive measurement error in the predictor in four steps, as explained in Buonaccorsi (2010).

Using notation from Section 3.3.1, assume an additive error in one predictor $\tilde{w}_t = \tilde{x}_t + \tilde{u}_t$. This has a measurement error variance $Var(\tilde{u}_t) = \sigma^2_{ut}$. Begin by defining $\theta_j(\lambda)$ as the expected (or limiting) value of the naive estimator of θ_j if $Var(\tilde{u}_t) = (1+\lambda)\sigma^2_{ut}$. Then true value of the j^{th} coefficient is $\theta_j = \theta_j(-1)$.

1. For each λ_m, where $m = 1, \ldots, M$ and λ represents the additional measurement error, generate: $\tilde{w}_{tb}(\lambda_m) = \tilde{w}_t + \lambda_m^{1/2} U_{bt}$ for $b = 1, \ldots, B$, where B is a large number and the U_{bt} are independent generated errors with mean 0 and variance $\hat{\sigma}^2_{ut}$. Because the parameter estimates in the generalized Schaefer production model are small in magnitude, precision is very important. When multiple covariates are measured with error the number of simulated data sets, B, will generally need to be larger to achieve acceptable levels of

Monte Carlo precision Carroll et al. (2012). This is because the Monte Carlo averaging in the simulation step is analogous to numerical integration.

Since $\tilde{w}_t|\tilde{x}_t$ already has variance σ^2_{ut}, the generated $\tilde{w}_{tb}$ would have exactly the variance $(1+\lambda_m)\hat{\sigma}^2_{ut}$ assuming $\hat{\sigma}^2_{ut} = \sigma^2_{ut}$. In practice we usually only have an estimate of σ^2_{ut}.

2. Find $\theta(\lambda_m, b)$, which is the naive parameter estimate for θ_j based on the simulated data. Then define: $\bar{\theta}(\lambda_m)$= $\sum_b \hat{\theta}(\lambda_m, b)/B$. This is the average of the B estimated $\hat{\theta}_j$'s at a particular λ_m .

3. For each $j^t h$ coefficient, fit a model $g_j(\lambda)$ for $\bar{\theta}_j$ (λ_m). Effectively, we regress the vector of $\bar{\theta}_j$ on the vector λ_m. This allows us to establish a trend between greater amounts of measurement error variance and the naive estimator.

4. Get the SIMEX estimate of θ_j using: $\hat{\theta}_j(j, SIMEX) = g_j(-1)$. Because the variance of $\tilde{w}_{tb}$ is exactly $(1+\lambda_m)\hat{\sigma}^2_{ut}$, at the point where $\lambda = -1$ the measurement error variance collapse to zero. This gives an approximately consistent estimate of the true parameter θ_j.

The last step of the SIMEX method is the extrapolation step. There are several functional forms which can be chosen, including the linear, quadratic and rational extrapolant functions. Figure 3.1 shows an example of both the linear and quadratic extrapolation functional forms. We choose a quadratic extrapolation function for this model, reporting results for both the naive estimates and quadratic extrapolation function. It should be noted that the choice of the extrapolation function will affect the SIMEX estimates.

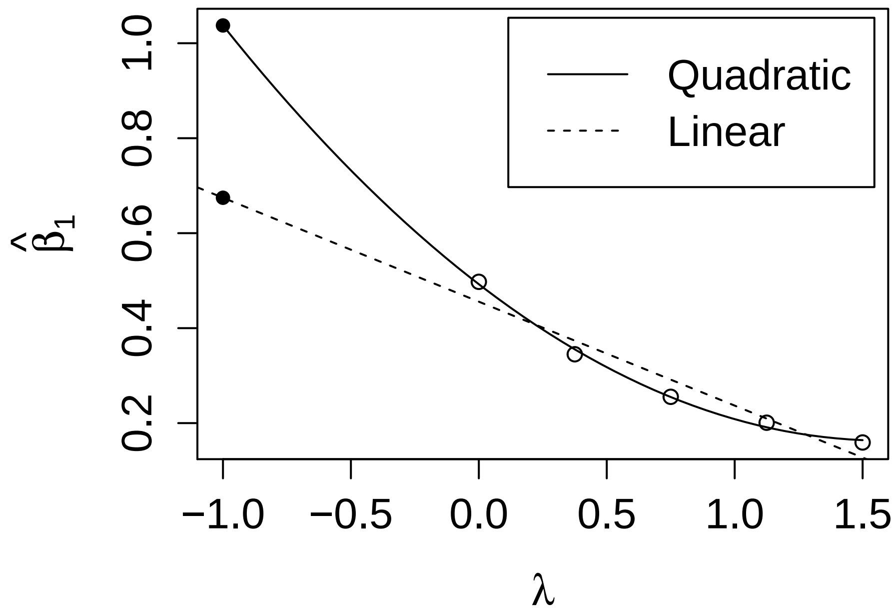

Figure 3.1: An example of the SIMEX extrapolation step

3.3.3 Standard Errors of SIMEX Estimates

Standard errors for the SIMEX parameter estimates are obtained using the two-stage bootstrap. The two-stage bootstrap is different from a one-stage bootstrap because it generates both a response from a regression model, and the mismeasured covariates, similar to a parametric bootstrap procedure. One advantage of the two-stage bootstrap is that it gives an estimate of bias. We report both standard errors, and measures of bias obtained from the two-stage bootstrap later in the results section.

Below we describe the steps for the two-stage bootstrap from (Buonaccorsi 2010). This procedure is used to obtain the standard errors for the estimated parameters in the second stage growth model. Two of the covariates in the model, $\hat{Y}_t$ and $\hat{Y}_t^{\,2}$, are measured with error. In the two-stage bootstrap we generate these variables in b repeated simulations and denote them, w_{bt1} and w_{bt2} respectively. The third covariate is the harvest, denoted x_{bt3}, is assumed not measured with error. For the b^{th} bootstrap sample, generate $[y_{bt}, w_{bt1}, w_{bt2}, x_{bt3}]$, where

$$y_{bt} = \beta_1 w_{bt1} + \beta_2 w_{bt2} + \beta_3 x_{bt3} + e_{bt} \tag{3.23}$$

and the two mismeasured covariates are

$$w_{bt1} = exp(\hat{c}_{t1} + u_{bt}) \tag{3.24}$$

and

$$w_{bt2} = (w_{bt1})^2 \tag{3.25}$$

where $u_{bt}|w_{bt1}, w_{bt2}, x_{bt3} \overset{\text{ind.}}{\sim} N\ (0, \sigma^2_{u_t})$ and $e_{bt} \overset{\text{i.i.d.}}{\sim} N\ (0, \sigma^2_{\epsilon})$.

The standard errors reported for the SIMEX estimates are from 5000 simulations using this two-step data generating process. Estimates of bias are calculated as

$$Bias(boot) = \frac{\sum_{b=1}^{B} \hat{\beta}_b}{B} - \hat{\beta} \tag{3.26}$$

The next section describes how we apply the generalized Schaefer Production Model to the panel data from the Mid-Atlantic surfclam fishery.

3.4 First Stage Production Function for Mid-Atlantic Surfclam

The first-stage is concerned with the estimation of the Cobb-Douglas production function. For the first-stage, we estimate a production function with a random intercept for each vessel. A time-varying stock index is created by assigning a binary variable $x_t = 1$ if observation comes from time period t, and 0 otherwise. A stock index (δ_t) is then estimated as a time-varying fixed effect in the production function. Because a random intercept is specified for each vessel, the model contains fixed and random effects. The variables for the model are defined as

- h_{it} = log_e(total bushels harvested by vessel i in year t),
- z_{it1} = log_e(total hours fished by vessel i in year t),
- z_{it2} = log_e(length of vessel i in year t),
- x_t = a dummy variable, equal to 1 for observations in year t, and 0 otherwise

Let $i = 1, \ldots, 70$ denote vessel and $t = 1, \ldots, t_{n_i}$ denote the n_i years in which vessel i is observed. The production function is specified as

$$h_{it}|b_i, e_{it} = \beta_1 z_{it1} + \beta_2 z_{it2} + \sum_{t=1}^{9} \delta_t x_t + b_i + e_{it} \tag{3.27}$$

with $b_i \overset{\text{ind.}}{\sim} N(0, \sigma_b^2)$ and $e_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_e^2)$, $i = 1, \ldots, 70$, $t = 1, \ldots, t_{n_i}$. This model can be estimated using Restricted Maximum Likelihood Estimation (REML) in R[1]. The estimated $\hat{\delta}_t$ in the model are the stock index for time t. Once the $\hat{\delta}_t$ are estimated, we can calculate $\hat{Y}_t = exp(\hat{\delta}_t)$, and it will then be possible estimate the growth model as shown in equation 3.13.

We note a couple of differences between the production function form for the Mid-Atlantic surfclam fishery logbook data and the functional form used in (Zhang and Smith 2011). Because the harvesting technology for surfclams is the same across all vessels in the fishery, we do not have gear-specific binary variables in the model. Additionally, because a majority of the harvesting takes place off the northern coast of New Jersey, area-specific binary variables are not appropriate. We also use data aggregated to vessel-year, instead of trip-level data, to reduce the computational burden.

---

[1]Estimation of the model is performed using the linear mixed effects models package ”nlme” (Pinheiro et al. 2013) in the R statistical software (R Core Team 2013). The estimates for σ_b^2 and σ_e^2 are found using Restricted Maximum Likelihood or REML. The fixed effects, $\beta's$, are computed using maximum likelihood under the assumption of normality. REML is a form of maximum likelihood estimation that uses a transformed version of the data so that nuisance parameters have no effect on the estimates. It has been shown to provide less biased estimates of the variance-covariance parameters than maximum likelihood McCulloch and Searle (2000).

3.5 Data

The data for the empirical analysis come from the National Marine Fishery Service logbook reporting system, which documents every harvesting trip taken by every vessel in the Mid-Atlantic surfclam fishery in the U.S. EEZ (3-200miles offshore). The logbook data are a panel data set containing approximately 24,000 vessel-trip observations, for years 2001-2009. The trip-level data set includes variables such as bushels harvested, time fishing, time-at-sea, and vessel characteristics such as vessel length, gross-tons and horsepower. There are a total of eighty-eight different vessels observed over the nine year period.

To simplify the correlation structure within each vessel and because biomass is observed annually, data are aggregated by vessel-year. The new data set has one observation for each vessel in a year. One trade off of using aggregated data is that trip-level variability is not observed. The resulting data are reduced to 70 vessels and 285 vessel-year observations. Summary statistics for the data are presented in Table 3.1.

	Obs	Mean	Std.Dev	Min	Max
Harvest (bushels)	285	93749	82007.6	864	442496
Time Fishing (hours)	285	1209.2	951.9	58	3959.4
Fuel (gallons)	285	65896	65979.3	876	388204
Length (feet)	70	85.7	18.4	28	162
Biomass (1000 metric tons)	9	1037	171.9	750	1294

Table 3.1: Summary statistics 2001-2009

A closer inspection of the data reveals that CPUE, seen in Figure 3.2, has been declining in the fishery over the nine year period. This downward trend reflects declining resource abundance due to repeated harvesting from a small area of the fishery, located off the northern New Jersey coast. As a result of this fishing behavior, harvesters are spending more and more time-at-sea.

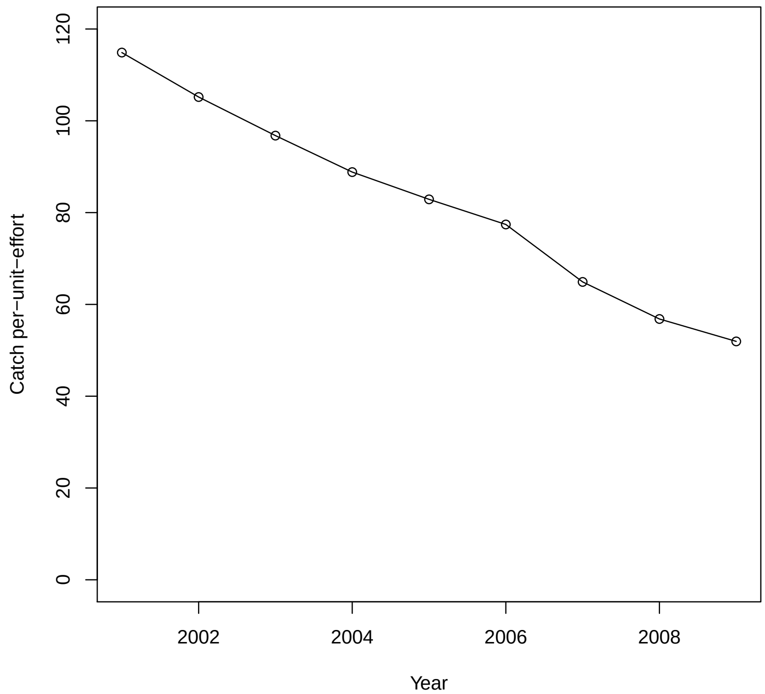

Figure 3.2: CPUE for Mid-Atlantic Surfclam Fishery (2001-2009)

The resulting scenario in which effort is increasing while stock abundance decreases can potentially affect the ability of the generalized Schaefer model to estimate the biomass with any precision. Hilborn (1979) notes that this ”one-way trip” of increasing effort and decreasing index of abundance can sometimes lead to uninformative results. Unfortunately, there are many regulated fisheries which share this fate. We proceed with estimation in with the knowledge that the biases in the estimates may be specific to the Mid-Atlantic surfclam fishery.

3.6 Results

In this section we present and discuss the results from three different models; the classic Schaefer Model, the generalized Schaefer production model, and the adapted Schaefer production model with SIMEX. We compare and contrast the various estimates of the biomass from each model, using the scientific estimates from the Northeast Fisheries Science Center (NEFSC) as estimates of the truth. To estimates the biomass, we convert surfclam bushels to pounds of surfclams using the standard conversion, 1 bushel = 17 pounds (Northeast Fisheries Science Center 2010). Residual plots for each of the estimated models can be found in Appendix B.

3.6.1 Classic Schaefer Production Model

In table 3.2 we present estimation results for the second-stage growth model from classic Schaefer Production Model. This model uses CPUE as a proxy for the stock. In the table below r represents intrinsic growth rate, K is carrying capacity and q is the catchability coefficient.

Variable	Estimate
r	0.532
	(0.379)
r/qK	-5.4E-04
	(0.002)
q	2.72E-06
	(5.1E-06)
$\hat{\sigma}_e^2$	7.416
*p < 0.1,	**p <0.05, ***p<0.01

Table 3.2: Classic Schaefer Model Estimates for Growth Equation

Using the CPUE estimator, we can estimate the biomass as $\hat{X}_t = \frac{\hat{y}_t}{\hat{q}}$. Biomass estimates from classic Schaefer Production Model are shown below in Table 3.3. The NEFSC estimates are treated as the truth in the table below. All biomass estimates are reported in thousands of metric tons. Variance and standard error estimates for the biomass are obtained using a bootstrap method. Confidence intervals are reported as 95% approximate Wald intervals.

Year	NEFSC est.	Sch est.	95% lower	95% upper	% Bias
2001	1294	520	312	728	-60%
2002	1207	476	290	662	-61%
2003	1128	438	265	611	-61%
2004	1104	402	239	565	-64%
2005	1079	375	209	541	-65%
2006	1013	350	207	493	-65%
2007	912	294	177	411	-68%
2008	827	257	152	362	-69%
2009	750	235	140	330	-69%

Table 3.3: Classic Schaefer Biomass Estimates (1000MT)

Using the NEFSC estimates as the true value, the average bias for the classic Schaefer Model is -65%. Clearly the model does a poor job of estimating the biomass for the fishery, with significant downward bias in each year. Although a fishery manager might not see this as a problem, given that estimates of MSY would be conservative, we cannot be sure this would be the case in another fishery. As Uhler (1980) points out, the biases in the model are too complex to estimate beforehand, and so a different fishery could certainly have significant bias in the opposite direction. It is certainly possible this result is limited to the Mid-Atlantic surfclam fishery. At the very least, these estimates of the biomass are not very informative because they are biased downward by a substantial amount.

3.6.2 Generalized Schaefer Production Model

Next we apply Zhang and Smith (2011)'s two-stage panel data estimator to the logbook data from the Mid-Atlantic surfclam fishery. Table 3.4 presents results from first stage estimation of generalized Schaefer production model.

Variable	Estimate
log(timefish)	1.051
	(0.022)***
log(length)	0.149
	(0.132)
Year 2001	3.770***
	(0.593)
Year 2002	3.602***
	(0.593)
Year 2003	3.480***
	(0.595)
Year 2004	3.388***
	(0.597)
Year 2005	3.283***
	(0.580)
Year 2006	3.155***
	(0.599)
Year 2007	2.965***
	(0.597)
Year 2008	2.883***
	(0.597)
Year 2009	2.801***
	(0.596)
$\hat{\sigma}_b^2$	0.070
$\hat{\sigma}_e^2$	0.061
*p < 0.05,	**p <0.01, ***p<0.001

Table 3.4: Generalized Schaefer Model Estimates

Using the results from table 3.4, we can estimate the second stage growth model, specified in equation 3.13. Table 3.5 presents results from second stage estimation of the growth model for the generalized Schaefer Production Model. Just as in the classic Schaefer Production Model, the parameter estimates are not significant due to a high degree of collinearity among the right-hand side variables.

Variable	Estimate
r	-.181
	(0.318)
r/qK	1.38E-04
	(0.005)
q	5.69E-07
	(1.46E-06)
$\hat{\sigma}_e^2$	0.940
*p < 0.1,	**p <0.05, ***p<0.01

Table 3.5: Generalized Schaefer Model Estimates for Growth Equation

Given the results in the table 3.5 the parameter estimate for q is used to estimate the biomass for the fishery $\hat{X}_t$, where $\hat{X}_t = \frac{\hat{y}_t}{\hat{q}}$. Table 3.6 below shows estimates of biomass using the Generalized Schaefer Model. All estimates are reported in thousands of metric tons (1000 MT). Variance estimates for the biomass were obtained using the bootstrap method. The average bias for the Generalized Schaefer Model is - 41%. While this is much smaller than the classic Schaefer Production Model, it does consistently underestimate the biomass in each year except 2001. The 95% confidence intervals for the generalized Schaefer model contain the NEFSC biomass estimate for 2001 only. If we are to believe that the NEFSC biomass estimates represent an unbiased picture of the fishery stock, then this would seem to be a significant problem for fishery managers. A fishery manager using this method with this particular data would likely set the total allowable catch too low or simply not be able to use this model for estimation of the biomass.

Year	NEFSC	Gen. Schaefer	95% Lower	95% Upper	% Bias
2001	1294	996	671	1321	-23 %
2002	1207	842	564	1120	-30 %
2003	1128	745	504	986	-34 %
2004	1104	681	461	901	-38 %
2005	1079	612	419	804	-43 %
2006	1013	539	366	712	-47 %
2007	912	446	300	592	-51 %
2008	827	411	278	544	-50 %
2009	750	368	242	494	-50 %

Table 3.6: Generalized Schaefer Model Biomass Estimates (1000MT)

The results from table 3.6 show that the confidence intervals from the generalized Schaefer Model do not contain the true biomass in any year. Next we perform a two-stage bootstrap for the standard errors, are able to determine the amount and direction of bias for all three parameter estimates in the growth equation. Results are presented in table 3.7. The bias in the parameter q is important, because it is used to estimate the biomass. Because the parameter estimate is smaller than the true value, the resulting biomass is too large.

Variable	Bias	% Bias
r	0.008	4.874
r/qK	4.650E-05	2.633
q	4.824E-07	-17.140

Table 3.7: Bias Estimates for Biological Parameters

3.6.3 Generalized Schaefer Production Model with SIMEX

Table 3.8 below shows estimates from the second stage growth model using the SIMEX with a linear extrapolation functional form. All SIMEX estimates are 10,000 simulations at each λ, resulting in a total of 40,001 simulated datasets. Standard errors reported are from using a two-stage bootstrap method. Again, it can be seen that none of the parameter estimates are statistically significant, due to high multicollinearity and small sample size. The important difference is the point estimates are different than the generalized Schaefer Production Model. The difference in the estimate of q is particularly important because it is used to estimate the biomass. These estimates can be seen in table 3.9.

Variable	Estimate
r	-0.103
	(0.620)
r/qK	1.36E-03
	(0.130)
q	2.67E-07
	(5.354E-06)
$\hat{\sigma}_e^2$	0.778
*p < 0.1,	**p <0.05, ***p<0.01

Table 3.8: Generalized Schaefer Model with SIMEX, Estimates for Growth Model

The biomass estimates, calculated in thousands of metric tons (1000MT), are shown in table 3.9. Variance and standard error estimates of the biomass were obtained using the two-stage bootstrap method.

Year	NEFSC est.	SIMEX est.	95% Lower	95% Upper	% Bias
2001	1294	1399	937	1861	8 %
2002	1207	1183	781	1585	-2%
2003	1128	1047	656	1438	-7%
2004	1104	956	603	1309	-13%
2005	1079	860	549	1171	-20%
2006	1013	757	481	1033	-25%
2007	912	626	400	852	-31%
2008	827	577	378	776	-30%
2009	750	531	336	726	-29%

Table 3.9: SIMEX Biomass Estimates (1000MT)

The average bias for the SIMEX estimates of biomass, using the NEFSC biomass estimates as truth, is -17%. This is less than half the amount of bias in the generalized Schaefer model estimates. Additionally, the 95% confidence interval in each year cover the NEFSC biomass estimate. These intervals are wider than the generalized Schaefer model, because they account for measurement error.

3.7 Discussion

This paper explores the impact of measurement error in the Schaefer Production Model and describes a Monte Carlo method for reducing bias in the parameters. The Schaefer production model is a two-stage model that estimates the biomass of a fishery using only information on catch and effort, usually obtained through vessel logbook data. We build on previous work by Zhang and Smith (2011) to reduce bias in the parameter estimates of a more generalized version of the Schaefer Production Model by using a Monte Carlo method called simulation extrapolation (SIMEX).

Using data from the Mid-Atlantic surfclam fishery from 2001-2009, along with scientific estimates of the biomass from NEFSC, we show that in small samples the remaining measurement error in the generalized Schaefer model can still contribute to significant bias in the estimates of biomass, up to -41%. A fishery manager using this method would not have much reason to be confident in setting a total allowable catch with such bias. Obtaining unbiased estimates of the biomass is crucial to effective fishery management. A total allowable catch that is set too high could put a fishery on a trajectory towards collapse, while a total allowable catch that is set too low can lead to economic loss of rents. We show that fishery managers can obtain more truthful estimates of the biomass after correcting for the measurement error with SIMEX. The results show that the bias in the biological parameters, and biomass estimates, is dramatically less than either the classic Schaefer model or the generalized Schaefer model.

The generalized Schaefer production model attempts to simultaneously deal with three problems inherent in the Schaefer production model, including; 1) the error in the production function, 2) the error in the second stage growth model, and 3) restrictive functional form in the production function. The idea behind the generalized Schaefer production model is that a stock index can be created using a fixed effects estimator for the production function. Zhang and Smith (2011) show that this stock index will be asymptotically consistent, reducing bias in the parameters of the second-stage growth model. To reduce bias using the SIMEX algorithm, we first estimate the measurement error variance in the first stage production function. This gives an estimate of the measurement error in the stock index, which serves as a proxy for the biomass in the second stage growth model. With the measurement error variance estimated, we apply simulation extrapolation (SIMEX) to the second stage growth model. The average remaining bias in the SIMEX estimates is -17%. This is less than half the average bias in the generalized Schaefer production model, at -41%. Additionally, all nine 95% confidence intervals from the SIMEX estimates cover the NEFSC estimate. Only one confidence interval from the generalized Schaefer production model covers the NEFSC estimate, while none of the intervals cover the truth using the classic Schaefer production model.

While these findings would suggest this additional measurement error correction method improves on previous models, it should be noted that these findings only apply to the Mid-Atlantic surfclam fishery. Because the surfclam is a slow-growing mollusk, the dynamics of the growth model may be very different from a pelagic fishery. We would need to apply this method in several different kinds of fisheries, including pelagic and other fisheries with significantly more variation in the biomass, before making any definitive conclusions about the model. We think this model is a good first step towards giving fishery managers a more reliable estimate of the biomass when they only have catch and effort data obtained through vessel logbooks.

CHAPTER 4

MEASURING CHANGES IN PRODUCTION EFFICIENCY: A BAYESIAN APPROACH

4.1 Introduction

Using panel data from the Mid-Atlantic surfclam industry from 2001-2009, this paper examines changes in production efficiency caused by changes in the industrial organization of the fleet. The motivation for this analysis is due to two important factors, 1) a market crisis that occurred in the fishing fleet in 2005 and 2) evidence of persistent declines in landings per-unit-effort (LPUE). We specify a stochastic frontier model with a time-varying inefficiency term and examine the changes in marginal productivity of inputs, as well as overall technical efficiency. We also show how a Bayesian stochastic frontier model will yield results consistent with traditional maximum likelihood estimation techniques (Battese and Coelli 1995).

In 1990, the Mid-Atlantic surflclam fishery became the first U.S. fishery to be regulated using Individual Tradable Quotas (ITQs). The implementation of ITQs resulted in initial increases in efficiency for the surfclam industry, during the period 1990-1995 (Brandt 2007, Walden et al. 2012), due to the exit of many inefficient vessels. Many vessel owners stopped harvesting after the implementation of ITQs but continued to own quota in the fishery. A more recent long-term analysis shows that productivity has since declined in the period 2000-2012, largely due to spatial changes in the biomass (Walden et al. 2012). The exit of many inefficient vessels and consolidation of ITQ by large upstream surfclam processors has resulted in a more consolidated industry. Much of the remaining fleet is either vertically integrated with processors or horizontally integrated fleets. Important questions about whether this consolidation is likely to continue and the impact it will have on fishing communities is beyond the scope of this study. However, we think that we can shed light on the economic impact of recent changes in the industry and fishery itself.

In 2005 a ”market crisis” occurred in the fishery, causing much of the existing surfclam fleet to leave the fishery (Mid-Atlantic Fishery Management Council 2010).This crisis was caused by upstream surfclam buyers switching to imported clams from Vietnam and Canada. During this time period an influx of smaller vessels is observed. After 2005, some vessels did return to the fishery, but the number of active vessels dropped substantially, from about forty vessels to approximately thirty-five. We find significant changes in technical efficiency and marginal productivity can be observed during this time period. In addition, the data reveal a significant decline in landings per-unit-effort (LPUE), caused by decreasing resource abundance (Weinberg 2005). This will likely result in more vessels exiting the fishery and lead to further consolidation of the industry. Our analysis of the data shows that mean technical efficiency for the fishery has been relatively flat, or possibly slightly increasing during the nine year period. Because ITQs create a barrier to entry for new firms, further gains in technical and allocative efficiency may not be realized. We consider these implications when discussing the results of this study.

Stochastic frontier models have been used in analysis of productivity and firm efficiency since first proposed by Meeusen and Broeck (1977) and Aigner et al. (1977). Meeusen and Broeck (1977) used an exponential inefficiency term, while Aigner et al. (1977) use a half-normal distribution to model the inefficiency of firms. Stochastic frontier analysis has been widely used to analyze productive, allocative and cost efficiency in the fisheries literature (Squires and Kirkley 1999, Pascoe and Coglan 2002, Brandt 2007). Data envelope analysis (DEA) is another popular technique used ti assess productive efficiency in fisheries (Felthoven 2002, Kirkley et al. 2004, Tingley et al. 2005). DEA also estimates a production frontier but uses linear programming methods to estimate a deterministic production frontier, rather than a stochastic frontier. Kirkley et al. (1995) points out that given the inherent stochastic nature of fisheries production data, including weather and other environmental shocks, stochastic frontier models are more appropriate. We proceed using this stochastic frontier methodology in both a Bayesian and maximum likelihood estimation framework.

More recently, frontier analysis has moved to using Bayesian methods for estimating technical and allocative efficiency. Previous studies of include Osiewalski and Steel (1998), Ehlers (2011), Griffin and Steel (2007), Fernandez et al. (1997), Tsionas (2005) and Van den Broeck et al. (1994). Van den Broeck et al. (1994) introduced Bayesian methods for stochastic frontier models and showed the advantages of exact, small-sample inference for efficiencies, as well as methods to incorporate prior information into the model. Ehlers (2011) showed how to estimate stochastic frontier models using JAGS (Plummer et al. 2003), which stands for Just Another Gibbs Sampler, using the statistical software R (R Core Team 2013). Griffin and Steel (2007) outlined estimation of a stochastic frontier model using another popular Bayesian software platform for social sciences, WinBUGS (Lunn et al. 2000). One particular advantage of using Bayesian methods is that posterior inferences for technical efficiency are easy to produce using readily available software, and often have smaller variances than frequentist methods.

To specify the functional form of the stochastic frontier we use a model similar to Brandt (2007). Our model differs in that we do not have data on vertical and horizontal integration of firms in the industry, and we use a time-varying half-normal inefficiency distribution. Estimation of the Bayesian model is accomplished using the R package, R2jags (Su and Yajima 2013), a package which calls JAGS from the R console. We compare these Bayesian estimates to a maximum likelihood model frontier model, which is estimated using the Frontier package (Coelli and Henningsen 2013) in R.We compare the two models to see if the estimates of marginal productivity and technical efficiency differ. For the Bayesian model we specify uninformative priors, so that the model estimates reflect the underlying data.

The rest of the paper will proceed as follows; Section 4.2 discuss the recent changes in the surfclam industry and describes the data in more detail. Section 4.3 builds the Bayesian framework for the stochastic frontier model, and describes the workings of the Gibbs Sampler. Section 4.4 presents the estimates from the two model, and Section 4.5 discusses the implications of the findings and areas of future research.

4.2 Motivation and Data

The motivation for this paper comes from recent evidence of declining productivity in the fishery (Walden et al. 2012) and evidence of a structural change in the surfclam fleet in 2005 (Mid-Atlantic Fishery Management Council 2010). In 2005, a ″market crisis″ caused a large number of vessels to leave the surfclam fishery. According to the Mid-Atlantic Fishery Management Council (2010), a combination of factors led to this substantial change in the fleet structure. Industry members report that major users of clam meats reduced their purchases and stopped advertising surfclam products, such as clam chowder. Additionally, an excess supply of clams, caused by increased imports from from Canada and Vietnam, led to a drop in the ex-vessel price for clam meat. A second factor in the market crisis was the reduced operations of the second largest surfclam processing plant, located in Mappsville, Virginia. In addition, vessels that were vertically integrated with this processor were sold to another firm, resulting in increased consolidation of the industry. As shown in table 4.1, prior to the restructuring of the surfclam fleet in 2005, about 35-40 vessels operated in a given year. Some of these vessels harvest surfclam only (SC), while others harvest both surfclam and ocean quahog (SC and OC). After 2005, the number of total vessels harvesting surfclam only (SC) drops initially, but then recovers. Additionally, the number of vessels harvesting both surfclams and ocean quoahog drops substantially as well.

Vessel Type	2001	2002	2003	2004	2005	2006	2007	2008	2009
SC and OQ	14	16	11	14	12	9	9	8	8
SC only	21	23	23	21	24	20	24	24	28
Total	35	39	34	35	36	29	33	32	36

Table 4.1: Number of Vessels by Year

(Mid-Atlantic Fishery Management Council 2010)

In 2005, the data shows that 14 previously unobserved vessels entered the fishery. These vessels were smaller, with a mean length of around sixty feet, well below the nine year average of eighty-five feet. These vessels then exited the fishery in the following year. The impact of these smaller vessels on the length distribution of the fleet in 2005 can be seen in figure 4.1.

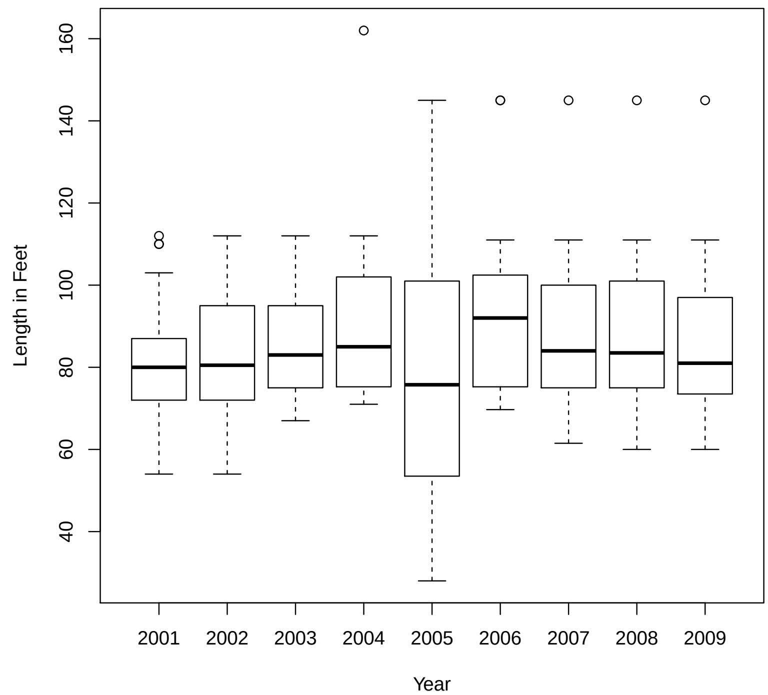

Figure 4.1: Distribution of Vessel Length by Year

In addition to the market crisis in 2005, there are several other factors which are driving the industry towards consolidation. Landings per-unit-effort (LPUE) is a widely used measure of effort in fisheries. There are substantial data indicating that LPUE has been declining at a rate of 10% per year during the 2001-2009 time period (Mid-Atlantic Fishery Management Council 2010). The result is vessels have had to spend more time-at-sea in order to catch their quota. Additionally, harvesting costs have increased over time as firms have to pay for additional fuel. The nature of the contracts between vessels and surfclam processors means that vessel captains only get to choose where they harvest, and how long they operate at sea. All other aspects of the harvesting trips are stipulated in the contractual agreements between the processor and vessel owner.

Table 4.2 shows how average time fishing has been increasing over the nine year time period, in addition to average length of the vessels. The one exception is 2005, when many smaller vessels entered the fishery during the market crisis.

Year	Length (ft)	Time Fishing (hrs)	Harvest (bu)
2001	82.2	751.6	86297.2
2002	83.5	820.1	86309.9
2003	85.3	1005.9	97393.2
2004	90.1	1134.7	100783.5
2005	76.9	1116.7	93627.7
2006	92.5	1408.9	109071.9
2007	87.8	1604.8	104217.8
2008	87.5	1712.3	97284.1
2009	85.9	1430.8	74294

Table 4.2: Average Vessel Length, Time Fishing, and Harvest

The data used for estimation come from the National Marine Fishery Service logbook reporting system, which documents every harvesting trip taken by every vessel in the Mid-Atlantic surfclam fishery in the U.S. EEZ (3-200miles offshore). The logbook data are a panel data set containing approximately 24,000 vessel-trip observations, for years 2001-2009. The trip-level data set includes variables such as bushels harvested, time fishing, time-at-sea, and vessel characteristics such as vessel length, gross-tons and horsepower. The trip-level data has a total of eighty-eight different vessels observed over the nine year period. In order to estimate the stochastic frontier, additional survey data collected by the National Oceanic and Atmospheric Administration's (NOAA) Northeast Fisheries Science Center (NEFSC)-Resource Evaluation and Assessment Division are used as an estimate of biomass.

Before estimating the model we dropped observations on non-harvesting vessels, such as trips by research vessels. The resulting data had a total of seventy vessels observed over the nine year time period. We then aggregated the data over trips in a year, resulting in 285 vessel-year observations. This was done to make estimation of the model easier. Before aggregating, we estimated the same frontier model using both the trip-level and aggregate data, to check for possible aggregation bias. Because the model results were not significantly different, we report summary statistics from the aggregate dataset, shown in table 4.3. The estimates of biomass come from the Northeast Fisheries Science Center Northeast Fisheries Science Center (2010), which uses a biological model to estimate the biomass. The data shows significant variability in total harvest and time fishing over the nine year period, but variability in vessel length is not as large. Because biomass is only estimated annually, we have only nine observations.

	Obs	Mean	Std.Dev	Min	Max
Harvest (bushels)	285	93749	82007.6	864	442496
Time Fishing (hours)	285	1209.2	951.9	58	3959.4
Fuel (gallons)	285	65896	65979.3	876	388204
Length (feet)	70	85.7	18.4	28	162
Biomass (1000 metric tons)	9	1037	171.9	750	1294

Table 4.3: Summary statistics 2001-2009

4.2.1 Bayesian Methods

There are numerous reasons why Bayesian methods are becoming more widely used in the social sciences. In the social sciences, programs such as WinBUGS and R can allow researchers with training in Bayesian methods to run sophisticated models without needing to know how to program a Gibbs sampler or other Markov Chain Monte Carlo (MCMC) routine. MCMC routines are a ubiquitous tool for estimating integrals over complicated probability distributions. Because Bayesian models rely on integration of probability distributions over a large number of dimensions, solutions to these integrals are very difficult to derive without using MCMC simulation methods. The computational burden of these methods was a limiting factor until the past few decades, but with advances in computing power Bayesian methods are now becoming widespread in the social sciences.

Another advantage of Bayesian methods is model flexibility. Using programs such as R allow researchers to specify hierarchical or nonlinear functional forms, which the MCMC methods can easily accommodate with little additional computational burden. For example, the stochastic frontier model can be specified with many different distributional assumptions on the inefficiency term. Another advantage to using Bayesian methods particular to the stochastic frontier model, is that credible intervals for the posterior estimates of technical efficiency can be obtained quite easily.

Western and Jackman (1994) provide a compelling argument for using Bayesian methods with observational data. Observational data, which are commonly used in economics, political science and other social science disciplines, are fraught with many problems, including measurement error and colinearity in the variables. This paper highlights two major issues surrounding the use of traditional frequentist methods for estimating linear regression models in social sciences. The first issue is that data often are not generated by a random process and often constitute all available observations from a population. This means that using frequentist inference, which relies on long-run behavior of a repeatable data mechanism, may not be appropriate. Given that our data contain all observations in the federally regulated surfclam fishery, this scenario applies to our data as well. Second, many times observational data used in the social sciences has colinearity among variables. Western and Jackman (1994) point out that unless the colinearity is extremely high, the only problem for the applied researcher occurs when the parameter estimates do not have the expected sign. In this situation, the problem is that the variables do not carry much additional information independently, and thus the parameter estimates may be imprecise. In economics, this can lead to parameter estimates with signs that are not consistent with economic theory.

Bayesian methods allow researchers to place prior distributions on parameters, thus giving them the expected signs using theory or previous studies as a guide. This allows for more ”sensible” results. In the frequentist paradigm, this is usually accomplished by imposing restrictions on the parameters. A common restriction could be excluding highly co-linear variables from the model, effectively setting the parameters equal to zero. Western and Jackman (1994) show that with the proper model setup, it is unnecessary to exclude these variables from the model, and that Bayesian methods can often result in smaller intervals for the estimated parameters.

In the next section we specify a Bayesian Model for measuring production efficiency, and derive the conditional distributions, which are functions of the priors, and the joint likelihood of the model and data. We also describe how the Gibbs sampler is used to estimate the posterior distributions for the parameters of interest using these conditional distributions.

4.3 Methodology

4.3.1 Modeling Production Efficiency

Production frontiers are used to model the maximum level of output a firm can obtain using a certain technology and given level of inputs. The stochastic production frontier model allows for a random error (e) that affects firm output, such as weather, and a one-sided error term (b) that captures observed heterogeneity due to factors that affect technical inefficiency. The fisheries literature has identified many factors that affect firm technical inefficiency, including vessel age (Pascoe and Coglan 2002) and skipper ability (Kirkley et al. 1998, Squires and Kirkley 1999). Kirkley et al. (1998) note that additional factors such as skipper and crew motivation are probably just as important to explaining firm technical inefficiency, but are difficult to measure. We allow the inefficiency term to vary by time in this model specification following (Kumbhakar 1990, Battese and Coelli 1995), allowing for an error components frontier. The motivation for a time varying inefficiency term, as opposed to a time-invariant term, comes from the logbook data and the institutional background of the Mid-Atlantic surfclam fishery. Because of significant changes in the surfclam fleet composition between 2001-2009, we think a time-varying inefficiency term is appropriate.

For panel data with N observed vessels and T time periods, the model can be expressed as

$$y_{it} = f(\mathbf{x}_{it}; \beta) - b_{it} + e_{it},\ i = 1, \ldots, N \ and\ t = 1, \ldots, T \tag{4.1}$$

where y_{it} is the natural logarithm of output, $\mathbf{x}_{it}$ is a vector of natural logarithms of inputs, including an intercept, β is a vector of coefficients, and e_{it} is independently and identically distributed with mean zero and variance σ_e^2. Additionally, e_{it} is assumed to be independent of b_{it}.

The functional form for the production frontier is assumed to be Cobb-Douglas, following (Brandt 2007). We specify the stochastic frontier model for vessel i in year t. Let $i = 1, \ldots, 70$ denote the vessel, and $t = t_1, \ldots, t_{n_i}$ denote the n_i years in which vessel i is observed. The stochastic frontier model for vessel i in year t is then

$$y_{it}|b_i, e_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{t3} - b_{it} + e_{it} \tag{4.2}$$

where e_{it} is assumed to be identically and independently distributed $N(0, \sigma^2)$, b_{it} is a non-negative half-normal random variable capturing vessel inefficiency, y_{it} is natural log of bushels, x_{it1} is natural log of time fishing, x_{it2} is natural log of vessel length, and x_{it3} is the natural log of biomass.

The b_{it} term measures the technical inefficiency of the i^{th} vessel in time period t. Following Lee and Schmidt (1993) we define the time-varying inefficiency as

$$b_{it} = \gamma(t) b_i \tag{4.3}$$

To model the effect of time on technical inefficiency, $\gamma(t)$, we follow Battese and Coelli (1992), who propose the model

$$\gamma(t) = exp[\eta(t - T)] \tag{4.4}$$

where positive η indicates increasing vessel efficiency over time. Technical efficiency of the i^{th} vessel in period t is estimated by TE_{it} = $exp(-b_{it})$.

To see the differences between frequentist and Bayesian methods, we first estimate this model with a time-varying efficiency term using maximum likelihood estimation (MLE) following (Battese and Coelli 1992). Next we estimate the Bayesian stochastic frontier model using the R2jags package in R. For the Bayesian analysis we report posterior means and 95% credible intervals from the parameters. A credible interval is the Bayesian equivalent to the confidence interval used in frequentist statistics. However, credible intervals can be interpreted as probabilities, unlike frequentist intervals, which rely on repeated sampling. For example, a 95% credible interval says that the probability the true parameter lies within the interval is 95%. Appendix C contains a summary of these posterior credible intervals and means.

In the next section we discuss the Bayesian model for production efficiency. Before we can estimate the Bayesian model we need to specify a full likelihood function for the data. In the next section we build the likelihood for the stochastic production frontier and then show how the complete conditional posterior distributions can be found. Last, we discuss the MCMC method used for estimating these posterior distributions, known as Gibbs sampling.

4.3.2 Bayesian Model for Production Efficiency

In the Bayesian model, the goal is to obtain posterior distributions for the parameters of interest. To do this we must first specify conditional distributions for the parameters of interest, conditioned on the other parameters in the model and the data. Following (Ehlers 2011), let β contain the parameters in the production function, θ is the set of hyperparameters in the prior distribution of b_{it}, and $\mathbf{X}$ is the matrix with log inputs. The vector of hyperparameters for $\mathbf{b_{it}}$ contain the variance of the inefficiency term, σ_b^2, and the half-normal prior assumption. Combining the likelihood and the joint prior distribution, the joint posterior distribution is

$$p(\beta, \sigma^2, \mathbf{b}, \theta|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{X}, \beta, \mathbf{b}, \sigma^2) \prod_{i=1}^{70} \prod_{t=1}^{t_{n_i}} p(b_{it}|\theta) p(\beta) p(\sigma^2) p(\theta) \tag{4.5}$$

where $p(\mathbf{y}|\mathbf{X}, \beta, \mathbf{b}, \sigma^2)$ is the joint likelihood of the data, using the model defined in equation 4.2. The joint posterior described above is proportional to the likelihood times the joint prior.

Complete conditional distributions are required for JAGS to simulate posterior distributions for each parameter. The complete conditional distributions for β, b_{it}, σ^2 and θ are respectively given by

$$p(\beta|\mathbf{y}, \theta, \sigma^2, \mathbf{b}) \propto p(\mathbf{y}|\mathbf{X}, \beta, \mathbf{b}, \sigma^2) p(\beta) \tag{4.6}$$

$$p(b_{it}|\mathbf{y}, \beta, \theta, \sigma^2) \propto p(y_{it}|\mathbf{x_{it}}, \beta, \sigma^2) p(b_{it}|\theta) \tag{4.7}$$

$$p(\sigma^2|\mathbf{y}, \beta, \theta, \mathbf{b}) \propto p(\mathbf{y}|\mathbf{X}, \beta, \mathbf{b}, \sigma^2) p(\sigma^2) \tag{4.8}$$

$$p(\theta|\mathbf{y}, \beta, \sigma^2, \mathbf{b}) \propto \prod_{i=1}^{70} \prod_{t=1}^{t_{n_i}} p(b_{it}|\theta)p(\theta) \tag{4.9}$$

Next we discuss the choice of prior distributions for the parameters in the stochastic frontier model.

4.3.3 Priors

We chose relatively uninformative priors for the model, seen in table 4.4, so that the estimated posterior moments reflect the underlying data. We tested the model with several different prior distributions for the production parameters and inefficiency terms. These did not affect the model estimates significantly.

Parameter	Distribution Family	Location and Scale
β	Normal	$\mu = 1, precision = 1E - 05$
$\mathbf{b}$	Truncated Normal	$\mu = 0, \sigma_b^2$
σ_b^2	Gamma	$\alpha = 3, \beta = 1$
σ_e^2	Gamma	$\alpha = 3, \beta = 1$
η	Normal	$\mu = 0, precision = 4$

Table 4.4: Priors for Bayesian Model

In the next section, we describe how the conditional distributions shown above are used by the Gibbs sampler to find the posterior distributions for the parameters of interest. Later we estimate expected values and standard deviations for these posteriors in order to make inferences about the model.

4.3.4 Gibbs Sampler

The Gibbs sampler is a MCMC algorithm that simulates posterior distributions for the parameters of interest. According to Gill (2002), the main idea of the Gibbs sampler is to get a marginal distribution for each parameter by iteratively conditioning on iterim values of the other parameters in a continuing cycle. The cycle continues until the samples generated empirical approximate the desired marginal distributions.

The Gibbs sampler is first defined by the conditional distributions for each parameter in the model, as shown in section 4.3.2. These are conditional distributions in that they depend on the other parameters in the model, including the priors, and the data. The model specification clearly affects these conditional distributions, and thus the resulting MCMC inferences. A ”transition kernel” for the Markov Chain is created by iteratively cycling through these distributions, drawing values that are conditioned on the latest draws of the dependencies, i.e. the Markovian property. It has been shown that when the MCMC algorithm is allowed to run long enough it will settle on the limiting distributions that characterize the marginal posteriors of the parameters in the model.

The MCMC sampling done in this paper using the software JAGS[1]. The JAGS software is run in the R platform using the package, R2Jags. The R2jags package also allows for diagnostics to be performed on the output analysis, to check the mixing properties of the MCMC chains. In the next section we present results from estimation of the Bayesian and maximum likelihood stochastic frontier models.

4.4 Results

In this section we display estimates of the stochastic frontier model from equation 4.2 using both the maximum likelihood (ML Model) and Bayesian estimator (Bayes Model). We first estimate using the entire dataset, years 2001-2009. We then break the data up into three separate time periods, to examine the impact of the 2005 market crisis. We look at time periods before the crisis, 2001-2004, and time periods during/after the crisis, 2005-2009 and 2006-2009. We then compare pa-

[1] JAGS was originally developed as a clone to the BUGS package (Bayesian Inference Using Gibbs Sampling) (Spiegelhalter et al. 1996) for graphical modeling.

rameter estimates for the production frontier, and estimates of technical efficiency across the time periods. Lastly, we look at how technical efficiency is affected by factors such as vessel age, hull material and region.

4.4.1 Model Estimates for 2001-2009

For the Bayesian Model, we place normal distributions on all priors for the fixed effects parameters, and a half-normal prior on the inefficiency distribution. In each set of estimates, the parameter λ captures the ratio of variability due to inefficiency to total variability in the model, and is defined as $\lambda = \frac{\sigma_b^2}{\sigma_e^2+\sigma_b^2}$. The first model shown below in table 4.5 is the maximum likelihood model, using methods described in (Battese and Coelli 1992).

	Estimate	Std. Error	95% Conf. Interval
Intercept	-7.60***	1.35	(-10.25,-4.95)
$log(timefish)$	1.05***	0.02	(1.01,1.09)
$log(length)$	0.21*	0.09	(0.03,0.39)
$log(biomass)$	1.59***	0.12	(1.35,1.83)
λ	0.81***	0.06	(0.69,0.93)
η	-0.04	0.03	(-0.10,0.02)
Mean T.E.	0.72		
σ_b^2	1.45		
σ_e^2	0.34***	0.10	
LogLik	-62.88		
Note: *p <.05 ,	**p <.01 ,	***p <.001	

Table 4.5: Maximum Likelihood Model

The estimated coefficients for $log(timefish)$, $log(length)$ and $log(biomass)$ are all positive, as economic theory would suggest. Because the model is in natural log form, the coefficients can be interpreted as elasticities. A $\lambda = 0.81$ suggests that much of the variability in the two-part error term is due to technical inefficiency, reinforcing the choice of a stochastic frontier model. The mean technical efficiency over the nine year time period is 0.72, however because the 95% confidence interval for η covers 0, this suggests no change in mean technical efficiency over time.

Next, we estimate the same model in a Bayesian framework. Table 4.6 shows estimates from the Bayesian Model with time varying technical efficiency. We chose to run the MCMC algorithm using four MCMC chains with different starting values. This has the advantage of increasing the mixing of the MCMC chains and speeding convergence. We report results from 50,000 simulations due to the MCMC chains showing good mixing. The table reports posterior means and 95% credible intervals for parameters in the model.

Variable	Mean	95% Credible Int.
Intercept	-7.45	(-10.01, -4.73)
$log(timefish)$	1.05	(1.00,1.09)
$log(length)$	0.21	(-0.01,0.43)
$log(biomass)$	1.56	(1.22,1.88)
λ	0.79	(0.67,0.88)
η	0.04	(-0.01,0.10)
Mean T.E.	0.72	
σ_b^2	0.30	(0.16,0.52)
σ_e^2	0.07	(0.06,0.09)

Table 4.6: Bayesian Model 2001-2009

The magnitudes of the production frontier estimates in tables 4.5 and 4.6 are very similar. The input elasticities for $timefishing$, $length$ and $biomass$ are all positive, which is consistent with economic theory. Mean technical efficiency exactly the same between the two models are 0.72. The 95% credible interval for η in the Bayesian model suggests that technical efficiency is not changing over time, which is also consistent with the ML model.

Recent studies have used much longer time periods, 1981-2012, and found that long-run efficiency may be declining in the fishery (Walden et al. 2012).Walden et al. (2012) uses a Malmquist Index (MI) to measure changes in industry productivity. One advantage of the MI is that changes in productivity can be broken into changes in technical efficiency, scale efficiency and technical change. The stochastic frontier model as currently specified can only measure changes in technical efficiency over time. One drawback of this model is that we cannot disentangle technical change from the technical efficiency changes in the current model. Walden et al. (2012) claim that technical change is the biggest factor in explaining why productivity in the industry has not dropped more substantially in the past decade. If over capitalization is still an issue, meaning the fleet is still too large, then further gains in technical efficiency may be possible from continued exit of excess capital. As the less technically efficient vessels exit the fishery, mean technical efficiency will continue to increase. The slow exit of excess capital in the Mid-Atlantic surfclam fishery has been well documented (Weninger and Just 1997).

4.4.2 Model Estimates for 2001-2004, 2005-2009 and 2006-2009

Next, we divide the data in different time periods, and look at estimates of technical efficiency and marginal productivity from before and after the fleet structural change in 2005. We divide the data into periods 2001-2004, 2005-2009 and 2006-2009. There are 133 observations in 2001-2004 time period, 152 in the 2005-2009 time period, and 124 observations in the 2006-2009 period. Table 4.7 shows estimates from the maximum likelihood stochastic frontier model. 95% confidence intervals are reported in the line below each variable.

ML Model	2001-2004		2005-2009		2006-2009	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Intercept	-6.81	3.78	-8.41***	1.92	-10.63***	2.23
	(-14.22,0.60)		(-12.17,-4.67)		(-15.00,-6.26)	
$log(timefish)$	0.93***	0.03	1.10***	0.03	1.06***	0.03
	(0.87,0.99)		(1.04,1.16)		(1.00,1.12)	
$log(length)$	0.89***	0.27	0.23*	0.11	0.65***	0.19
	(0.36,1.42)		(0.01,0.45)		(0.28,1.02)	
$log(biomass)$	1.17*	0.50	1.64***	0.26	1.73***	0.33
	(0.19,2.15)		(1.13,2.15)		(1.08,2.38)	
λ	0.91***	0.03	0.79***	0.07	0.71***	0.11
	(0.85,0.97)		(0.65,0.93)		(0.49,0.93)	
η	-0.15*	0.06	0.10	0.06	0.2*	0.08
	(-0.27,-0.03)		(-0.02,0.22)		(0.04,0.36)	
Mean T.E.	0.69		0.68		0.72	
σ_b^2	4.35		0.87		0.42	
σ_e^2	0.43***	0.13	0.23**	0.07	0.18**	0.06
LogLik	-16.97		-29.15		-24.82	
Note: *p $<$.05 ,	**p $<$.01 ,	***p $<$.001				

Table 4.7: Maximum Likelihood Model

Since the output and inputs are on the natural log scale, the coefficients on the production function can be interpreted as elasticities. For example, an estimate for β_1 of 1.05 indicates that a 1% increase in time fishing causes a 1.05% increase in bushels harvested, all else constant. The elasticities for timefishing, length and biomass are all positive as economic theory would suggest. The sum of the elasticities for timefishing and length is greater than one, suggesting that firms are operating in the increasing returns-to-scale portion of the production function. This would not make economic sense if we assume firms are profit maximizers, however given the nature of the industry it is more likely they operate as cost minimizers, subject to a quantity that is determined by their quota holdings.

Table 4.8 shows estimates from the Bayesian stochastic frontier model for periods 2001-2009. We report posterior means and 95 % credible intervals for the parameters in the model. All three models showed good MCMC mixing at 50,000 simulations.

	2001-2004	2005-2009	2006-2009
Intercept	-7.68	-7.42	-9.13
	(-18.78,2.47)	(-11.45,-3.26)	(-14.60,-3.83)
$log(timefish)$	1.03	1.10	1.06
	(0.96,1.10)	(1.04,1.16)	(0.99,1.14)
$log(length)$	0.33	0.22	0.68
	(-0.07,0.71)	(-0.03,0.46)	(0.23,1.15)
$log(biomass)$	01.53	1.51	1.48
	(0.13,3.07)	(0.94,2.05)	(0.69,2.28)
λ	0.63	0.76	0.72
	(0.47,0.78)	(0.62,0.87)	(0.57,0.85)
η	0.17	-0.06	-0.12
	(-0.11,0.58)	(-0.17,0.06)	(-0.27,0.04)
Mean T.E.	0.79	0.68	0.70
σ_b^2	0.17	0.23	0.21
	(0.09,0.32)	(0.12,0.407)	(0.11,0.38)
σ_e^2	0.10	0.07	0.07
	(0.08,0.13)	(0.05,0.09)	(0.06,0.10)

Table 4.8: Bayesian Model

The parameter estimates from both models are very similar. We assumed normal priors on the fixed effects in the Bayesian model, with large variances. This means that the priors have little impact on the posterior results, and the data drives the resulting estimates. It can be shown that the maximum likelihood estimates and Bayesian estimates will be the same when the priors placed on the Bayesian model are uninformative.The advantage of the Bayesian model remains in the inference, where a 95% credible interval can be interpreted as a probability interval, rather than an interval which relies on repeated sampling. One difference is seen in the estimates of the time effect (η) where the maximum likelihood model shows a declining technical efficiency in the 2001-2004 period, followed by significant increases in technical efficiency in the 2006-2009 period. The Bayesian model posterior estimates for η do not show a significant increase or decrease over the entire time period.

4.4.3 Marginal Productivity of Time Fishing and Length

It is well established that LPUE in the fishery has been declining at about 10% per year (Walden et al. 2012, Mid-Atlantic Fishery Management Council 2010). Given this fact, we would expect marginal productivity for time fishing to decrease. Firms may try to substitute capital (i.e. length of vessel) for time fishing in an attempt to find a cost minimizing solution, so we also examine marginal productivity of length.

Because the model is in natural logs, we can interpret the β's as elasticities. To get the marginal productivity of time fishing we take the derivative of $log(harvest)$ with respect to time fishing (x_1) and get

$$\frac{1}{y}\frac{\partial y}{\partial x_1} = \frac{\beta_1}{x_1} \tag{4.10}$$

which can be rewritten

$$\beta_1 = \frac{\partial y}{\partial x_1}\frac{x_1}{y} \tag{4.11}$$

where $\frac{\partial y}{\partial x_1}$ is the marginal productivity of time fishing, and β_1 represents the elasticity with respect to time fishing, all else constant. Similarly, the marginal productivity of length can be found by taking the derivative $log(harvest)$ with respect to x_2. Both marginal productivities are displayed for each year in table 4.9. The results show how marginal productivities (MP) of time fishing and length are declining over the nine year time period

Year	MP Time Fishing	MP Length
2001	120.6	325.4
2002	110.5	320.6
2003	101.6	353.8
2004	93.3	346.8
2005	88.0	377.4
2006	81.3	365.5
2007	68.2	368.0
2008	59.7	344.6
2009	54.5	268.3

Table 4.9: Marginal Productivities of Time Fishing and Length 2001-2009

This decline in LPUE reflects both an increase in average time fishing (x_1), and a roughly constant harvest (y) during the nine year period. The result of declining LPUE is an increasing ratio of $\frac{x_1}{y}$ over time, which is shown on the second part of right-hand side of equation 4.11. The modest increase in $\beta 1$ over the nine year time period indicates that the declines in marginal productivity are being outpaced by the decline in LPUE. The result is that vessel costs are increasing, as captains have to spend more on fuel and maintenance to meet their harvest quotas. Additionally, the model estimates show that the input elasticity of capital, or the coefficient on $log(length)$, is greater in the periods before and after 2005. This suggests that during the restructuring of the fishery in 2005, when many smaller vessels entered the fishery, vessel $length$ had a smaller marginal effect on total catch. This is confirmed by the data on these vessels, which shows they were indeed smaller on average.

The combination of increasing technical efficiency with declining LPUE on firm profitability is difficult to measure. The fact that the model suggests that firms are becoming more technically efficiency over time suggests that they would be more profitable, all else equal. However, the continued consolidation of the industry is likely due to increasing costs and efficiencies of scale. Under a fixed quota system, there are three ways a vessel can become more profitable, by increasing productivity, changing output mix, buying additional quota, or some combination of the three (Walden et al. 2012). The combination of decreasing LPUE with declining demand for surfclams (Mid-Atlantic Fishery Management Council 2010) will more than likely facilitate further consolidation of the industry, with the most efficient vessels remaining active. In the next section we examine how technical efficiency has changed over time and look at how the 2005 market crisis affected these estimates.

4.4.4 Technical Efficiency Estimates

In this section we look at estimates of mean technical efficiency for the four time periods, 2001-2009, 2001-2004, 2005-2009, and 2006-2009. We compare and contrast the different time periods from the ML model and Bayesian model.

Year	ML Model	Bayes Model
2001	0.74	0.74
2002	0.73	0.73
2003	0.75	0.75
2004	0.73	0.73
2005	0.77	0.76
2006	0.70	0.70
2007	0.70	0.70
2008	0.71	0.70
2009	0.70	0.69

Table 4.10: Mean Technical Efficiency Estimates for 2001-2009

Comparing the two sets of estimates in table 4.10, we can see that technical efficiency was relatively flat until 2005. Both models show that technical efficiency increases in 2005, and then declines in the following year. We perform a Wilcoxin rank sum test of difference in median technical efficiency between 2005 and 2006. The results show that median technical efficiency in 2005 is statistically greater than 2006 at the 10% level. This is consistent with (Walden et al. 2012) who find an increase in technical efficiency, technical change and scale efficiency in 2005, followed by a drop in all three in 2006. Again, while the models do not find a significant time effect for technical efficiency, from table 4.10 it is clear that mean technical efficiency in the years after 2005 is lower than before. This suggests that the market crisis and subsequent consolidation of the industry did not positively impact overall technical efficiency.

In tables 4.11 we display mean technical efficiency estimates for the period 2001-2004. With the 2001-2004 data, the model predicts much higher levels of technical efficiency for the Bayesian model than the ML model.

Year	ML Model	Bayes Model
2001	0.72	0.83
2002	0.70	0.80
2003	0.70	0.78
2004	0.64	0.74

Table 4.11: Mean Technical Efficiency Estimates 2001-2004

In tables 4.12 we display mean technical efficiency estimates for the periods 2005-2009 and 2006-2009. Using the estimates in table 4.12, the drop in technical efficiency from 2005 to 2006 is significant at the 10% level.

Year	ML 05-09	ML 06-09	Bayes 05-09	Bayes 06-09
2005	0.71	N/A	0.71	N/A
2006	0.63	0.64	0.64	0.64
2007	0.66	0.69	0.66	0.68
2008	0.70	0.75	0.69	0.72
2009	0.71	0.78	0.69	0.73

Table 4.12: Mean Technical Efficiency Estimates for 2005-2009 and 2006-2009

The impact of the market crisis can be seen in the smaller datasets, with both models showing a significant drop in overall technical efficiency in 2006. In the next section we use data on vessels characteristics, such as vessel age, vessel hull material and home port, to look at factors that affect technical efficiency.

4.4.5 Technical Efficiency Factors

In this section we analyze how technical efficiency varies by vessel age, hull material and home port state using the technical efficiency estimates obtained from the Bayesian model for years 2001-2009. We report these estimates because they are virtually identical to the results from the ML model. We find that the estimates are consistent with economic theory and the background of the fishery. We find that older vessels are significantly less technically efficient than newer vessels. This can also be seen when comparing the type of hull material, with fiberglass (FBG) vessels significantly more technically efficiency than older, wooden or steel vessels. When examining technical efficiency by home port state, we see significant differences between vessels operating from NJ and vessels operates from a northern port (RI, NY, MA,NH) or a southern port (NC, VA, MD).

Figure 4.2 shows a boxplot of technical efficiency versus vessel age. We perform a test to see if significant differences are present by vessel age. The age categories are; less than 10 years (n=29), 10-20 years (n=48), 20-30 years (n=130), 30-40 years (n=71) and 40+ years old (n=6). Because the data are not normally distributed we use a non-parametric test, the Kruskal-Wallis test. This test is the non-parametric equivalent to a one-way analysis of variance (ANOVA). The test for differences in vessel age is significant at the 1% level. The boxplots show that newer vessels, i.e. less than ten years old, are more technically efficient, as economic theory would suggest.

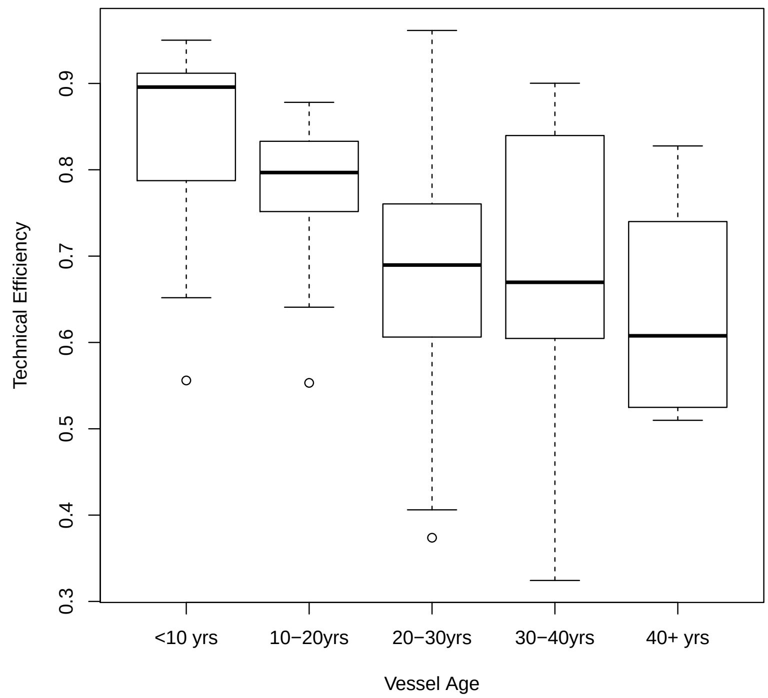

Figure 4.2: Technical Efficiency by Vessel Age

Figure 4.3 shows a boxplot of technical efficiency versus hull material (note: FBG = fiberglass hull). Over the nine year time period we observe fiberglass vessels only three times (n=3), while steel vessels are the most common (n=276) and wood vessels are second most common (n=6). A Kruskal-Wallis rank sum test reveals that there are significant differences in technical efficiency by hull material at the 1% level of significance. A pairwise comparison of medians reveals that fiberglass hulled vessels are statistically more technically efficient than either steel or wooden hulled vessels. This also reflects a difference in age, as the newer vessels are constructed of fiberglass. This also suggests that firms with significant amounts of capital will have an advantage because they can purchase newer vessels, which represent a large financial investment.

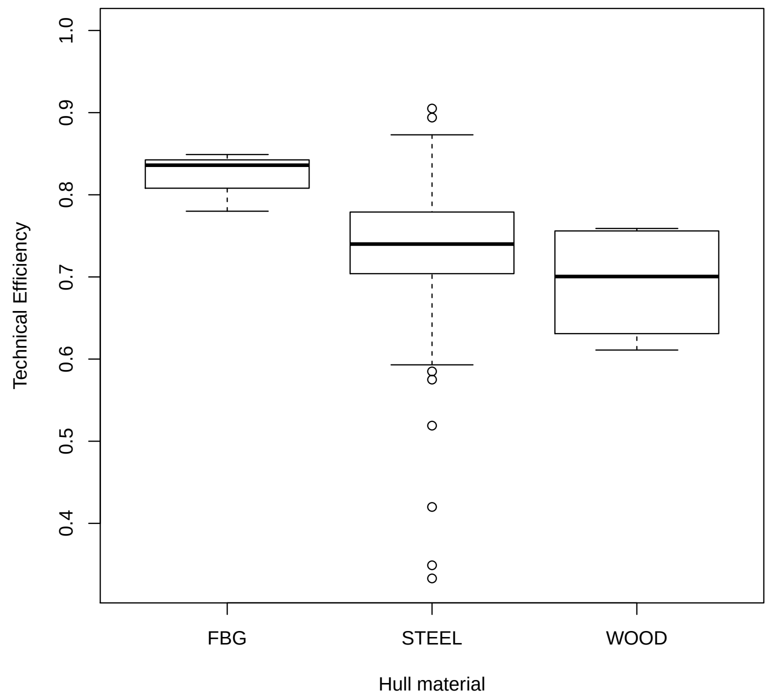

Figure 4.3: Technical Efficiency by Hull Material

Figure 4.4 shows a boxplot of technical efficiency by home port state. In terms of vessel trip frequency by home port state, the order from most frequent to least is: New Jersey, New York, Maryland, Rhode Island, Massachusetts, North Carolina, New Hampshire and Virginia. To analyze the regional differences in technical efficiency, we group the vessels into three categories based on home port. Vessels operating from NY, NH, MA and RI are put in the northern port category (n=61), vessels operating from MD, VA and NC are put in the southern category (n=35) and vessels operating from NJ (n=189) are put in a separate category. We give NJ a separate category because the majority of the fishing fleet are based out of ports in this region.

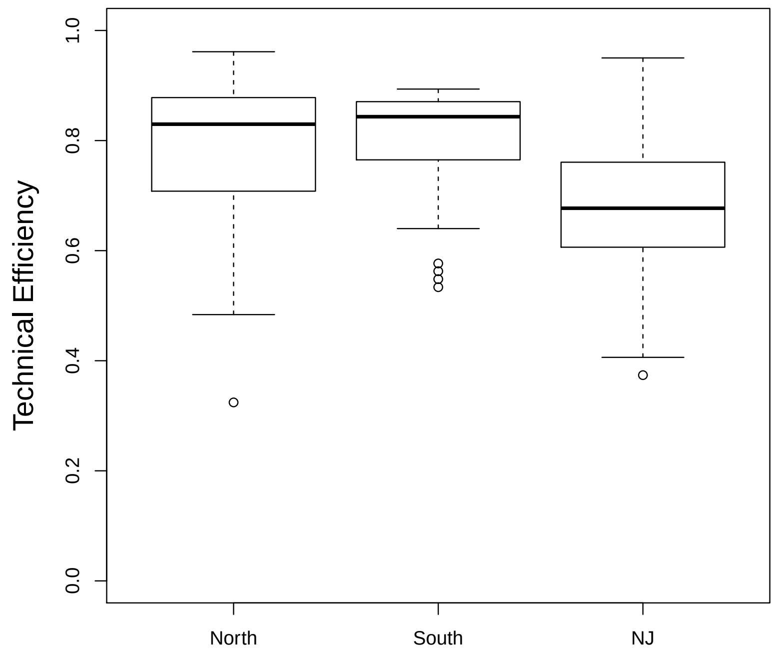

Figure 4.4: Technical Efficiency by Home Port State

Using a Kruskal-Wallis test, differences in technical efficiency by region are significantly different at the 1% level. Figure 4.4 shows that vessels from NJ have significantly lower technical efficiency. Additional boxplots for technical efficiency factors for the data periods of 2001-2004, 2005-2009 and 2006-2009 are provided in Appendix C.

4.5 Discussion

The mid-Atlantic surfclam industry has undergone significant changes since the introduction of ITQs in 1990. Considerable consolidation in the fishing fleet has occurred, and recently more vessels have exited the fishery due to changes in market conditions for surfclam. In 2005, a market crisis led to a substantial part of the fishing fleet leaving the fishery. This crisis resulted in many smaller vessels entering in the fishery in 2005, and then leaving the following year. After the 2005 crisis the average fleet was reduced to less than 40 vessels, indicating that many vessels had permanently exited the fishery. We specify a stochastic frontier model for the industry and use both Bayesian and maximum likelihood estimation methods to examine changes in production efficiency caused by the fleet restructuring in 2005. We also examine changes in productivity throughout the 2001-2009 time period. To accomplish this, we specify a time-varying technical efficiency term in a stochastic frontier model. This allows us to capture changes in technical efficiency for each vessel across all time periods.

The 2005 market crisis caused significant changes in technical efficiency for the surfclam fleet. Both the Bayesian and maximum likelihood models show a statistically significant decline in technical efficiency from 2005 to 2006, the year after the market crisis. However both models show that the effect of time on technical efficiency is not significant. This finding is somewhat consistent with (Walden et al. 2012), who find that technical efficiency in the industry is no longer increasing, and may be decreasing over time. It is also possible that the fleet is still overcapitalized and too many vessels are operating to maximize efficiency. Weninger (1998) predicted that the Mid-Atlantic surfclam and ocean quoahog fleet would eventually contract to between 21-25 vessels. Given that we currently observe approximately 35 active vessels each year, there may be more consolidation in the near future.

Additionally, factors impacting vessel technical efficiency include age of the vessel, hull material and home port state. We find that fishery managers should expect to see older vessels leave the fishery in the near future, given that they are statistically less technically efficient. Regional differences in technical efficiency can be seen, with vessels operating from home ports in NJ less technically efficient. When looking at the entire nine year period there is also clear evidence of declining marginal productivities of time fishing and length, most likely driven by declines in landings per-unit-effort (LPUE). The impact of continued declines in LPUE, accompanied by increasing harvesting costs, will likely drive the industry to become further consolidated in the future.

CHAPTER 5

CONCLUSION

This dissertation looks at estimation issues in two economic models widely used in fisheries, the stochastic frontier model and the Schaefer production model. The context for these two models is Mid-Atlantic surfclam fishery, for which we have panel data from the vessel logbook reporting system for years 2001-2009. In exploring estimation issues of these two models, surrounding measurement error in Chapters 2 and 3, and Bayesian methods for measuring production efficiency in Chapter 4, the goal of this research is to inform fishery policy. We show why these important econometric issues should be looked at carefully when determining fishery policy, to ensure that the best possible information is available.

Chapter 2 looks at how measurement error in a single covariate, the $log(biomass)$, can affect parameter and technical efficiency estimates in a stochastic frontier model. In a stochastic frontier model for a fishery, the biomass captures the resource abundance, much as acres of irrigated land would capture the abundance of land in agricultural. The motivation for the measurement error problem comes from the fact that the biomass of a fishery is never exactly known. We show that a naive estimation of this model, without accounting for the measurement error, would lead to biased and incorrect estimates of technical efficiency and production output elasticties. To correct for the measurement error, we use a Monte Carlo method for reducing bias known as Simulation Extrapolation (SIMEX). We assume an additive measurement error model in the $log(biomass)$ and estimate a stochastic frontier model using SIMEX.

In comparing the results of the naive and SIMEX estimates we see that the output elasticities for $log(timefishing)$, $log(biomass)$ and $log(length)$ are biased downwards in the naive model, while $log(gallons)$ is biased upwards. The naive estimates of technical efficiency also shown to be significantly greater than the SIMEX estimates, thus overstating the mean technical efficiency for the industry. In addition, the SIMEX estimates reflect that technical efficiency varies with vessel age, hull material and region in a manner consistent with economic theory. We conduct a Monte Carlo study to test the large sample properties of the SIMEX estimator with additive measurement error. We find that the SIMEX estimator has the smallest bias and root mean square error. These results show that naive estimation of a stochastic frontier model with additive measurement error in a covariate will give fishery managers poor information about the health of the industry, and what factors are driving technical efficiency. In particular, the regional differences in technical efficiency confirm that vessels in the southern region of the fishery are not as profitable, due to declining resource abundance and increasing harvesting costs.

Chapter 3 attempts to improve on the generalized Schaefer production model, using a two-stage method recently proposed by (Zhang and Smith 2011). The two-stage Schaefer production model uses catch and effort data from a fishery to estimate the stock or biomass. This model is widely is by fishery managers to determine the biological status of the stock, and to determine optimal levels of catch. The measurement error problem in the model comes from the use of a proxy variable for the latent stock term in the first stage production function. The generalized Schaefer production model proposed by (Zhang and Smith 2011) uses a fixed effects model to create a stock index in the first stage production function. The stock index is then substituted into the second stage growth model. Their study shows that estimates of the stock are consistent as the number of cross sections and time periods in the catch and effort data increase towards infinity. We show that the SIMEX estimator can be used to reduce bias in the stock estimates, making use of the variance in the first stage production function as a estimate of the measurement error variance.

Using logbook data from the Mid-Atlantic surfclam fishery we estimate the biomass of the fishery from 2001-2009. We do this with three models 1) the classic Schaefer production Model, 2) generalized Schaefer production model and 3) the SIMEX model. 95% confidence intervals were estimated for the biomass for each model. We comparing the 95% confidence intervals for each model to the biomass estimates from the Northeast Fisheries Science Center (NEFSC), which we consider to be the true biomass. The SIMEX estimates cover the NEFSC estimates in 9 out of 9 years, while the generalized Schaefer model covers it in one year. 95% confidence intervals for the classic Schaefer model estimates do not cover the NEFSC estimates in any year.

These results show that the SIMEX model does much better at estimating the fishery biomass. The other two models provide stock estimates which are severely biased downwards. We show that a fishery manager using this data would have very poor information to make decisions about the total catch, or biological status of the fishery. We do caution that while this model does well for this fishery, further research will be needed to assess how well the SIMEX model performs with pelagic fisheries. Given that the surfclam grows very slowly and does not move, modeling the growth for this fishery is very different from other fisheries. In future research we hope to prove that this method can be extended to data from many different types of fisheries.

Chapter 4 revisits the stochastic frontier model, and explores the advantages of Bayesian methods for estimation. In this paper we specify a time-varying stochastic frontier model following (Battese and Coelli 1992) and estimate both a Bayesian and maximum likelihood model. The motivation for the time-varying efficiency model comes from the data, which is a panel data set from years 2001-2009. The surfclam fleet, which has been consolidating since 1990, experienced a "market crisis" in 2005 which caused much of the fleet to exit the fishery. Using data from the mid-Atlantic surfclam fleet between 2001 and 2009, and we look at the impact of fleet consolidation on estimates of technical efficiency and marginal productivity. To estimate the Bayesian model we run the MCMC using a program called JAGS (Just Another Gibbs Sampler) (Plummer et al. 2003). JAGS estimates posterior densities of the stochastic frontier model in R (R Core Team 2013) through the package R2jags (Su and Yajima 2013).

We estimate this model for periods before, during and after 2005. By specifying a time-varying technical efficiency term in the model, we capture changes in the industrial organization of the industry throughout the nine year time period. There is strong evidence that marginal productivities for time fishing and length are decreasing, in spite of the decline in active vessels. This is most likely being caused by declining LPUE and climate change. Both the Bayesian and maximum likelihood model estimates are consistent with a significant drop in technical efficiency following the 2005 fleet restructuring. These results are consistent with **?** who find that technical efficiency is likely decreasing in the industry due to spatial changes in the biomass.

As these three chapters show, fishery managers face many difficulties in setting effective policy. When using economic models to inform these decisions, managers should be aware that issues such as measurement error can give misleading results. Additionally, changes in the industrial organization of fisheries make it a challenge to maximize social welfare while preserving the resource. These issues will likely grow in importance as many developing nations move towards managing their open-access fisheries to prevent over-fishing.

APPENDIX A

CHAPTER TWO APPENDIX

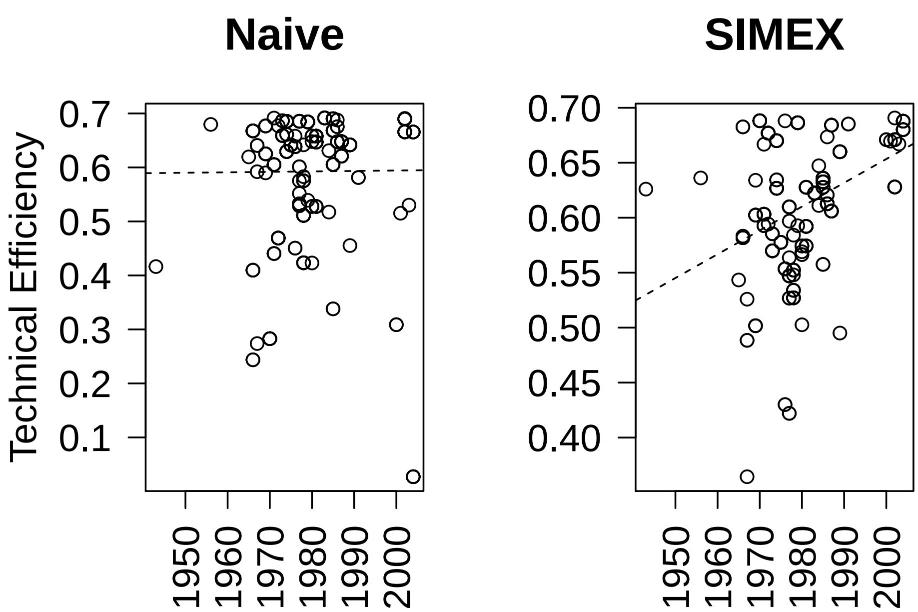

Figure A.1: Technical Efficiency vs. Year Built

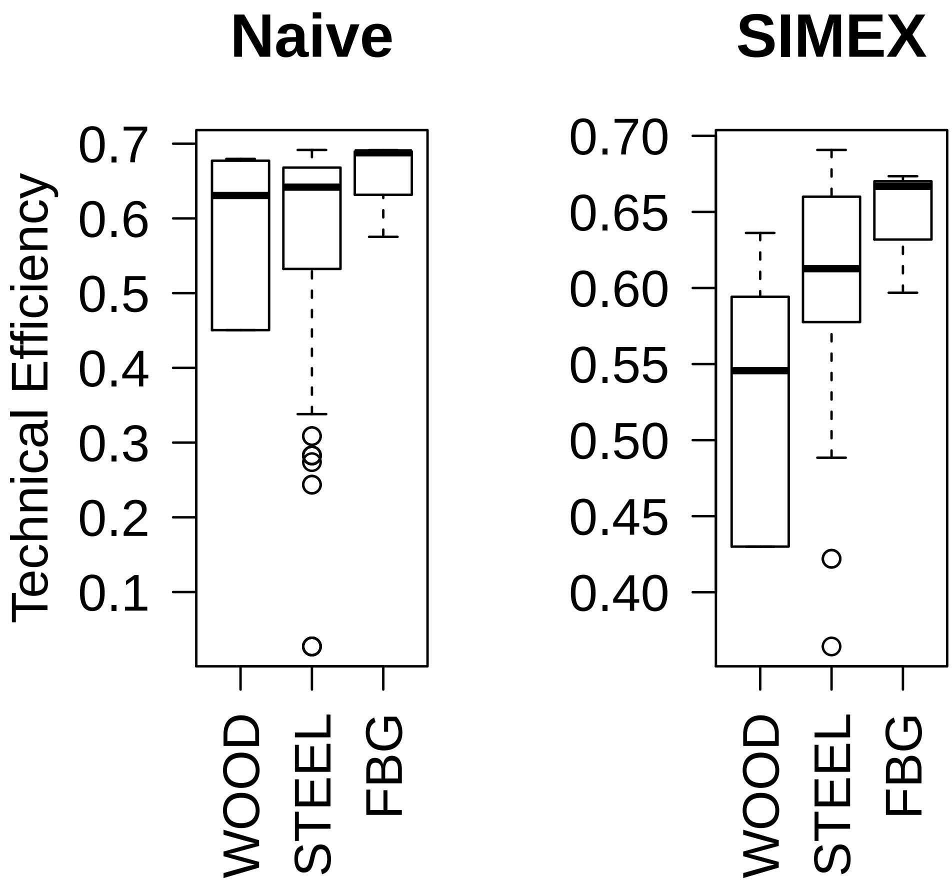

Figure A.2: Technical Efficiency vs. Hull Material

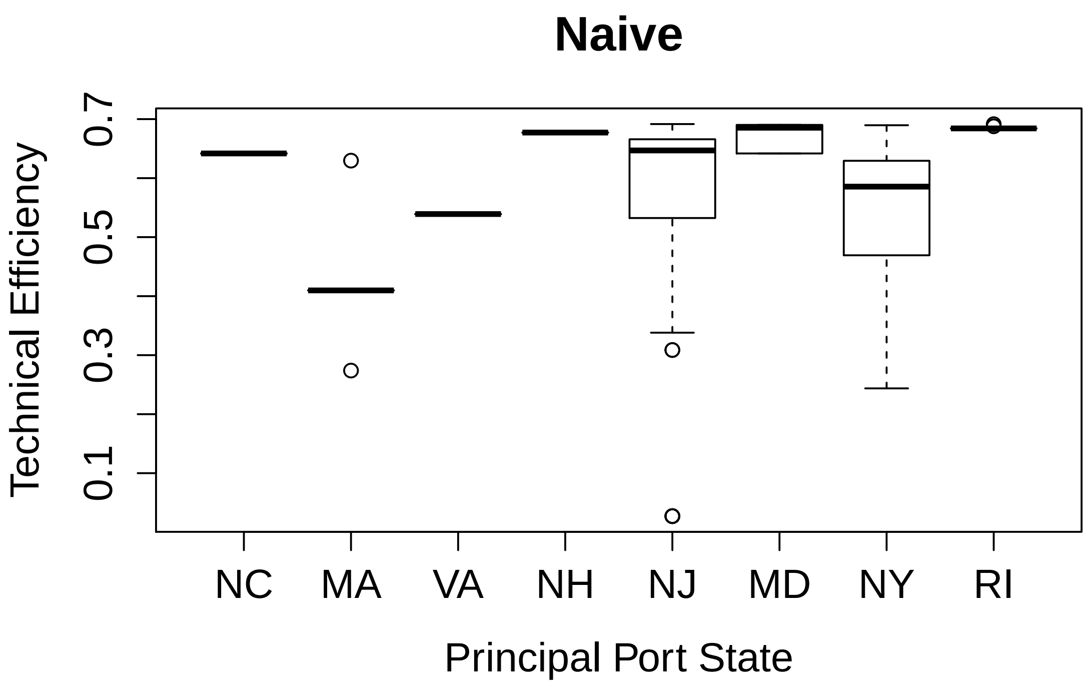

Figure A.3: Technical Efficiency by Region

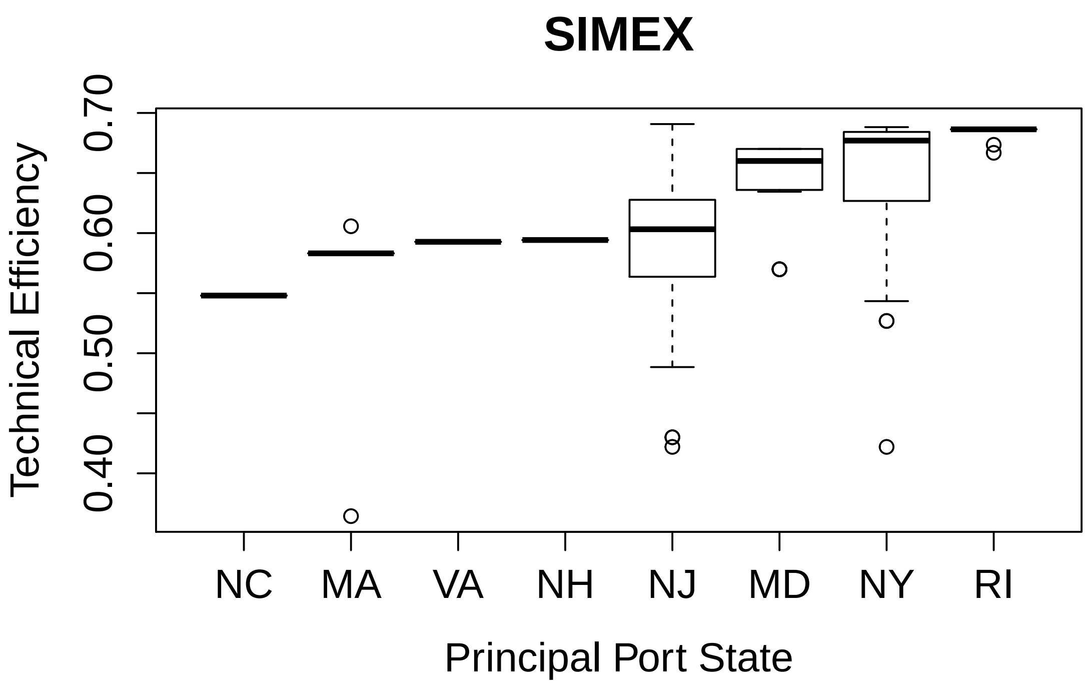

Figure A.4: Technical Efficiency by Region

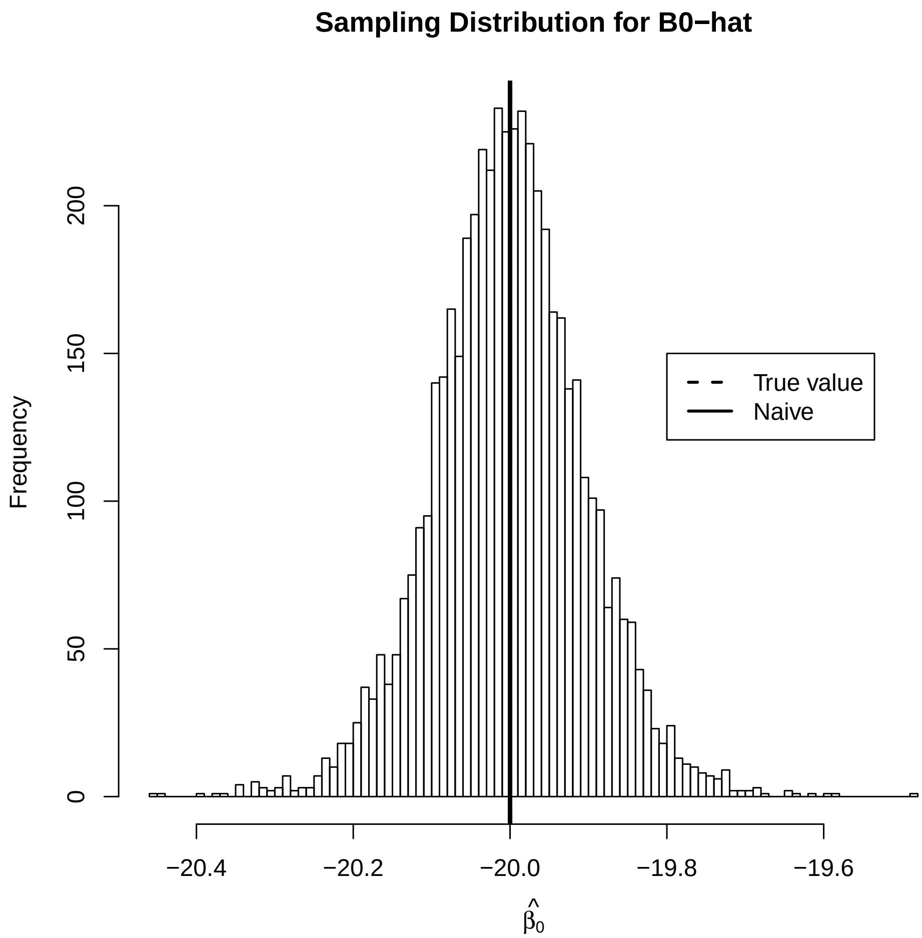

Figure A.5: Simulation Results for $\hat{\beta}_0$

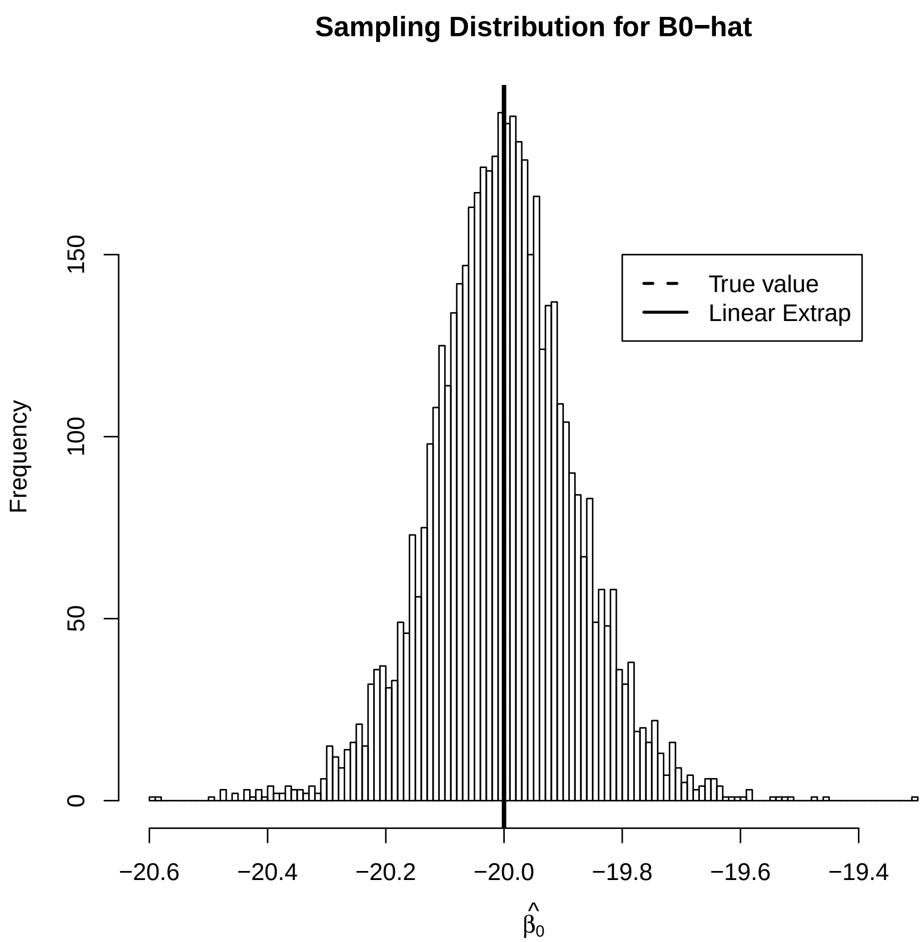

Figure A.6: Simulation Results for $\hat{\beta}_0$

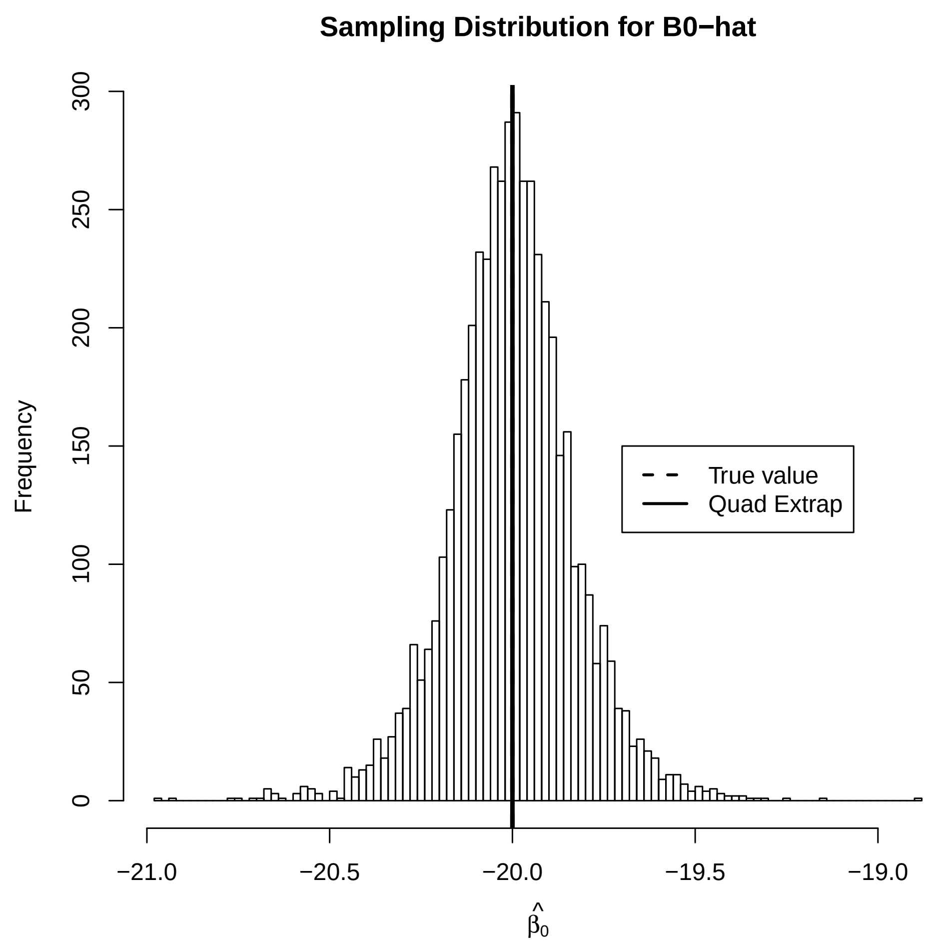

Figure A.7: Simulation Results for $\hat{\beta}_0$

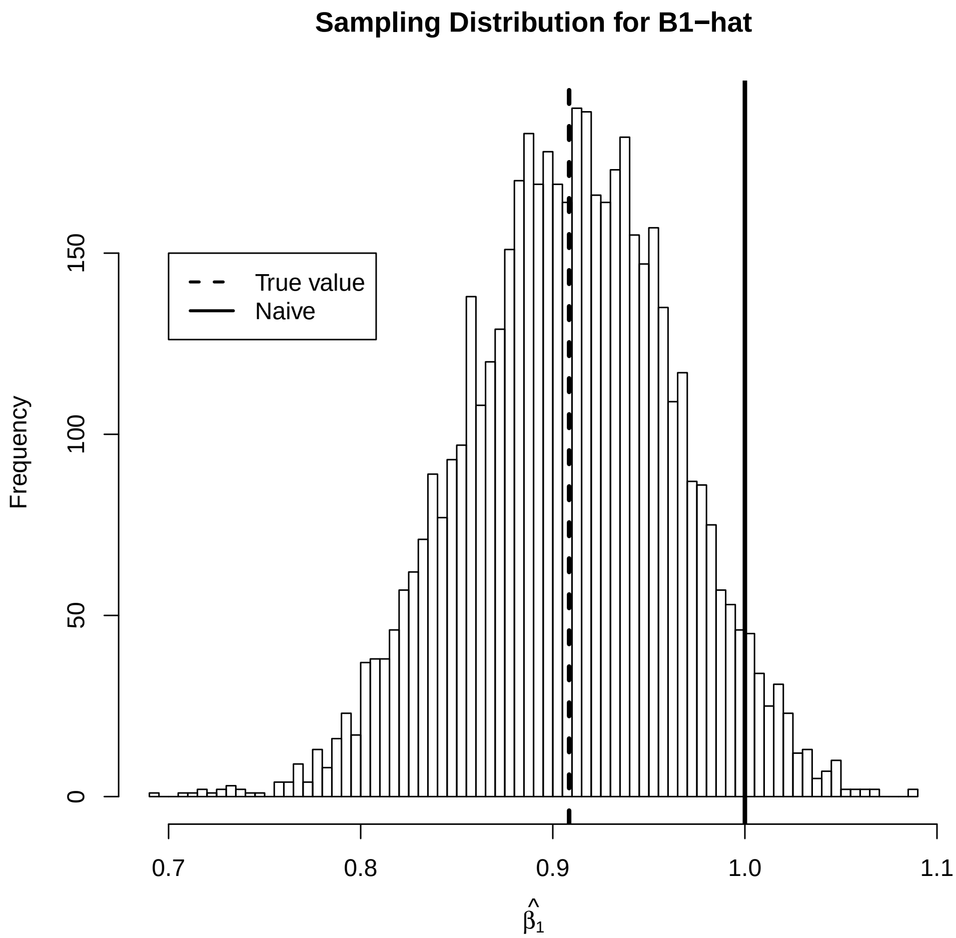

Figure A.8: Simulation Results for $\hat{\beta}_1$

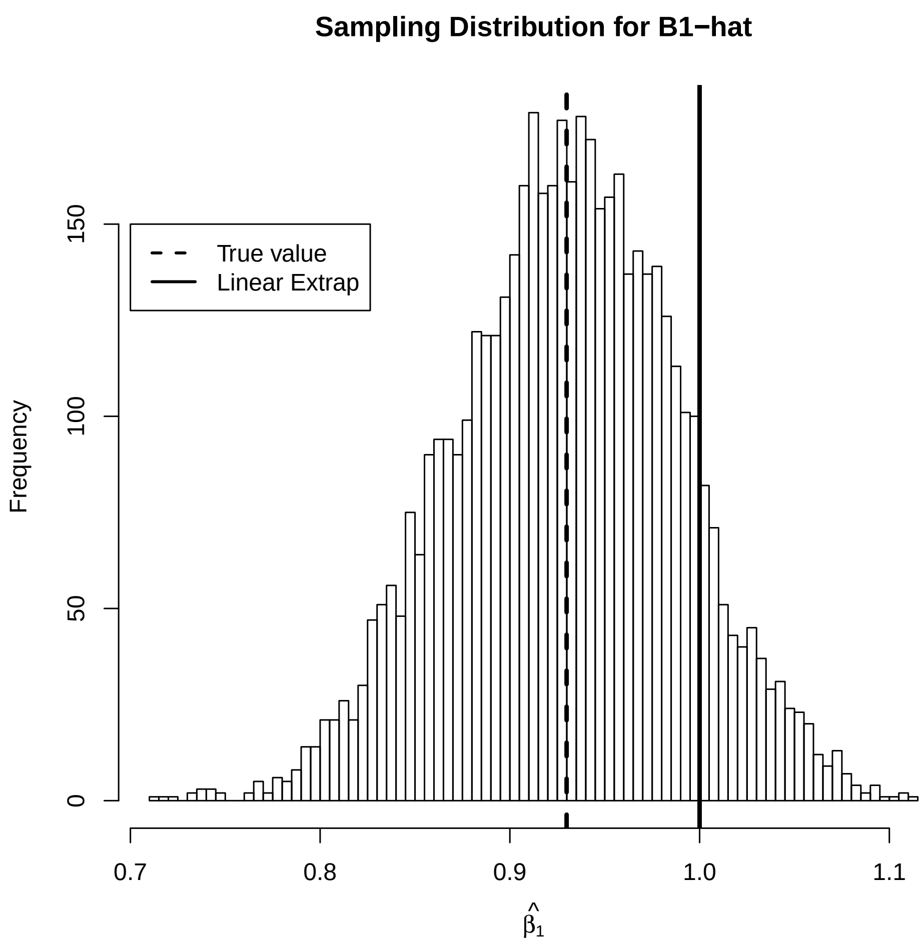

Figure A.9: Simulation Results for $\hat{\beta}_1$

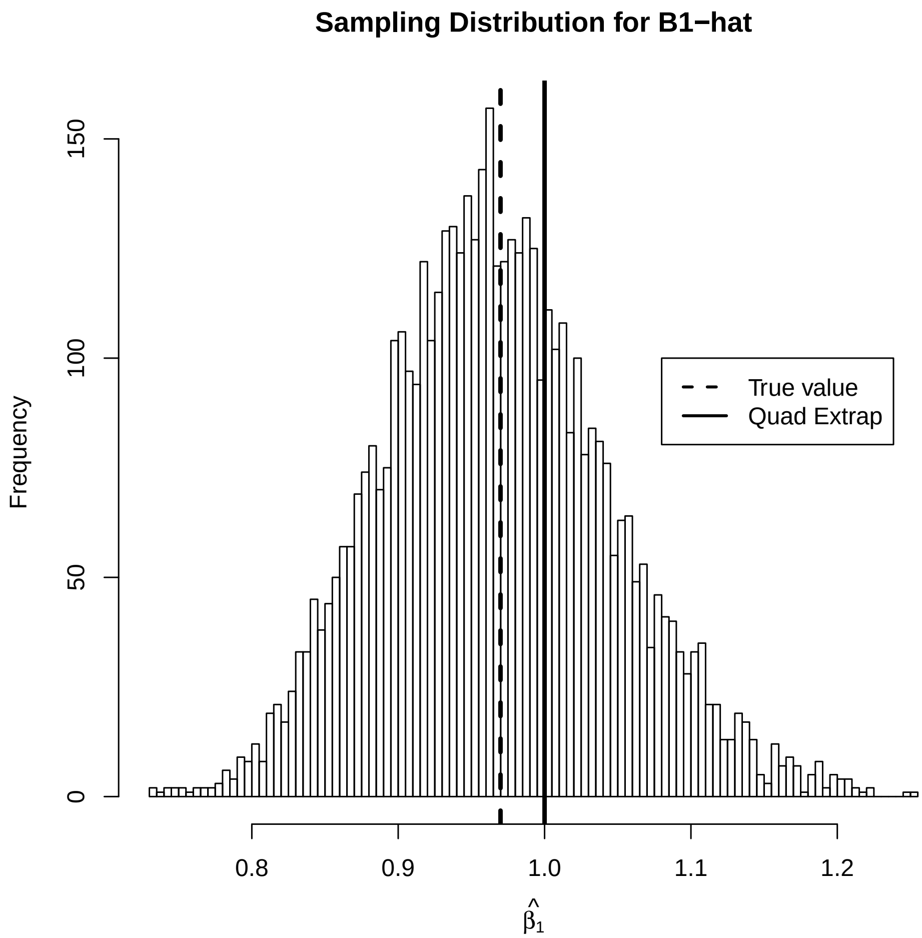

Figure A.10: Simulation Results for $\hat{\beta}_1$

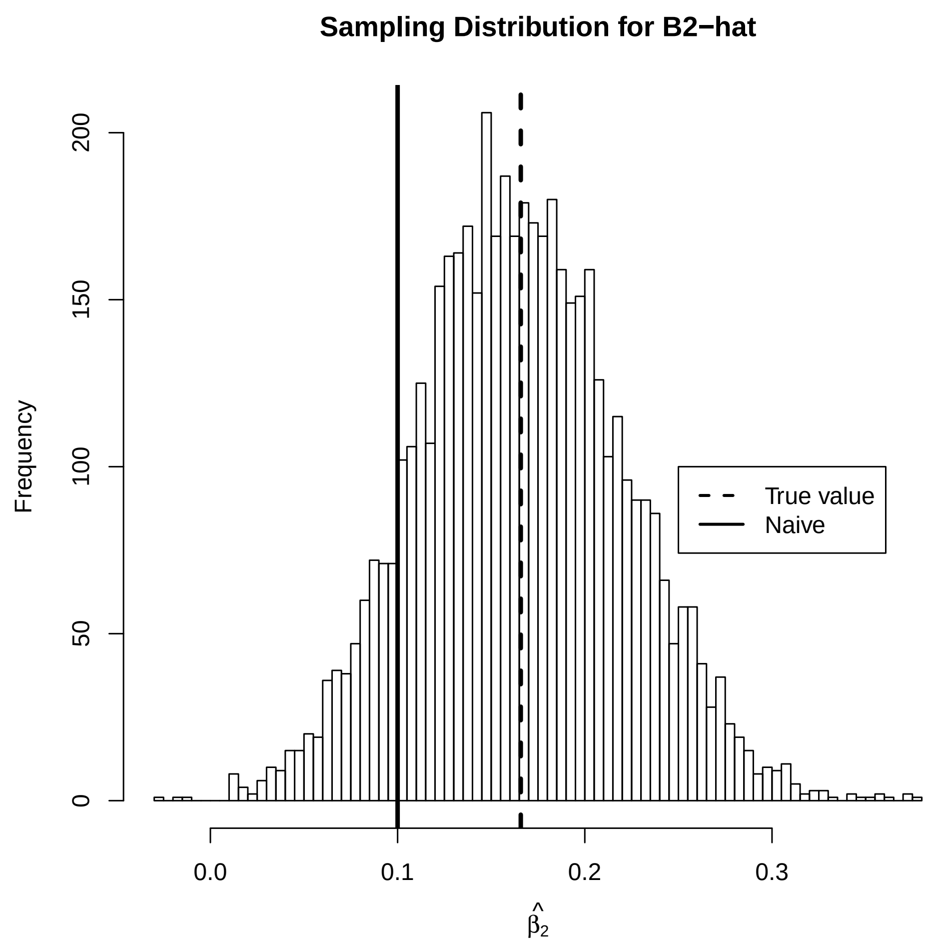

Figure A.11: Simulation Results for $\hat{\beta}_2$

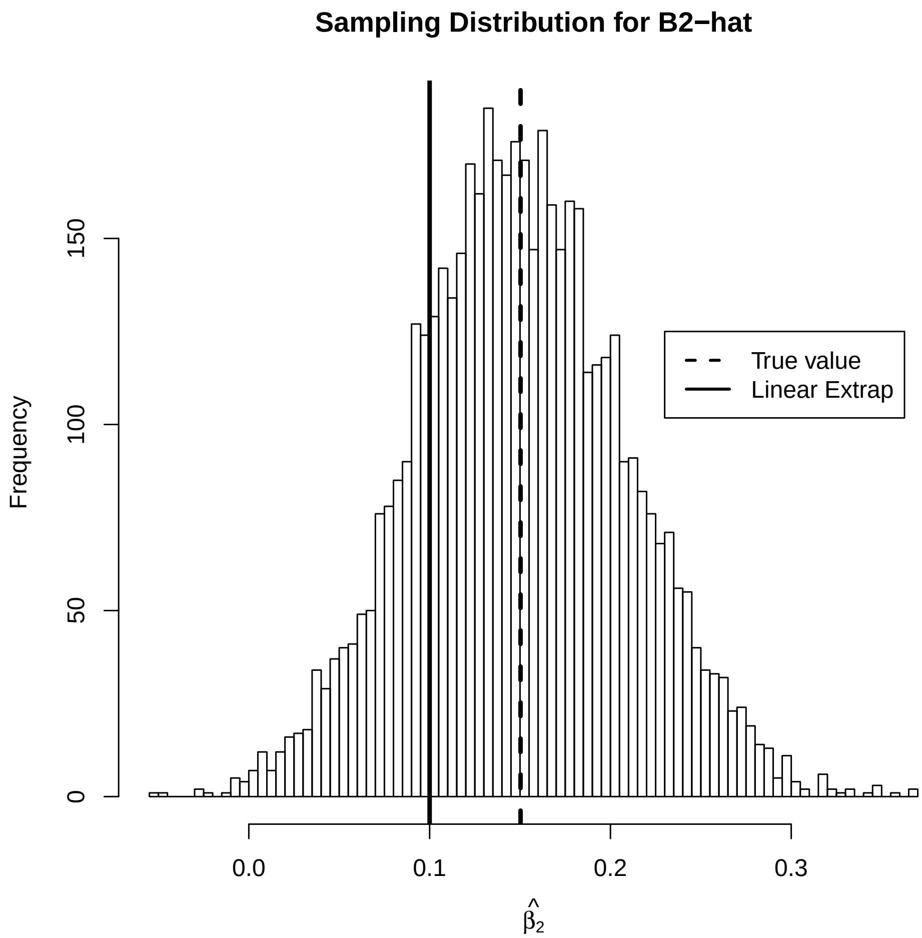

Figure A.12: Simulation Results for $\hat{\beta}_2$

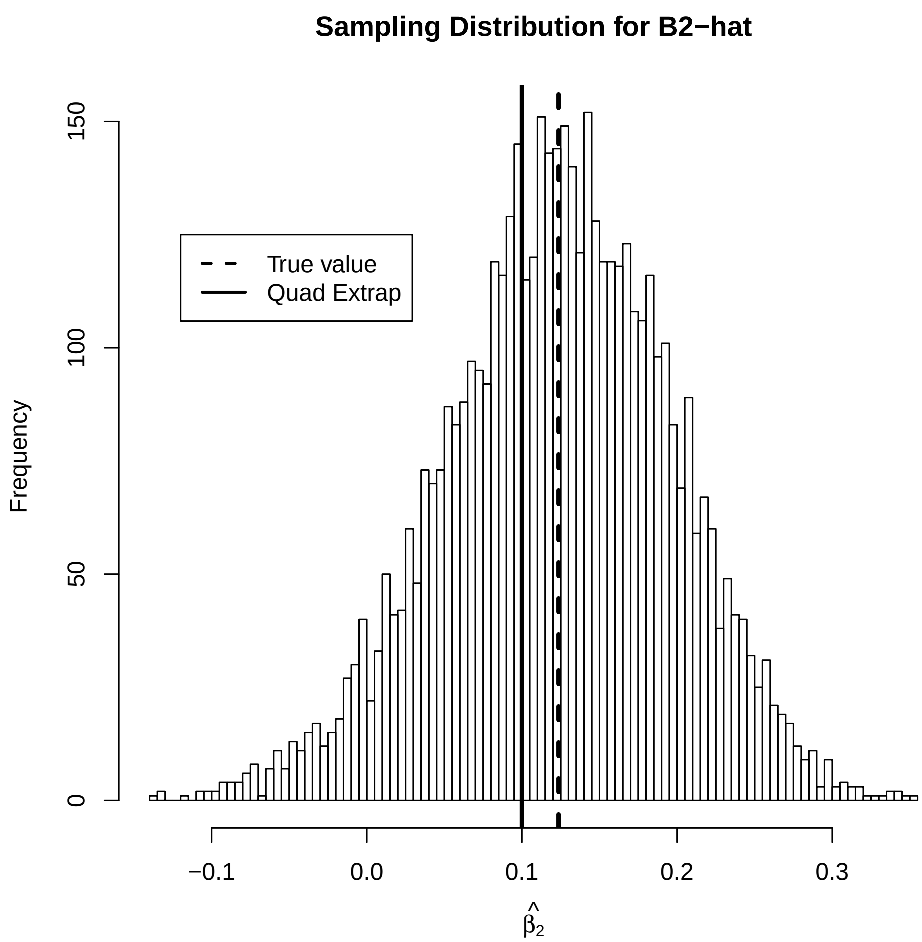

Figure A.13: Simulation Results for $\hat{\beta}_2$

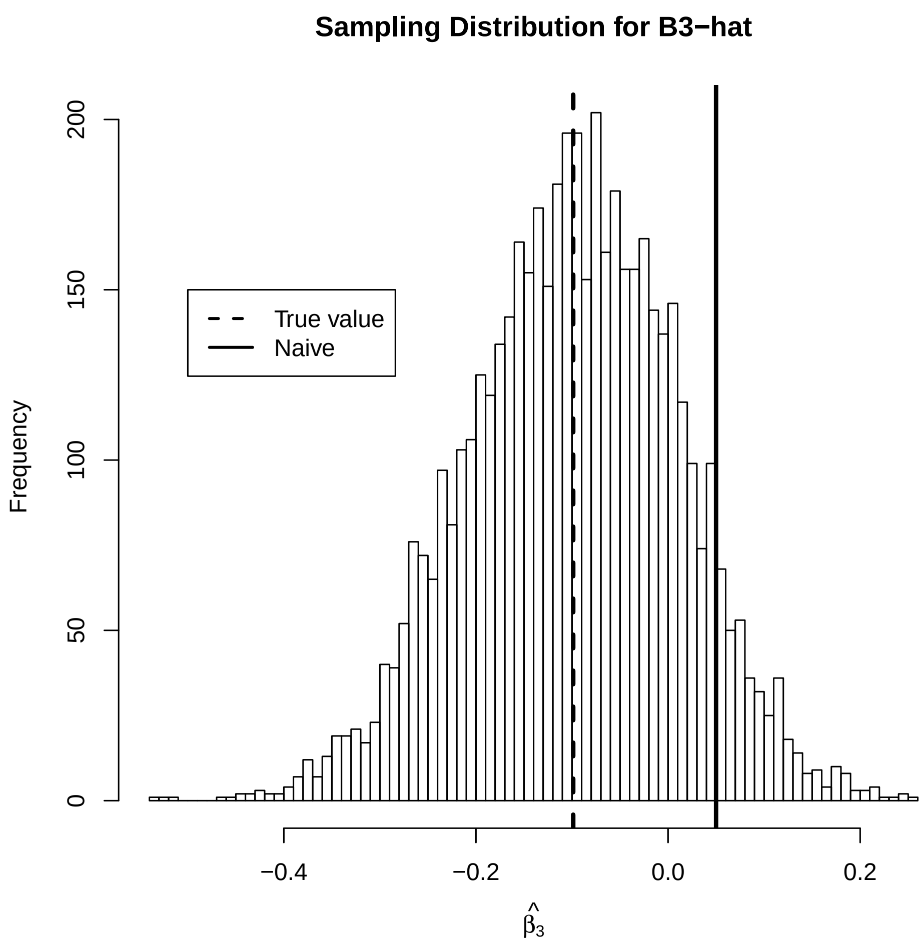

Figure A.14: Simulation Results for $\hat{\beta}_3$

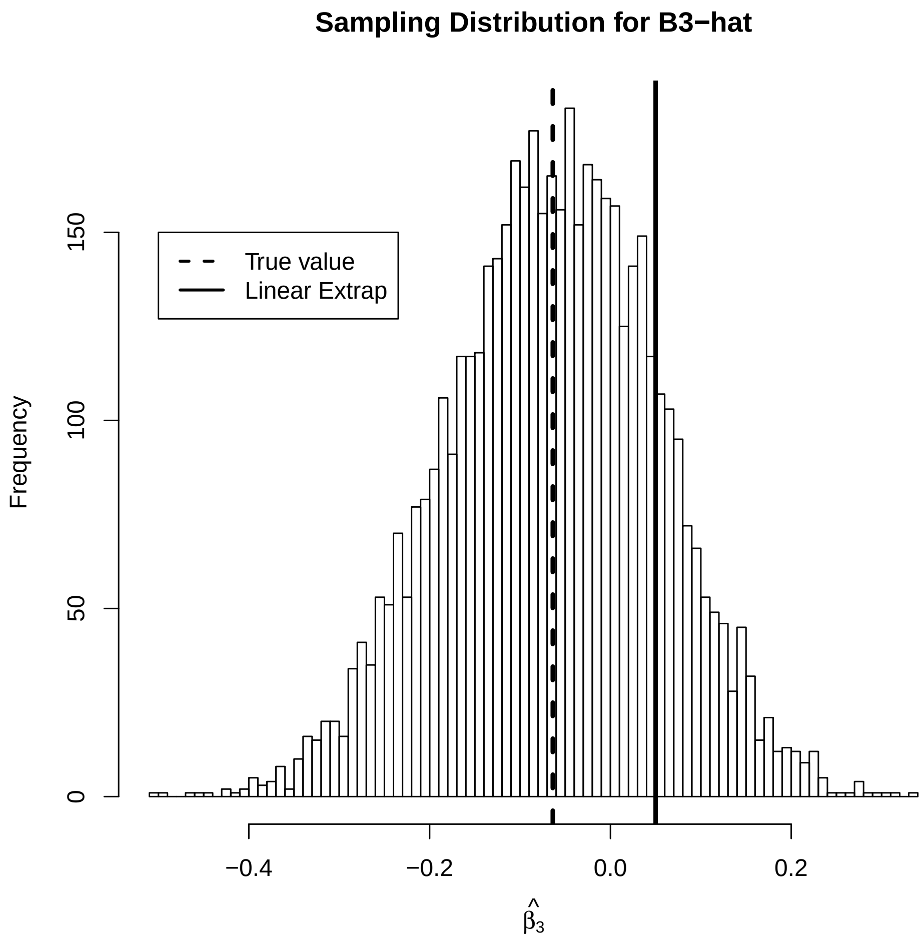

Figure A.15: Simulation Results for $\hat{\beta}_3$

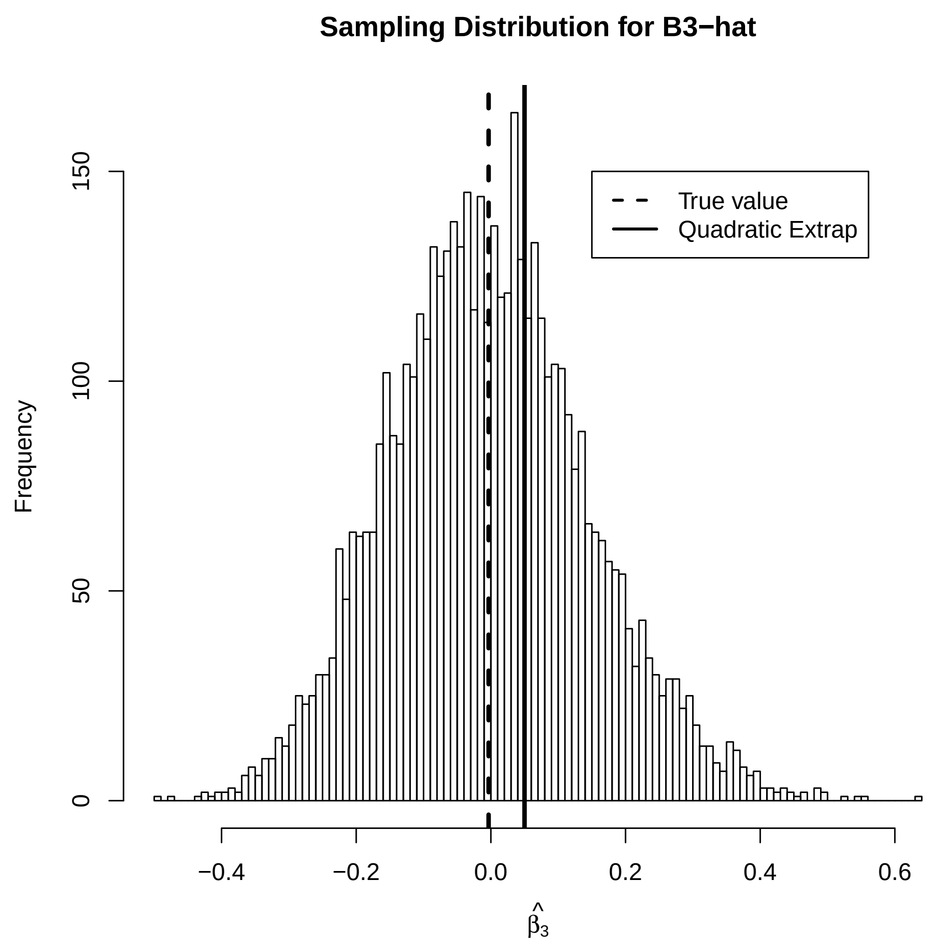

Figure A.16: Simulation Results for $\hat{\beta}_3$

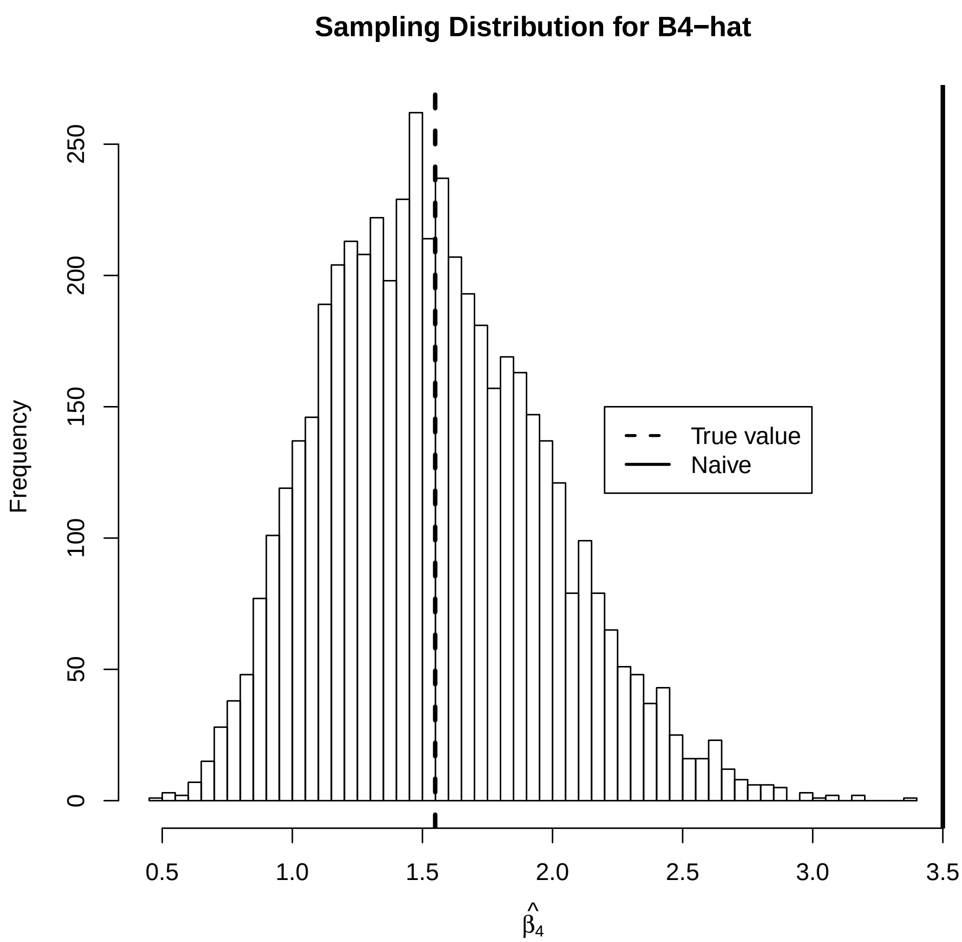

Figure A.17: Simulation Results for $\hat{\beta}_4$

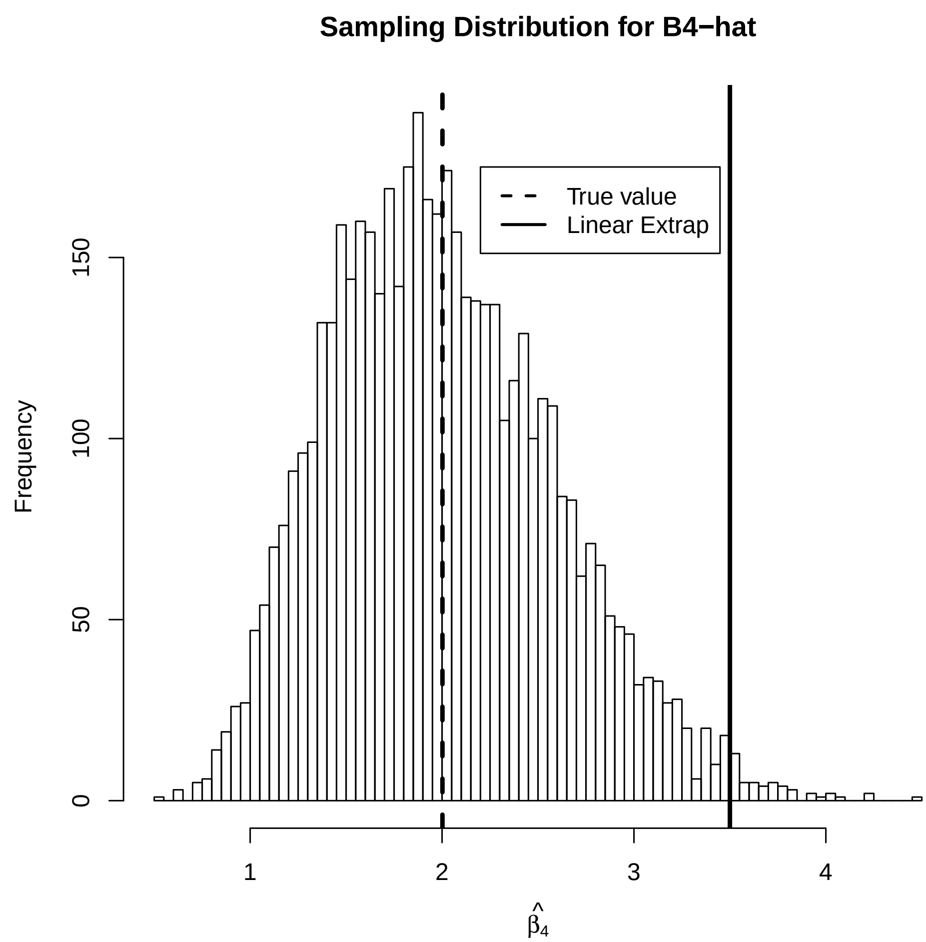

Figure A.18: Simulation Results for $\hat{\beta}_4$

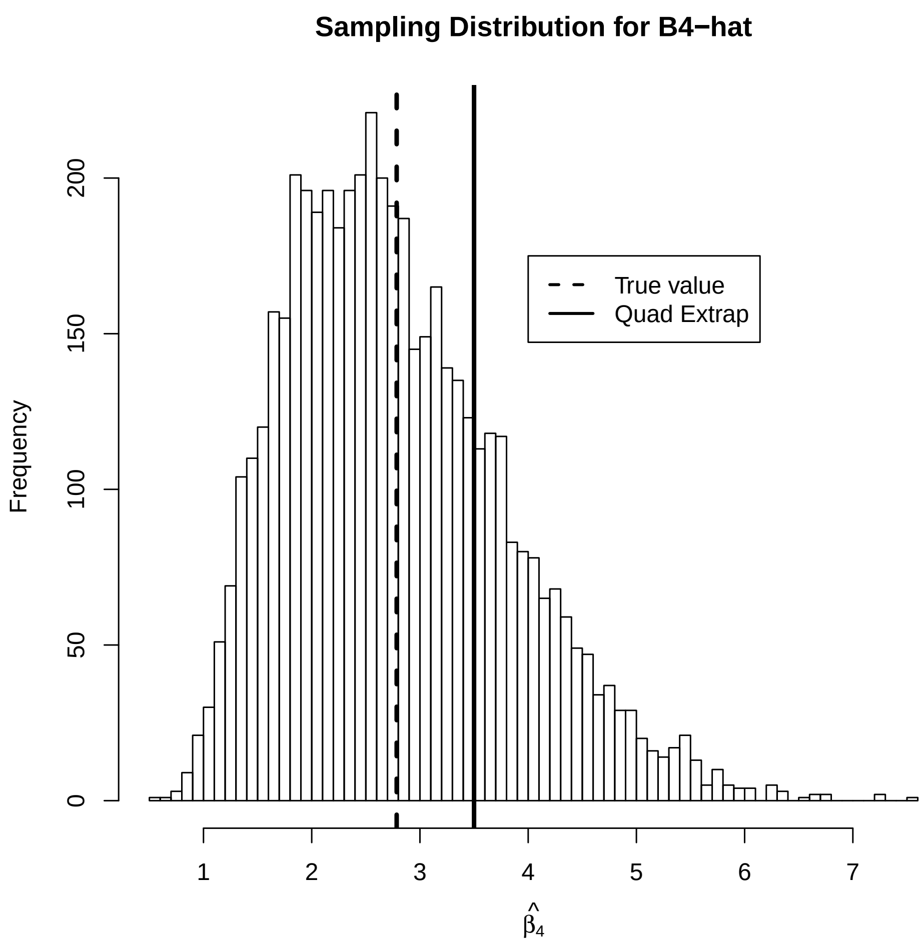

Figure A.19: Simulation Results for $\hat{\beta}_4$

APPENDIX B

CHAPTER THREE APPENDIX

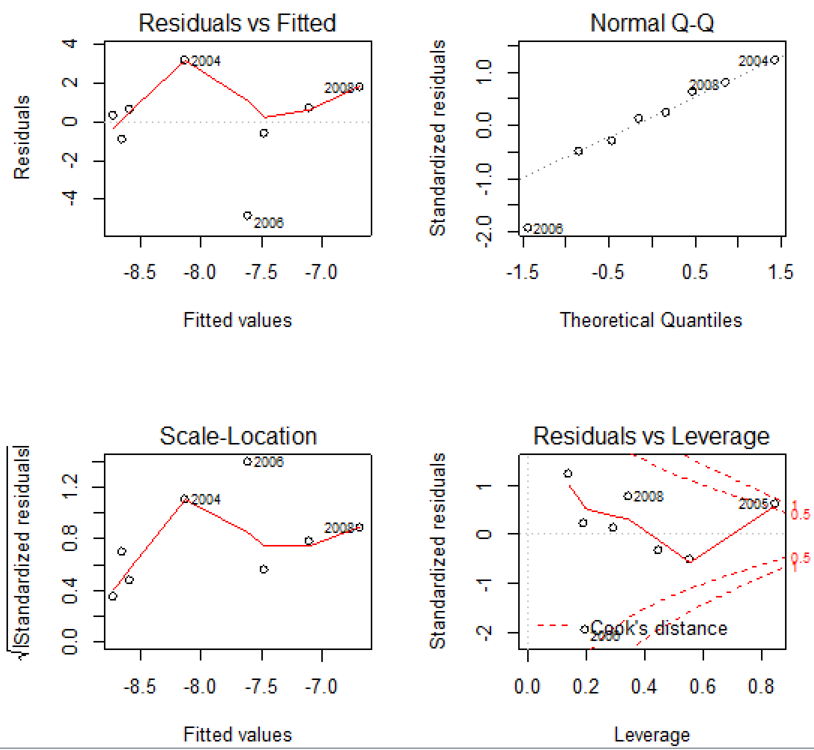

Figure B.1: Residuals Plot for Classic Schaefer Growth Model

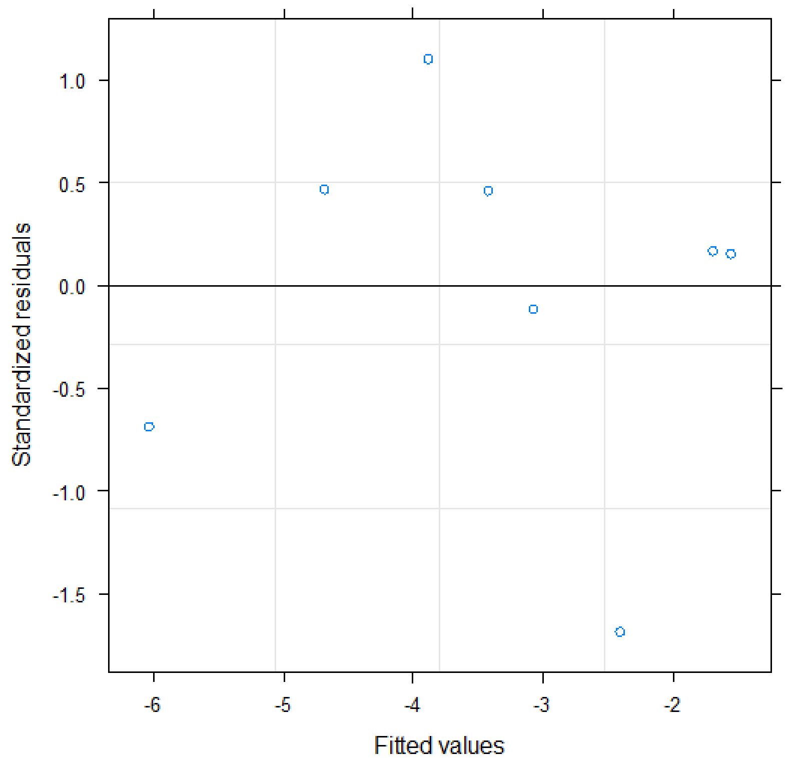

Figure B.2: Residuals Plot for Generalized Schaefer Growth Model

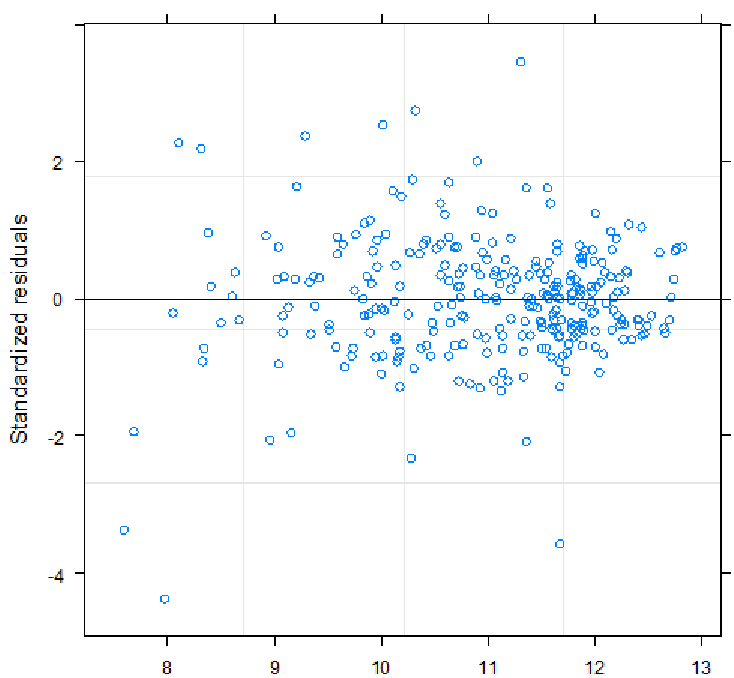

Figure B.3: Residuals Plot for Generalized Schaefer Production Function

APPENDIX C

CHAPTER FOUR APPENDIX

Summary plots for all four Bayesian Models, 2001-2009, 2001-2004, 2005-2009 and 2006-2009.

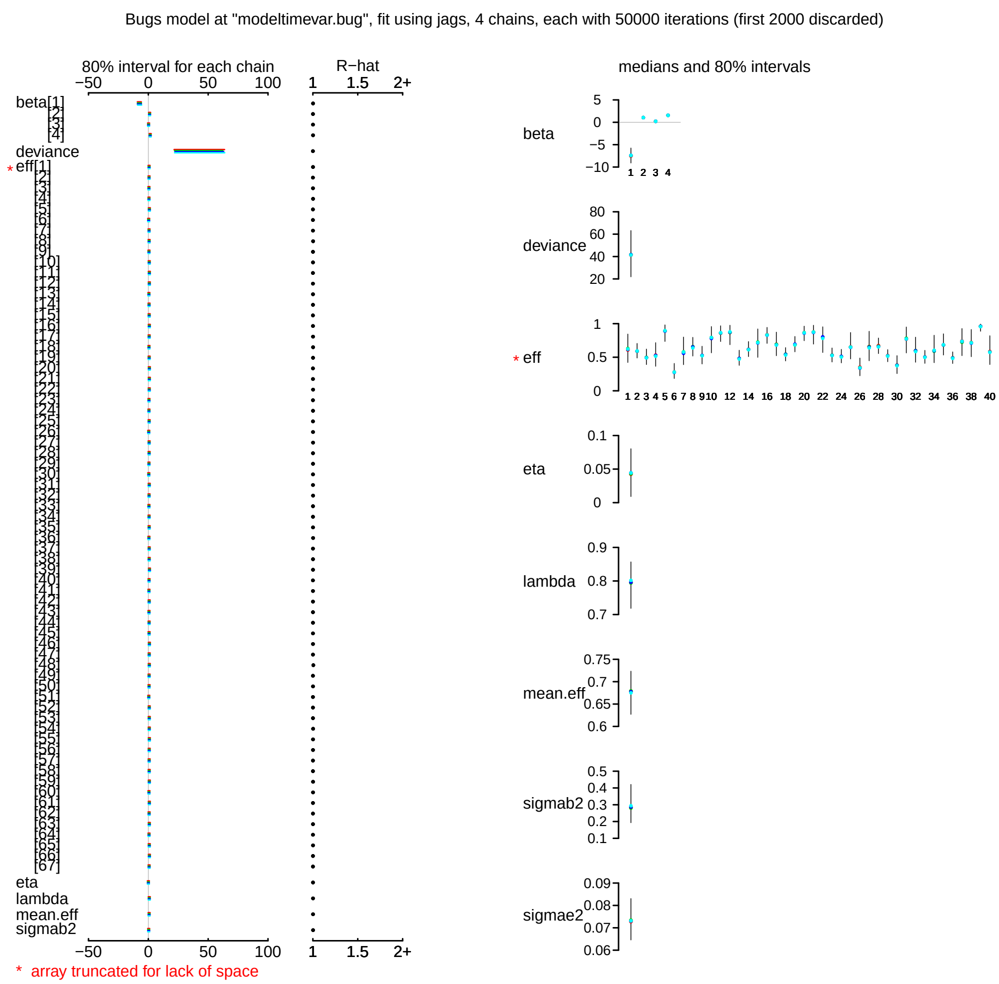

Figure C.1: Summary Plot 2001-2009

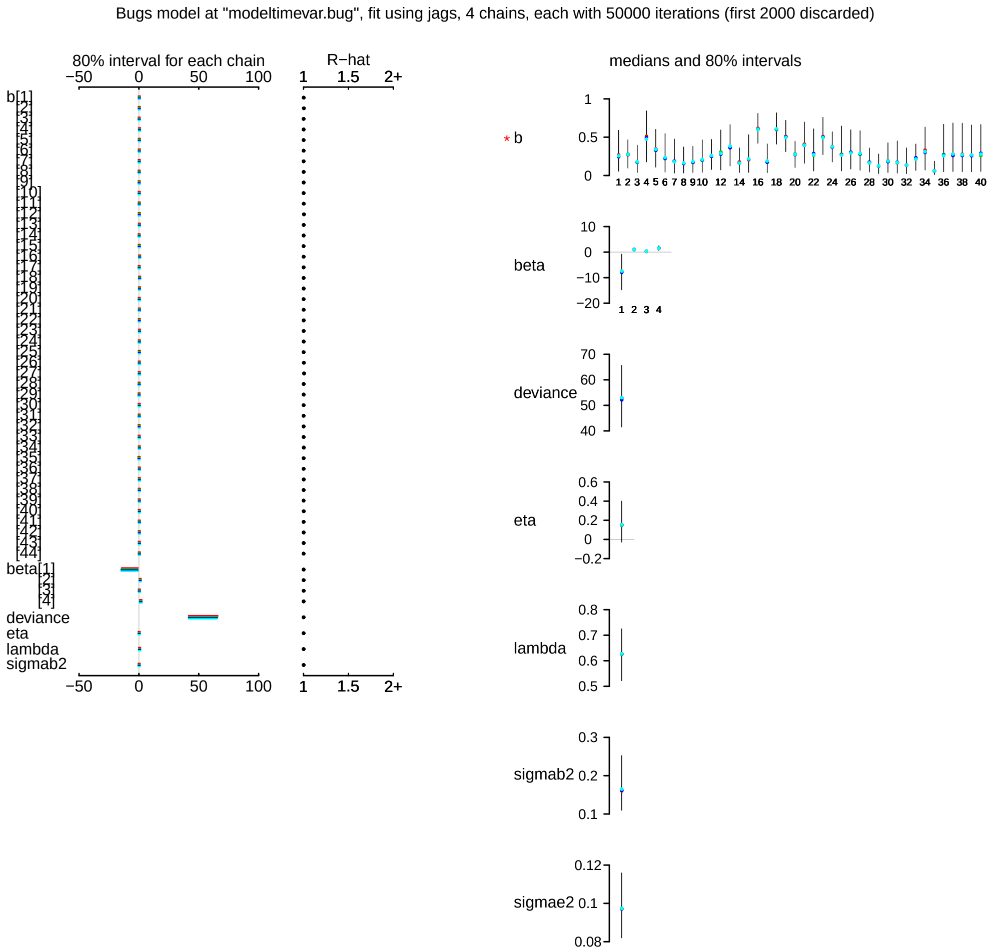

Figure C.2: Summary Plot 2001-2004

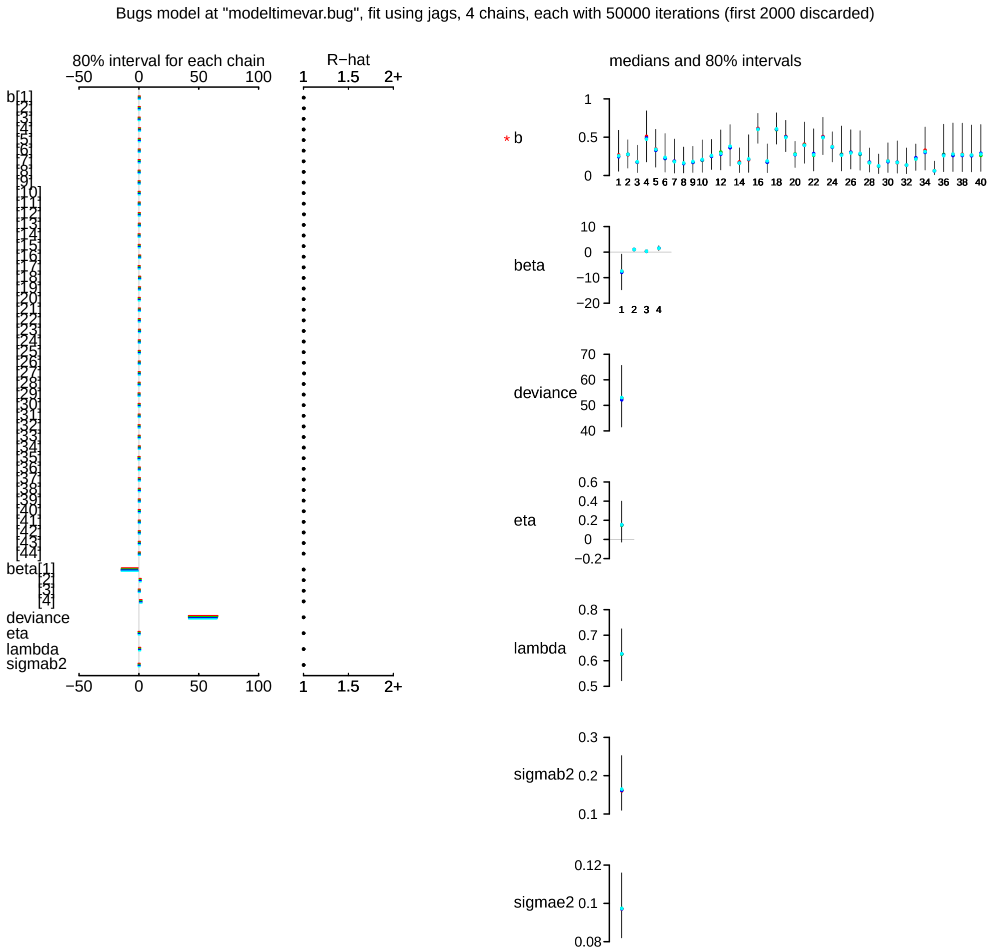

Figure C.3: Summary Plot 2005-2009

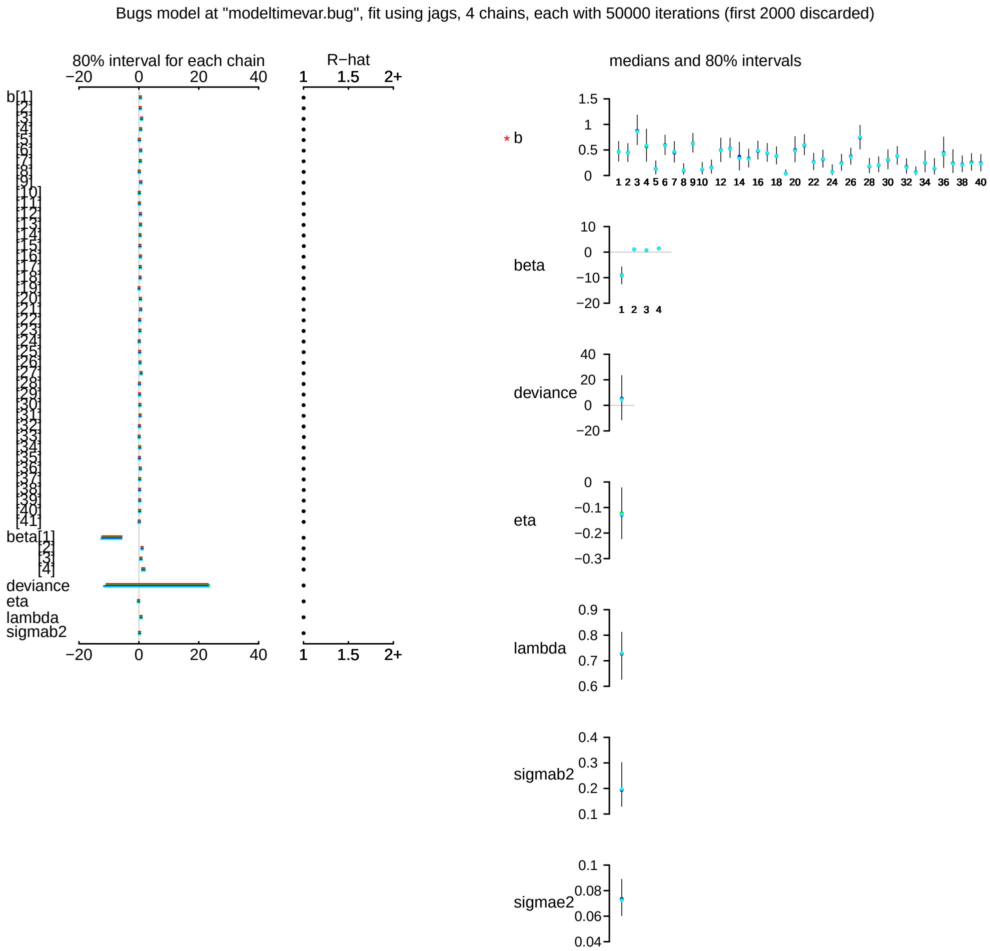

Figure C.4: Summary Plot 2005-2009

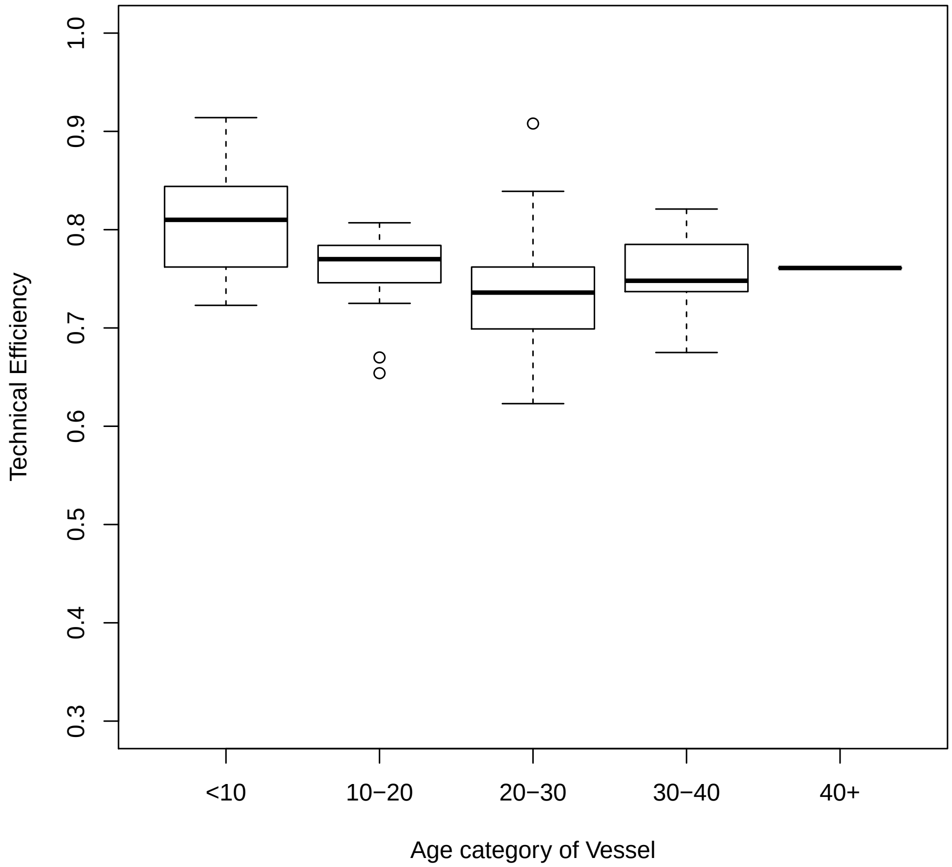

Figure C.5: Technical Efficiency by Age 2001-2004

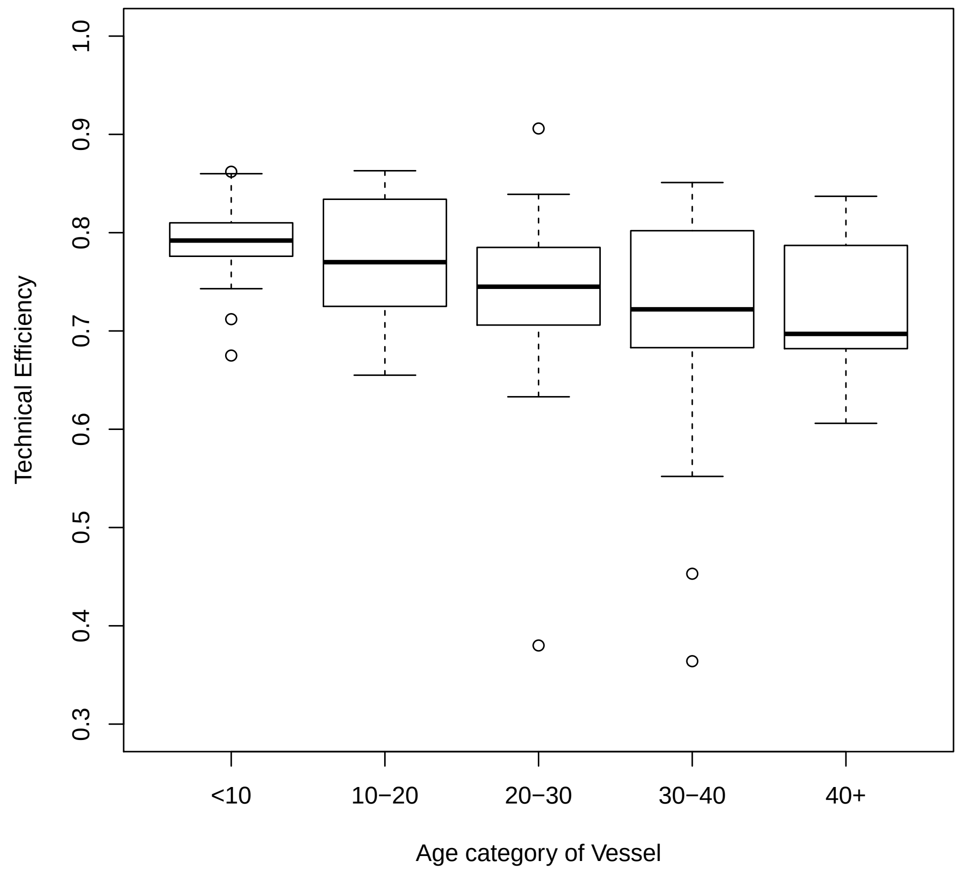

Figure C.6: Technical Efficiency by Age 2005-2009

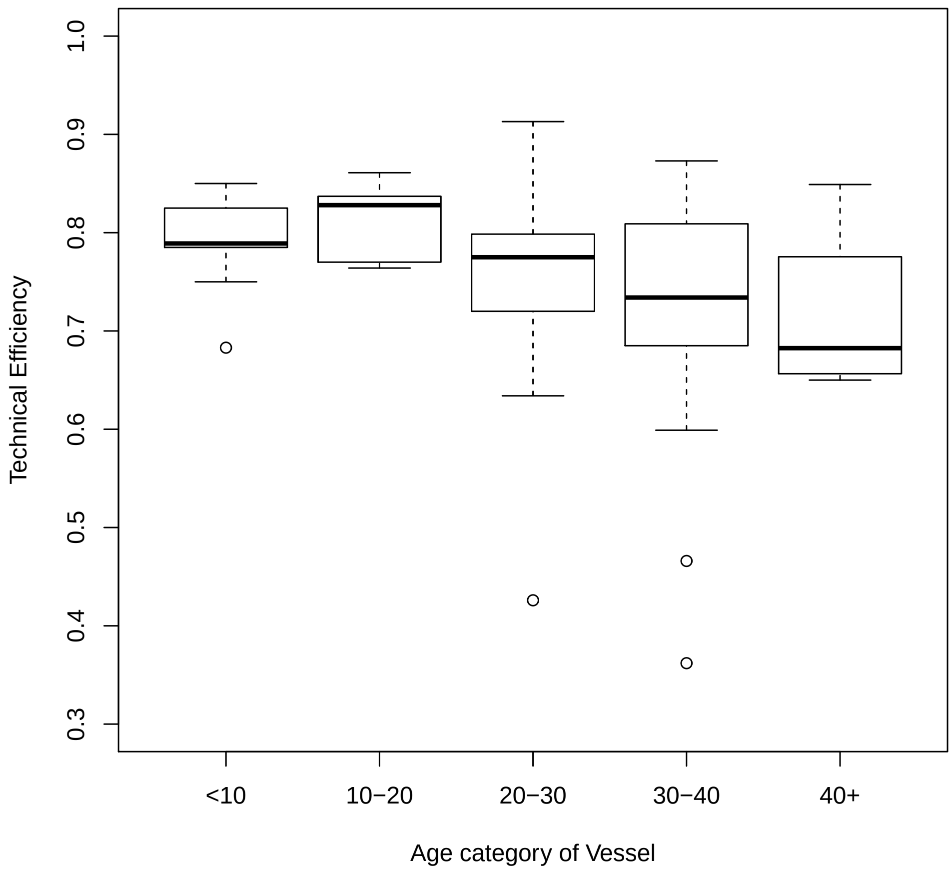

Figure C.7: Technical Efficiency by Age 2006-2009

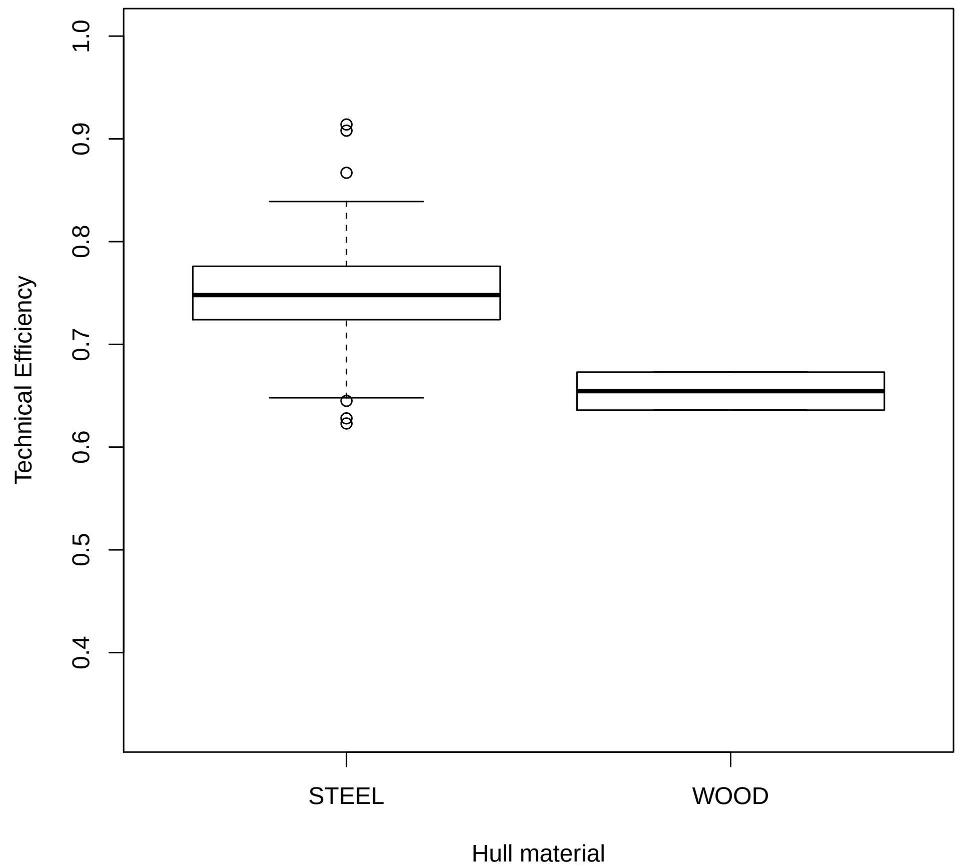

Figure C.8: Technical Efficiency by Hull Material 2001-2004

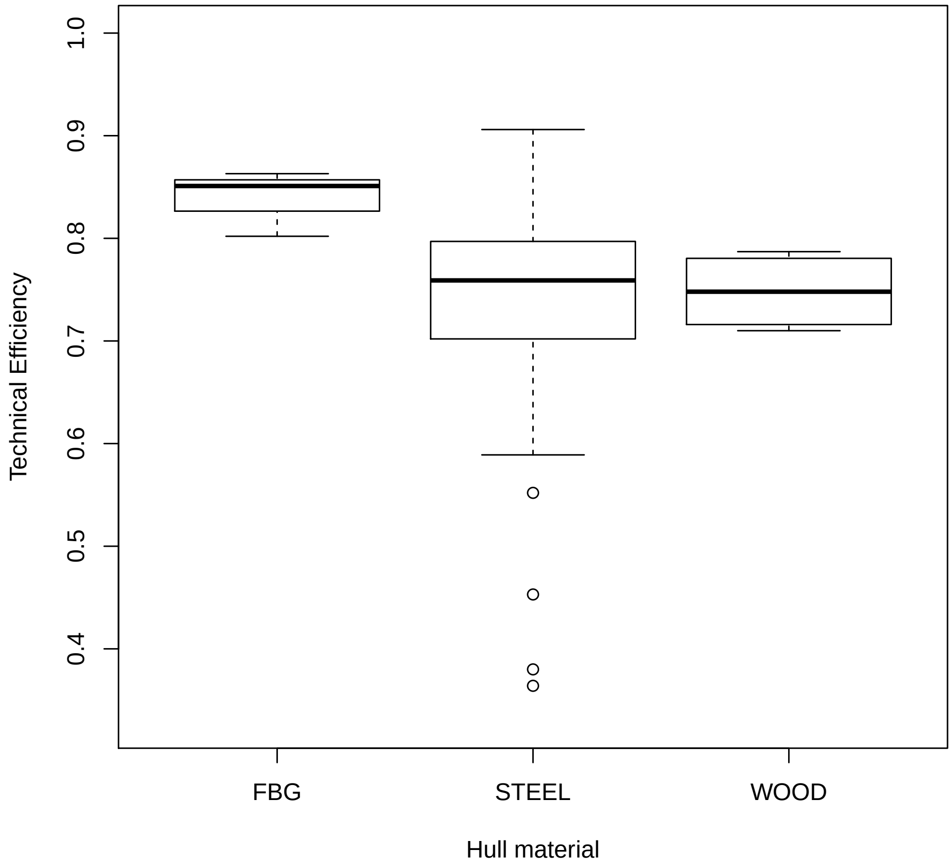

Figure C.9: Technical Efficiency by Hull Material 2005-2009

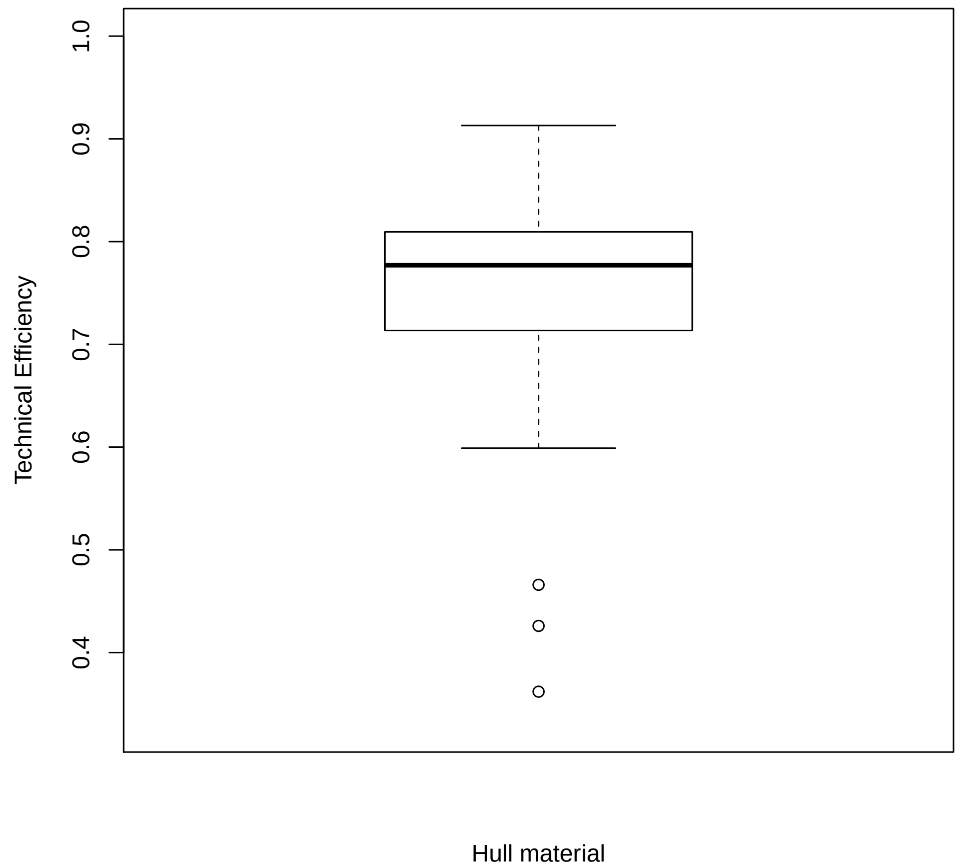

Figure C.10: Technical Efficiency by Hull Material 2006-2009

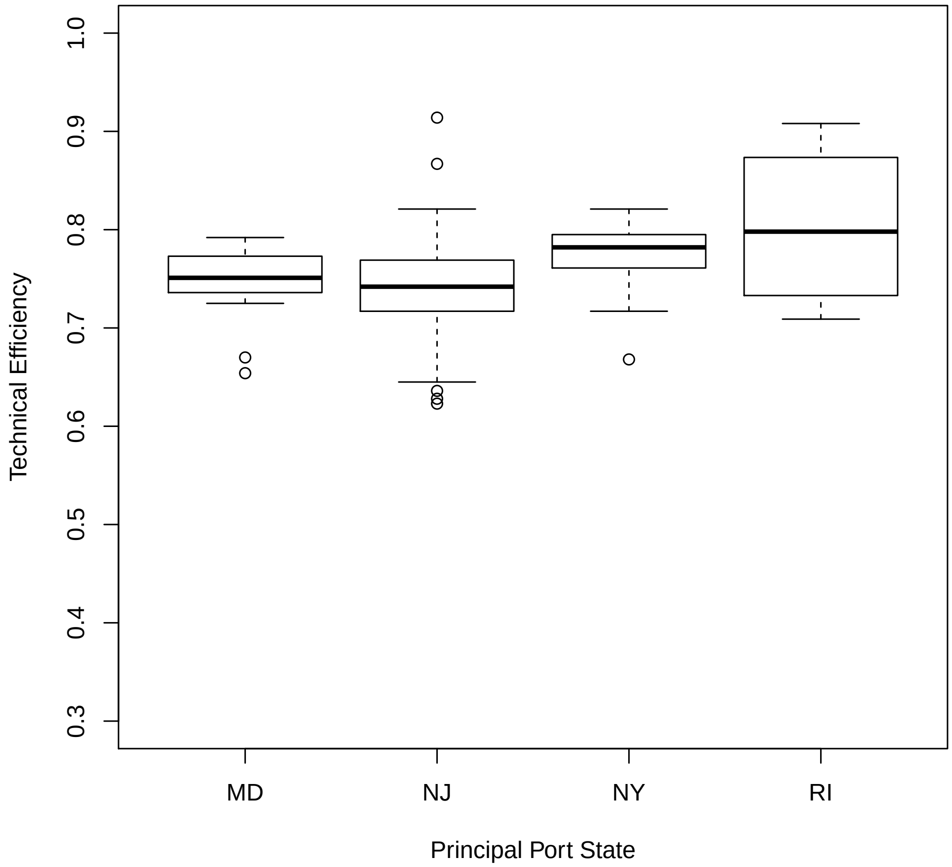

Figure C.11: Technical Efficiency by Home Port State 2001-2004

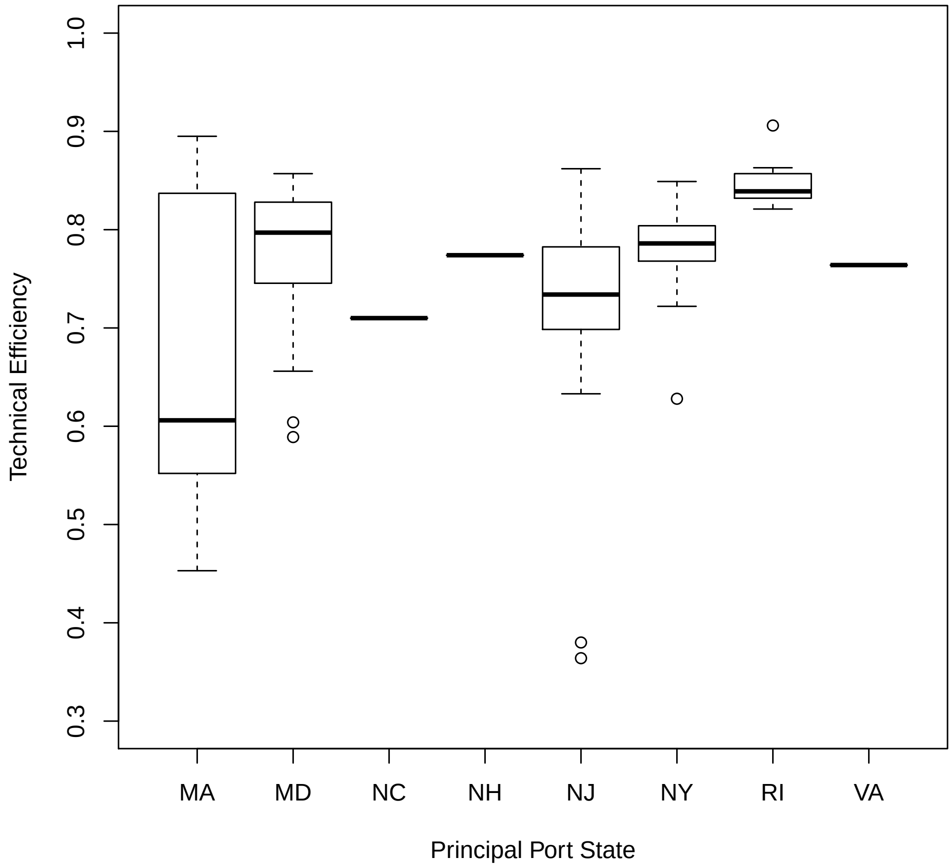

Figure C.12: Technical Efficiency by Home Port State 2005-2009

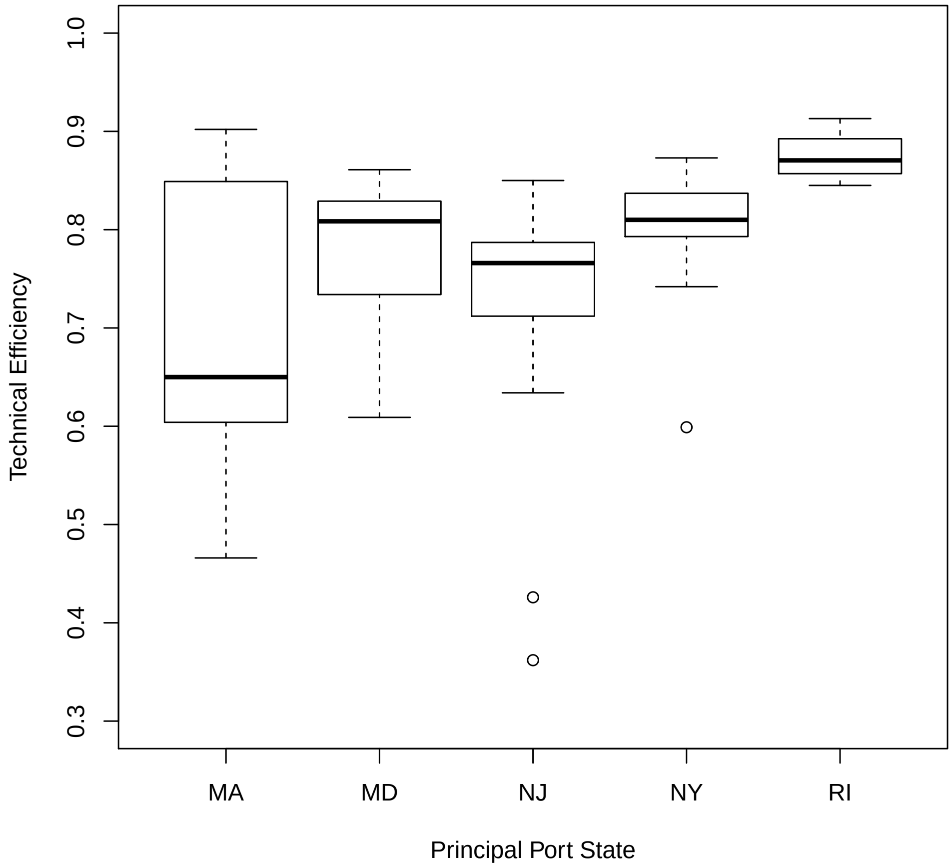

Figure C.13: Technical Efficiency by Home Port State 2006-2009

BIBLIOGRAPHY

Acheson, J. M. (1988). *The lobster gangs of Maine*. Upne.

Aigner, D., C. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6*(1), 21–37.

Battese, G. E. and T. J. Coelli (1992, June). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis 3*(1-2), 153–169.

Battese, G. E. and T. J. Coelli (1995, June). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics 20*(2), 325–332.

Brandt, S. (2007, May). Evaluating tradable property rights for natural resources: The role of strategic entry and exit. *Journal of Economic Behavior & Organization 63*(1), 158–176.

Buonaccorsi, J. (2010). *Measurement error : models, methods, and applications*. Boca Raton: CRC Press.

Butterworth, D. and P. Andrew (1984). Dynamic catch-effort models for the hake stocks in icseaf divisions 1.3-2.2. *Coli. Scient. Pap. Int. Comm. SE Atl. Fish 11*, 29–58.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2012). *Measurement error in nonlinear models: a modern perspective*. CRC press.

Casey, K. E., C. M. Dewees, B. R. Turris, and J. E. Wilen (1995). The effects of individual vessel quotas in the british columbia halibut fishery. *Marine Resource Economics 10*(3).

Cochrane, K. L. (2002). *A fishery manager's guidebook: management measures and their application*. Number 424. FAO.

Coelli, T. and A. Henningsen (2013). *frontier: Stochastic Frontier Analysis*. R package version 1.1-0.

Cook, J. R. and L. A. Stefanski (1994, December). Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association 89*(428), 1314–1328.

Ehlers, R. S. (2011). Comparison of bayesian models for production efficiency. *Journal of Applied Statistics 38*(11), 2433–2443.

Felthoven, R. G. (2002). Effects of the american fisheries act on capacity, utilization and technical efficiency. *Marine Resource Economics 17*(3).

Fernandez, C., J. Osiewalski, and M. F. Steel (1997). On the use of panel data in stochastic frontier models with improper priors. *Journal of Econometrics 79*(1), 169–193.

Fox, W. W. (1975). Fitting the generalized stock production model by least-squares and equilibrium approximation. *Fish. Bull 73*(1), 23–37.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. CRC press.

Grafton, R. (2006). *Economics for fisheries management*. Aldershot England ;;Burlington VT: Ashgate Pub.

Grafton, R. Q., D. Squires, and K. J. Fox (2000). Private Property and Economic Efficiency: A Study of Common-Pool Resource. *Journal of Law & Economics 43*.

Greene, W. H. (1990). A Gamma-distributed stochastic frontier model. *Journal of Econometrics 46*(1), 141–163.

Greene, W. H. (2003). *Econometric Analysis*, Volume 97. Prentice Hall.

Griffin, J. E. and M. F. Steel (2007). Bayesian stochastic frontier analysis using winbugs. *Journal of Productivity Analysis 27*(3), 163–176.

Griliches, Z. and J. A. Hausman (1986). Errors in variables in panel data. *Journal of econometrics 31*(1), 93–118.

Gulland, J. A. (1961). Fishing and the stocks of fish at iceland. *Fishery investigations. Ser. 2 23*(4).

Hausman, J. A. (2001). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives 15*(4), 57–67.

Hilborn, R. (1979). Comparison of fisheries control systems that utilize catch and effort data. *Journal of the Fisheries Board of Canada 36*(12), 1477–1489.

Hilborn, R., T. P. Quinn, D. E. Schindler, and D. E. Rogers (2003). Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences 100*(11), 6564–6568.

Hsiao, C. (2003). *Analysis of panel data*, Volume 34. Cambridge university press.

Jondrow, J., C. Knox Lovell, I. S. Materov, and P. Schmidt (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics 19*(2), 233–238.

Kim, Y. and P. Schmidt (2000). A review and empirical comparison of bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. *Journal of Productivity Analysis 14*(2), 91–118.

Kirkley, J., C. J. Morrison Paul, and D. Squires (2004). Deterministic and stochastic capacity estimation for fishery capacity reduction. *Marine Resource Economics 19*(3).

Kirkley, J., D. Squires, and I. E. Strand (1998, March). Characterizing Managerial Skill and Technical Efficiency in a Fishery. *Journal of Productivity Analysis 9*(2), 145–160.

Kirkley, J. E., D. Squires, and I. E. Strand (1995, August). Assessing Technical Efficiency in Commercial Fisheries: The Mid-Atlantic Sea Scallop Fishery. *American Journal of Agricultural Economics 77*(3), 686–697.

Kumbhakar, S. (2000). *Stochastic frontier analysis*. Cambridge [England] ;New York: Cambridge University Press.

Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics 46*(1), 201–211.

Lee, Y. H. and P. Schmidt (1993). A production frontier model with flexible temporal variation in technical efficiency. *The measurement of productive efficiency: Techniques and applications*, 237–255.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing 10*(4), 325–337.

McCay, B. J., S. Brandt, and C. F. Creed (2011, May). Human dimensions of climate change and fisheries in a coupled system: the Atlantic surfclam case. *ICES Journal of Marine Science 68*(6), 1354–1367.

McCulloch, C. E. and S. R. Searle (2000, December). *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Meeusen, W. and J. Broeck (1977, June). Technical efficiency and dimension of the firm: Some results on the use of frontier production functions. *Empirical Economics 2*(2), 109–122.

Mid-Atlantic Fishery Management Council, M. (2010). "Surfclam and OCean Quahog Specifications for 2011, 2012, and 2013 Including: Draft Environmental Assessment, Regulatory Review, and Initial Regulatory Flexibility Analysis." Mid-Atlantic Fishery Management Council Quota Setting Document. Mid-Atlantic Fishery Management Council.

Moloney, D. G. and P. H. Pearse (1979). Quantitative rights as an instrument for regulating commercial fisheries. *Journal of the Fisheries Board of Canada 36*(7), 859–866.

Northeast Fisheries Science Center, N. (2010). 49th Northeast Regional Stock Assessment Workshop 49th Northeast Regional Stock Assessment Workshop (49th SAW). Technical Report February.

Osiewalski, J. and M. F. Steel (1998). Numerical tools for the bayesian analysis of stochastic frontier models. *Journal of Productivity Analysis 10*(1), 103–117.

Pascoe, S. and L. Coglan (2002). The contribution of unmeasurable inputs to fisheries production: an analysis of technical efficiency of fishing vessels in the English Channel. *American Journal of Agricultural Economics 84*(3), 585–597.

Pella, J. J. and P. K. Tomlinson (1969). *A generalized stock production model*. Inter-American Tropical Tuna Commission.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team (2013). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-111.

Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*, pp. 20–22.

Polacheck, T., R. Hilborn, and A. E. Punt (1993). Fitting surplus production models: comparing methods and measuring uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences 50*(12), 2597–2607.

Punt, A. (1992). Selecting management methodologies for marine resources, with an illustration for southern african hake. *South African Journal of Marine Science 12*(1), 943–958.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schaefer, M. B. (1954). *Some aspects of the dynamics of populations important to the management of the commercial marine fisheries*. Inter-American Tropical Tuna Commission.

Schmidt, P. and R. C. Sickles (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics 2*(4), 367–374.

Schnute, J. (1977). Improved estimates from the schaefer production model: theoretical considerations. *Journal of the Fisheries Board of Canada 34*(5), 583–603.

Spiegelhalter, D. J., A. Thomas, N. G. Best, W. Gilks, and D. Lunn (1996). Bugs: Bayesian inference using gibbs sampling. *Version 0.5,(version ii) http://www. mrc-bsu. cam. ac. uk/bugs 19*.

Squires, D. and J. Kirkley (1999). Skipper skill and panel data in fishing industries. *Canadian Journal of Fisheries and Aquatic Sciences 56*(11), 2011–2018.

Su, Y.-S. and M. Yajima (2013). *R2jags: A Package for Running jags from R*. R package version 0.03-11.

Tingley, D., S. Pascoe, and L. Coglan (2005). Factors affecting technical efficiency in fisheries: stochastic production frontier versus data envelopment analysis approaches. *Fisheries Research 73*(3), 363–376.

Tsionas, E. G. (2002). Stochastic frontier models with random coefficients. *Journal of Applied Econometrics 17*(2), 127–147.

Tsionas, E. G. (2005). An introduction to efficiency measurement using bayesian stochastic frontier models. *Global Business and Economics Review 3*(2), 287–311.

Uhler, R. S. (1980, August). Least Squares Regression Estimates of the Schaefer Production Model: Some Monte Carlo Simulation Results. *Canadian Journal of Fisheries and Aquatic Sciences 37*(8), 1284–1294.

U.S. Department of Commerce, M.-S. F. (2007). Management reauthorization act of 2006. *US Public Law 109 479*.

Van den Broeck, J., G. Koop, J. Osiewalski, and M. F. Steel (1994). Stochastic frontier models: A bayesian perspective. *Journal of Econometrics 61*(2), 273–303.

Walden, J. B., J. E. Kirkley, R. Färe, and P. Logan (2012). Productivity change under an individual transferable quota management system. *American Journal of Agricultural Economics 94*(4), 913–928.

Walters, C. J. and R. Hilborn (1976). Adaptive control of fishing systems. *Journal of the Fisheries Board of Canada 33*(1), 145–159.

Wang, N., X. Lin, R. G. Gutierrez, and R. J. Carroll (1998, March). Bias Analysis and SIMEX Approach in Generalized Linear Mixed Measurement Error Models. *Journal of the American Statistical Association 93*(441), 249–261.

Weinberg, J. R. (2005). Bathymetric shift in the distribution of atlantic surfclams: response to warmer ocean temperature. *ICES Journal of Marine Science: Journal du Conseil 62*(7), 1444–1453.

Weitzman, M. L. (2002). Landing fees vs harvest quotas with uncertain fish stocks. *Journal of Environmental Economics and Management 43*(2), 325–338.

Weninger, Q. (1998). Assessing efficiency gains from individual transferable quotas: an application to the mid-atlantic surf clam and ocean quahog fishery. *American Journal of Agricultural Economics 80*(4), 750–764.

Weninger, Q. and R. E. Just (1997). An analysis of transition from limited entry to transferable quota: non-marshallian principles for fisheries management. Technical report.

Western, B. and S. Jackman (1994, June). Bayesian Inference for Comparative Research. *The American Political Science Review 88*(2), 412.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

Wilen, J. E. (1985). Towards a theory of the regulated fishery. *Marine Resource Economics 1*(4), 369–388.

Zhang, J. and M. D. Smith (2011, May). Estimation of a Generalized Fishery Model: A Two-Stage Approach. *Review of Economics and Statistics 93*(2), 690–699.