The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Correcting for Measurement Error in a Stochastic Frontier Model: A Fishery Context

Christopher Burns

University of Massachusetts-Amherst

cbburns@acad.umass.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Associations 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4-6, 2013.

Copyright 2013 by Christopher Burns. All rights reserved.Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Correcting for Measurement Error in a Stochastic Frontier Model: A Fishery Context

Chris Burns

Abstract

Using data from the Mid-Atlantic surfclam fishery, this study examines the effect of measurement error on the analysis of technical efficiency. After specifying a stochastic frontier model and estimating it under naive analysis, a measurement error correction technique known as Simulation Extrapolation (SIMEX) is used to obtain less biased estimates of technical efficiency and production parameters. The SIMEX estimates of the stochastic frontier model agree with economic theory and show that important relationships between technical efficiency and vessel characteristics are present, something the naive estimates do not. Both sets of estimates show regional variation in technical efficiency, possibly due to declining landings per-unit-effort, suggesting future fishery management should take this into account.

Key words: commercial fisheries, stochastic frontier, technical efficiency, measurement error

Introduction

Technical efficiency is measured as the ability of a firm to produce the maximum output given a level of inputs (Kumbhakar and Lovell 2000[12]). In natural resource-based industries such as fishing, the measurement of technical efficiency is important to policy decisions regarding fishery management and preventing overfishing. Often times empirical researchers performing this type of analysis use data that contain unknown amounts of measurement error. This study uses a unique data set with approximately known measurement error, and applies a bias reduction method known as simulation extrapolation (SIMEX) to assess the impact of measurement error on estimation of technical efficiency.

Measurement error or error-in-variables for regression models is a vexing problem for empirical researchers. When measurement error is present in a covariate 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.* 2006[6]).

The data set is a panel of logbook data from the Mid-Atlantic surfclam fishery. It describes input and output relationships for eighty-eight separate vessels that harvested surfclams during the years 2001-2009. Using a stochastic frontier approach, both naive and SIMEX estimates are obtained, and then technical efficiency estimates are derived using methods described in Kumbhakar and Lovell (2000[12]). Technical efficiency is measured as a ratio of realized output to the potential output. It is assumed that a firm's potential output is obtained by following the best practice methods, given the technology. A firm that operates far from the production frontier, given its level of inputs, will risk going out of business in the long run.

One motivation for measuring technical efficiency in this fishery is because it is regulated by Individual Transferable Quotas (ITQs), and vessels that are less efficient would be predicted to exit the fishery under the assumptions of a perfectly competitive market. Another area of interest is what factors are important to explaining technical efficiency. In the fisheries literature there has been an interest in assessing technical efficiency to establish what factors make vessels more efficient (Kirkley *et al.* 1998[11]), and whether changes in management can improve efficiency (Brandt 2007[3]). Because measurement error can lead to loss of statistical power and can mask important relationships in the data, an important question is whether there are significant differences in the naive and SIMEX estimates. Factors such as vessel age should be important in explaining technical efficiency, according to economic theory. Additionally, there is evidence of declining landings per-unit-effort (LPUE) in the southern areas of the fishery. This may suggest regional variations in technical efficiency should be observed in the data. Whether the different estimation methods show different amounts of regional variation is of interest to future fishery management decisions.

After some preliminary data analysis, the functional form for the stochastic frontier model (Aigner *et al.* 1977[1], Meeusen and van den Broeck 1977[13]) is specified as 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 one of the covariates, $log(biomass)$ is measured with additive error.

Data on the biomass come from the Northeast Fisheries Science Center (NEFSC) biomass survey (NEFSC 2009[15]). These data contain an estimate of the biomass in each year and also a measure of the sampling variability. This estimate of sampling variability is used to reduce the measurement error bias with a Monte Carlo method called simulation extrapolation (SIMEX) (Cook and Stefanski 1994[7]). This method has been shown to provide approximately consistent parameter estimates under a variety of measurement error models. SIMEX estimates of the standard errors are obtained using the sandwich estimator (Carroll *et al.* 2006[6]). After obtaining both the naive and SIMEX estimated production parameters, standard errors, returns to scale, and technical efficiency measures, the two sets of estimates are compared. Not only are substantial biases in the parameter estimates found, but significant differences in factors that explain technical efficiency are found between the two sets of estimates.

The remainder of the paper proceeds as follows: In Section 2, the logbook and biomass data are described in more detail, and the measurement error problem is motivated. In Section 3, the theoretical framework for the stochastic frontier model, measurement error model, and SIMEX are presented. In Section 4, results from the naive and measurement error corrected models are provided. Section 5 contains a summary of the paper and presents topics for future research.

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, see Figure 2. Atlantic surfclams are a fast-growing bivalve mollusk distributed along the coast of North America. In 1990 the fishery transitioned from limited entry to individual transferable quotas 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 logbooks that describe each fishing trip. The 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)[3], 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, making the fishery one of the most valuable single-species fisheries in the US(NEFSC 2009[15]).

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 is 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 88 different vessels observed over the nine year period.

To simplify the correlation structure within each vessel and because biomass is observed only once in a year, 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 3. 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 was more appropriate. After dropping observations for vessels that harvested only a few times, the resulting data are reduced 70 vessels, with 285 vessel-year observations. Summary statistics for the data can be found in Table 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 1: Summary statistics 2001-2009

In order to estimate the stochastic frontier, additional survey data collected by the National Oceanographic and Atmospheric Association's (NOAA) Northeast Fisheries Science Center (NEFSC)-Resource Evaluation and Assessment Division is used. This survey data has an estimate of the biomass (NEFSC 2009)[15] in each year. The survey data are gathered by stratified random sampling and plugged into a biological model, known as the KLAMZ model, in order to obtain an estimate of the biomass for the entire fishery. Estimates of the sampling variability of the biomass are obtained using both the delta method and bootstrapping. This sampling variability is crucial to the measurement error correction model, described in section 3. A boxplot of the bootstrapped biomass estimates from the KLAMZ model for years 1980-2008 are shown in table 3. The sampling variability for the biomass changes by year, and this is later addressed in the measurement error model. Because the biomass is measured using stratified random sampling, it is assumed that the sampling variability is independent from year to year.

Methodology

Observed data are available for the following variables:

- 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_{i3} = log_e(length of vessel i), and
- w_{t4} = log_e(biomass in year t).

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.

Note that x_{i3}, $log(length)$, is constant over t. w_{t4} is observed $log(biomass)$, which is measured with error, but constant over each vessel i. The true $log(biomass)$, x_{t4}, will be used to specify the measurement error model later in this section.

Before modeling the stochastic frontier, an assumption must be made for a functional form of production. The Cobb-Douglas model 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 % change in output due to a 1 % increase in input k. It also imposes constant elasticity of substitution on inputs, which may be an unrealistic assumption in certain industries. With these limitations in mind we proceed with estimation of the model. In the next section a random effects model for production, which is a type of linear mixed model, is motivated. Following Kumbhakar and Lovell(2000)[12], the the random effects model can be transformed into a stochastic frontier model.

Linear Mixed Model of Production

Let b_i be a vessel-level effect, and let e_{it} be within vessel errors. The vessel-level effect can be specified as a fixed or random effect, depending on the assumptions of the model (Kumbhakar and Lovell 2000[12]). A Hausman test for random effects fails-to-reject that the random effects model is inconsistent, with a p-value of 0.41 . A linear mixed model for log(total bushels harvested by vessel i in year t) that includes a one-way random effect is

$$y_{it}|b_i, e_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{i3} + \beta_4 w_{t4} + b_i + e_{it} \tag{1}$$

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 assumption is not crucial in this specification and can be relaxed.

Next, notation for a version of the model for vessel i is defined, and last a model for the whole fleet. Let

$$\mathbf{y_i} = (y_{it_1}, \ldots, y_{it_{n_i}})^T, \mathbf{X_i} = \begin{pmatrix} 1 & x_{it_1 1} & x_{it_1 2} & x_{i_1 3} & w_{t_1 4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{it_{n_i} 1} & x_{it_{n_i} 2} & x_{i_{n_i} 3} & w_{t_{n_i} 4} \end{pmatrix}, \tag{2}$$

$$\beta^T = (\beta_0, \ldots, \beta_4), \mathbf{e_i}^T = (e_{it_1}, \ldots, e_{it_{n_i}}), \text{ and } \mathbf{1}_{n_i} = \text{ a vector of length } n_i \text{ of all 1s.} \tag{3}$$

Using that notation, a vessel level model is

$$\mathbf{y}_i|b_i, \mathbf{e}_i = \mathbf{X_i}\beta + \mathbf{1}_{n_i} b_i + \mathbf{e_i} \tag{4}$$

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

$$\mathbf{y} = \begin{pmatrix} \mathbf{y_1} \\ \vdots \\ \mathbf{y_{70}} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} \mathbf{X_1} \\ \vdots \\ \mathbf{X_{70}} \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} \mathbf{1_{n_1}} & \mathbf{0_{n_1}} & \dots & \dots & \mathbf{0_{n_1}} \\ \mathbf{0_{n_2}} & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots \mathbf{0_{n_{69}}} & \\ \mathbf{0_{n_{70}}} & \dots & \dots & \mathbf{0_{n_{70}}} & \mathbf{1_{n_{70}}} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_{70} \end{pmatrix}, \text{ and } \mathbf{e} = \begin{pmatrix} \mathbf{e_1} \\ \vdots \\ \mathbf{e_{70}} \end{pmatrix}, \tag{5}$$

Using matrix notation a linear mixed model for the fleet is

$$\mathbf{y}|\mathbf{b}, \mathbf{e} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e} \tag{6}$$

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

Stochastic Production Frontier

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 unobserved, such as captain's fishing knowledge or ability. In theory, these unobserved random effects will determine how close the vessel operates to its potential output. The productivity literature (Kumbhakar and Lovell 2000[12]) provide a method for transforming the random effects in order to calculate technical efficiency. Before calculating technical efficiency the b_i's which are the empirical Best Linear Unbiased Predictors (eBLUPs), need to be normalized. Let

$$\hat{b}_{i^*} = max_j[\hat{b}_j] - \hat{b}_i \tag{7}$$

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 efficiency term. The productivity literature typically uses the half-normal, truncated normal and exponential. Following Kirkley, Squires and Strand(1995)[10] the technical efficiency term b_{i^*} is assumed half-normal, $|N(0,\sigma_b^2)|$. The stochastic production frontier model is then specified as

$$y_{it}|b_{i^*}, e_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{i3} + \beta_4 w_{t4} - b_{i^*} + e_{it} \tag{8}$$

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]

Calculating Technical Efficiency

Following Jundrow *et al.*(1982)[9], 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

$$TI_i = \frac{\sigma_b \sigma_e}{\sigma}\left[\frac{\frac{\phi(\epsilon_i \lambda)}{\sigma}}{1 - \Phi(\frac{\epsilon_i \lambda}{\sigma})} - (\frac{\epsilon_i \lambda}{\sigma})\right] \tag{9}$$

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

$$TE_i = exp(-TI_i) \tag{10}$$

After calculating technical efficiency for each vessel it will be possible to rank these vessels from least efficient to most efficient. It will also be possible to compare the estimates of technical efficiency under both the naive and SIMEX corrected model, to see if there is a significant difference in vessel rankings or the distribution of TE_i.

Measurement Error Model

The biomass of the surfclam fishery is estimated through a stratified random sampling method, therefore it is assumed to be independent from year to year. For the purposes of the model, vessels

[1] Estimation of the model is performed using the linear mixed effects models 'nlme' package[8] in R(2012)[16]. 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 2008 [14]).

are assumed to face a constant biomass in a time period t. The additive measurement error model for $log(biomass)$ is specified as

$$w_{t4} = x_{t4} + v_t \tag{11}$$

where $E[w_{t4}|x_{t4}] = x_{t4}$. I specify the distribution for measurement error as $v_t \overset{\text{ind.}}{\sim} N(0, \sigma^2_{vt})$. Because the true $\sigma^2_{vt}$is not observed, it is estimated by $\hat{\sigma}^2_{vt}$.

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 model. Recall that the model for the fleet is

$$\mathbf{y}|\mathbf{b}, \mathbf{e} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e} \tag{12}$$

Let $\hat{\beta}$ be the MLE esimator for β. The naive estimator of this model with additive measurement error in $log(biomass)$ has the result

$$\text{plim}_{n \to +\infty} \hat{\beta} \neq \beta \tag{13}$$

The result is that the model with observed $log(biomass)$ will have inconsistent parameter estimates, as well as incorrect standard errors. The direction of bias for parameter estimates will depend on the correlation structure of the covariates in the model. A comprehensive explanation of measurement error in linear mixed models can be found in (Carroll *et al.* 2006[6]) or (Buonaccorsi, Demidenko and Toteson 2000[4]). It can be shown in the case of a linear mixed model, the inconsistency in the parameter estimates will also lead to inconsistent estimates of the random effects. The consequences for our model of interest will be that technical efficiency measures are biased. Wang, Lin, Gutierrez, *et al.* (1998)[17] show that SIMEX can be used to correct for measurement error in an generalized linear mixed measurement error model (GLMMeM), using a quadratic extrapolant. In the next section,a Monte Carlo method known as SIMEX is described, which will correct for the measurement error in the model.

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 (Carrol *et al.* 2006)[6]. 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)[5].

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: $\hat{\theta}_j(j, SIMEX) = g_j(-1)$.

The last step of SIMEX is the extrapolation step, and there are several fitting methods that can be chosen. A quadratic extrapolation function was used because it fit the SIMEX estimated values best. However, it should be noted that the choice of the extrapolation function can affect the SIMEX corrected estimates. SIMEX was also used to obtain the corrected eBLUPS, OLS errors, and standard errors via the sandwich estimator. 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. This method accounts for the variability in the Monte Carlo simulation as well as the variability in the estimator.(Carroll *et al.* 2006 [6]).

Results

This section presents naive and SIMEX estimates of the production parameters, standard errors and technical efficiency. Model estimates are compared and contrasted, and the significance of the results are discussed later in this section. Table 2 shows SIMEX estimates obtained from 250 simulations at each λ_m.

	Naive	SIMEX
Intercept	-6.634	-20.434
	(0.956)***	(1.910)***
log(timefish)	0.498	1.029
	(0.086)***	(0.095)***
log(fuel)	0.551	0.106
	(0.085)***	(0.089)
log(length)	-0.599	0.047
	(0.171)***	(0.167)
log(biomass)	1.598	3.344
	(0.095)***	(0.158)***
$\hat{\sigma}_b^2$	0.070	0.036
$\hat{\sigma}_e^2$	0.054	0.020
note: *p < 0.1, **p <0.05, ***p<0.01		

Table 2: Model Estimates

The two sets of parameter estimates show distinct differences in the coefficients of $log(biomass)$, $log(timefishing)$ and $log(length)$, which all are biased downwards. The coefficient on $log(timefishing)$ is closer to one in the SIMEX estimates, suggesting it has a larger impact on $log(bushels)$ than in the naive model. Conversely, the naive coefficients for the intercept and $log(gallons)$ are biased upwards. The coefficient on $log(length)$ changes signs, going from negative and significant in the naive model, to positive and not significant in the SIMEX result. The lack of significance on

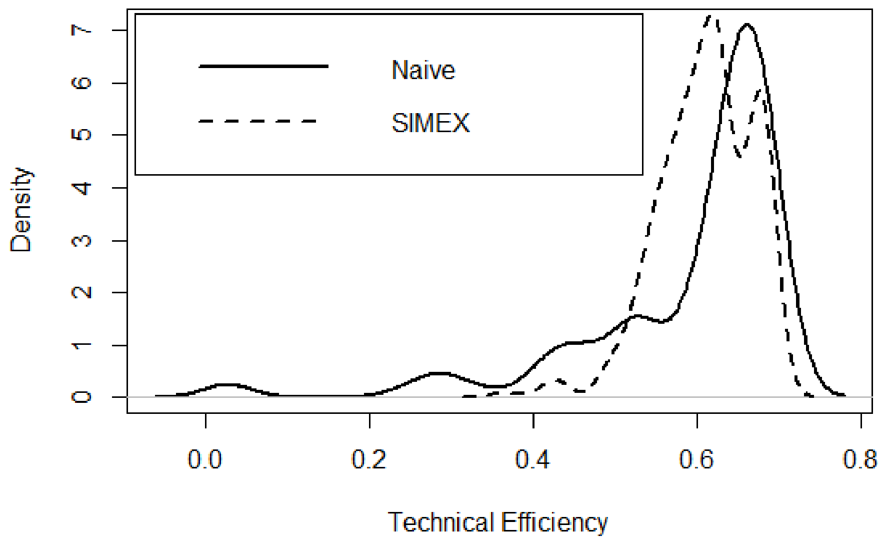

Figure 1: Technical Efficiency Density Plot

$log(length)$ may be due to low variability in the length of the vessels. The variable is left in the model because the Cobb-Douglas functional form includes a measure of capital, and length is typically used in the fisheries literature. Additionally, the positive coefficient on $log(length)$ is what would be expected from economic theory.

Table 1 above shows a density plot of the technical efficiency estimates. Mean technical efficiency under naive estimation is 59%, and 61% under SIMEX. 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, with the median for the naive estimates significantly greater than the median for the SIMEX estimates. This result suggests the naive model tends to overstate the mean technical efficiency of the surfclam industry.

Figures 2 and 3 show how measurement error can significantly mask important relationships in the data. In particular, the naive estimates do not show significant relationships between technical efficiency and age of the vessel, or hull material. The SIMEX estimates do show a significant relationship, both in age of the vessel and hull material. As theory would suggest, older vessels are less technically efficient. Similarly, vessels with fiberglass (FBG) hulls are significantly more technically efficient than boats with steel or wood hulls. This result would suggest that firms with

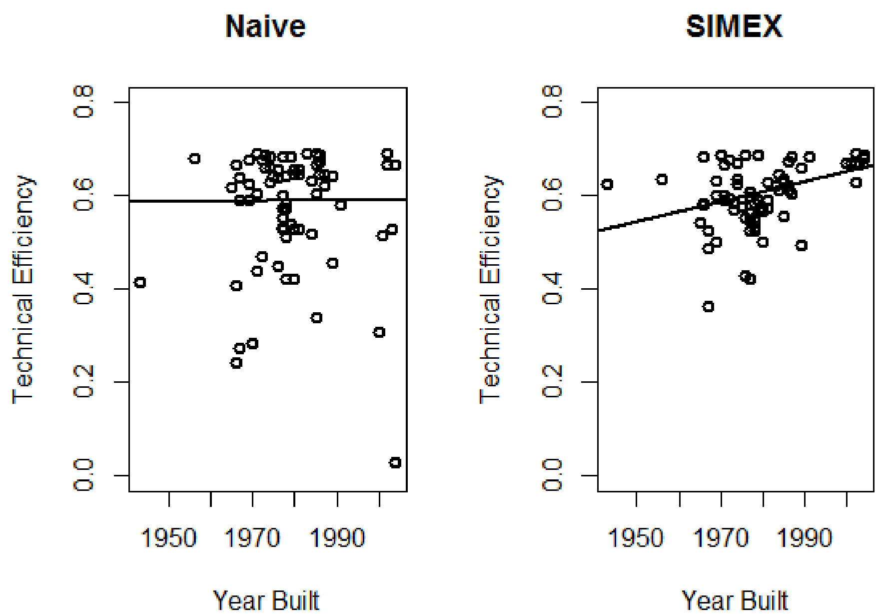

Figure 2: Technical Efficiency vs. Year Built

larger amounts of capital would have an advantage in the fishery, since they can purchase newer vessels, made of more advanced materials. Lastly, figure 4 shows that there are distinct differences in technical efficiency by region. Vessels whose home ports are in the southern region of

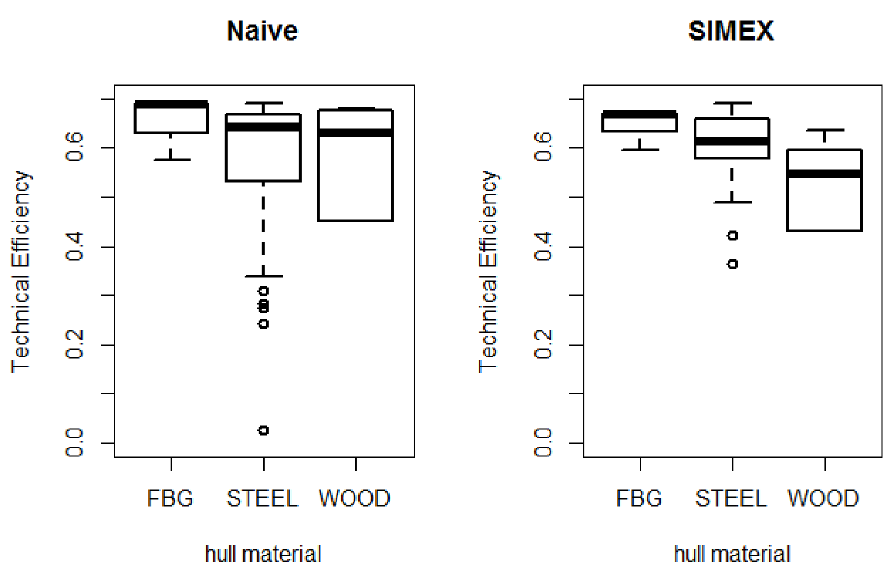

Figure 3: Technical Efficiency vs. Hull Material

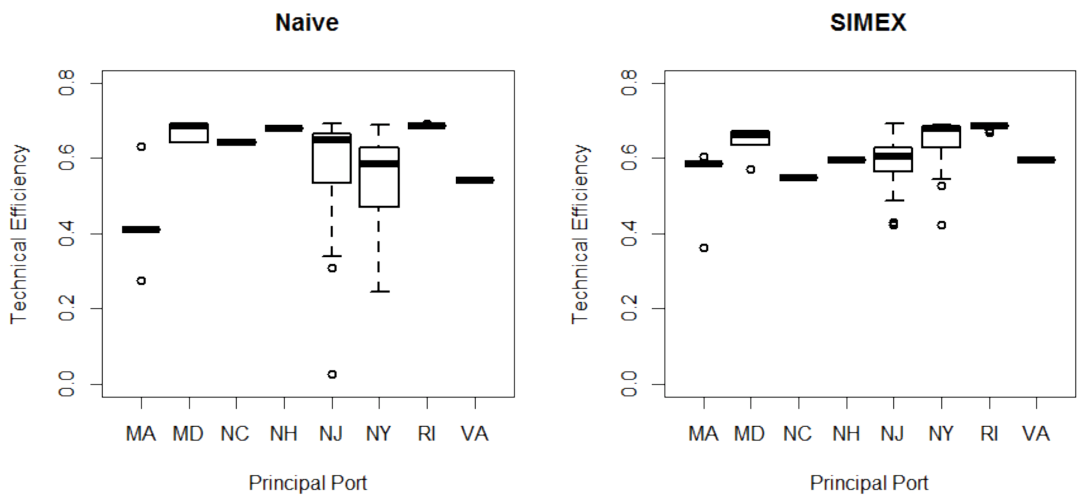

Figure 4: Technical Efficiency by Region

the fishery, such as North Carolina, have significantly lower technical efficiency than vessels in the northern part, such as New York or Rhode Island. This may be due to significant declines in LPUE in the southern part of the fishery.

Another quantity of interest is the linear combination of the coefficients on $log(timefishing)$, $log(gallons)$ and $log(length)$. Given the Cobb-Douglas functional form, this linear combination is a measure of the returns-to-scale for the surfclam industry. The value of this linear combination can tells us if the industry is operating under decreasing returns-to-scale, constant returns-to-scale, or increasing returns-to-scale. 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. Because profit maximizing behavior suggests firms should be operating in the decreasing returns-to-scale portion of the production function, the SIMEX estimates may suggest that firms operating in the fishery are quantity constrained. This would agree with anecdotal evidence of ITQ consolidation in the market, leading some vessels to not have enough quota to be profitable.

Discussion

This paper examined the consequences of measurement error in a stochastic production frontier model using logbook and biomass survey data from the Mid-Atlantic surfclam fishery. A stochastic frontier model was estimated using a Cobb-Douglas functional form specified by economic theory. The measurement error was assumed to be an additive component for the variable $log(biomass)$. The results show that naive estimation of the model will lead to inconsistent estimates of the parameters, standard errors and technical efficiency. The measurement error problem was 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 in the naive model. They 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 age of the vessel, hull material and region of home port. In particular, older vessels are less technically efficient, as are boats made of steel or wood, as compared to fiberglass boats. Vessels whose home ports are located in the southern end of the fishery have lower technical efficiency on average, suggesting that fishery managers must take this into account in future management decisions. The important finding is that not accounting for the measurement error would mask these relationships, leading a researcher to incorrect conclusions about factors affecting technical efficiency. Future research will look at how the bias in the estimated parameters can change under different measurement error model specifications and how estimation would proceed in a Bayesian framework.

References

[1] Aigner, Dennis, C.A.Knox Lovell, and Peter Schmidt. (1977).Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, 6, 2137.

[2] ”Atlantic Surfclam - Status of Fishery Resources off the Northeastern US.” Atlantic Surfclam - Status of Fishery Resources off the Northeastern US. N.p., n.d. Web. 23 Apr. 2013.

[3] Brandt, Sylvia. (2007). Evaluating Tradable Property Rights for Natural Resources: The Role of Strategic Entry and Exit. *Journal of Economic Behavior & Organization* , 63, 158176.

[4] Buonaccorsi, John P., Demidenko, Eugene, and Tostenson, Tor. (2000). Estimation in Longitudinal Random Effects Models With Measurement Error. *Staistica Sinica* ,10, 885-903.

[5] Buonaccorsi, John P. (2010). *Measurement Error: Models, Methods, and Applications*. CRC Press.

[6] Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. CRC Press.

[7] Cook, J. R., and L. A. Stefanski. (1994).Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association*, 89, 13141328.

[8] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar and the R Development Core Team (2012). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-104.

[9] Jondrow, J., K. Lovell, I. Materov, and P. Schmidt. (1982) On the Estimation of Technical Efficiency in the Stochastic Frontier Production Model. *Journal of Econometrics*, 19, 233-28.

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

[11] Kirkley, J.E., Squires, D., and Strand, I.E. (1998). Characterizing managerial skill and technical efficiency in a fishery. *Journal of Productivity Analysis*, 9, 145-160.

[12] Kumbhakar, Subal C., and C. A. Knox Lovell. *Stochastic Frontier Analysis*. Cambridge University Press, 2000.

[13] Meeusen, Wim, and Julien van Den Broeck. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, 18, 435444.

[14] McCulloch, Charles E., Shayle R. Searle, and John M. Neuhaus. (2008). *Generalized, Linear, and Mixed Models*. 2nd ed. Wiley-Interscience

[15] NEFSC (2009). 49th Northeast Regional Stock Assessment Review Committee (SARC 49). Atlantic Surfclam Stock Assessment Report.

[16] R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[17] Wang, N., Lin, X., Gutierrez, R. G. & Carroll, R. J.(1998). Generalized Linear mixed Measurement Error Models. *Journal of the American Statistical Association*, 93, 249-261.

Year	SVA	DMV	NJ	LI	SNE	Other	All areas
1991			142	95	40	9	141
1992		199	124	119			126
1993		183	131	143	28	390	137
1994		232	111	132			121
1995		229	115		46		120
1996		184	108	85	37		112
1997		182	108	122			112
1998		217	116	114	30		115
1999		115	134	135	14		132
2000		142	135	137	36		134
2001		168	124	126	71		129
2002	74	106	120	118	108		116
2003		78	112	115	197		110
2004		61	94	94	306		98
2005		81	90	82	179		93
2006		61	94	93	112		89
2007		56	76	92	62		73
2008		52	67	73	101		65
Min	74	52	67	73	14	9	65
Max	74	232	142	143	306	390	141
Mean	74	138	111	110	91	199	112

Figure 1: Landings Per Unit Effort by Region
(NEFSC 2009)[15]

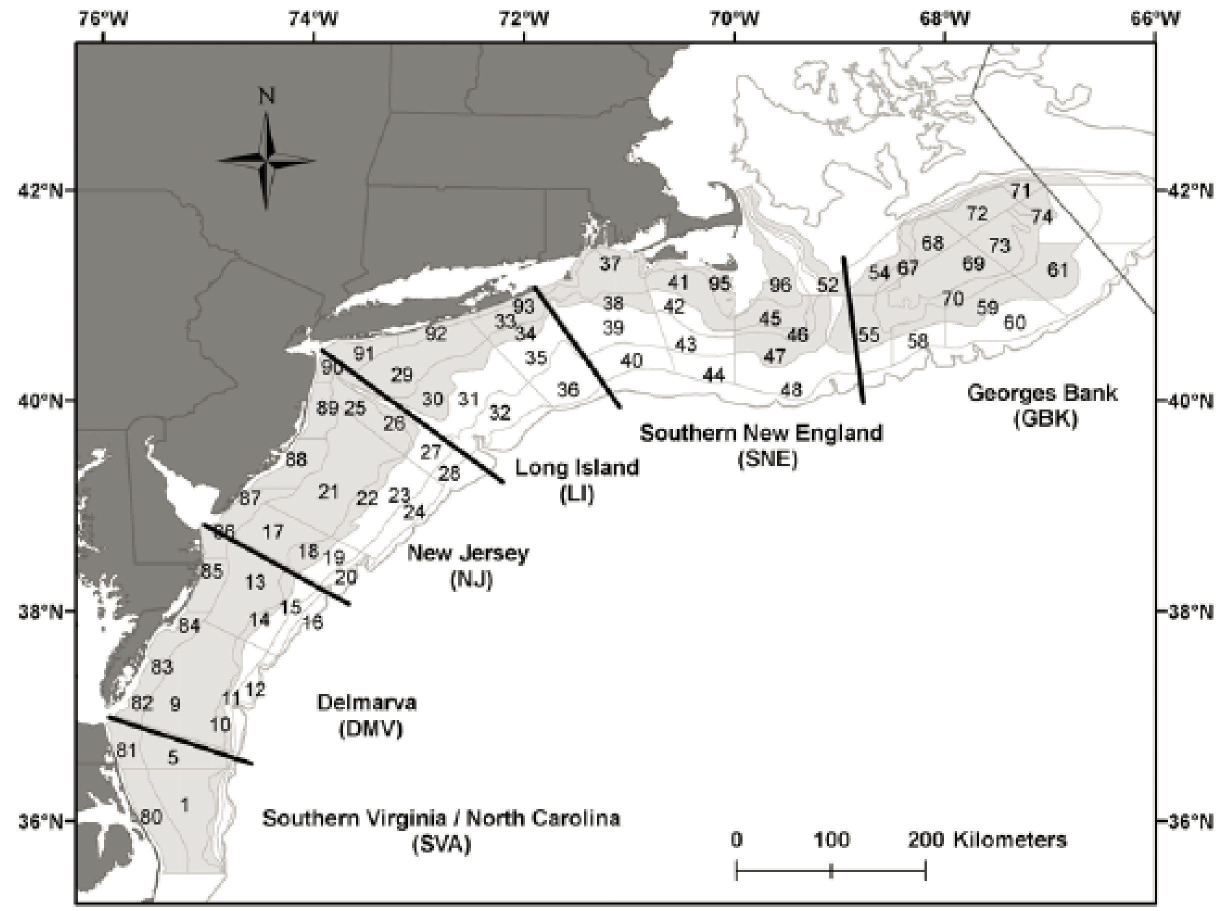

Figure 2: Mid-Atlantic Surfclam Fishery
(NEFSC 2009)[15]

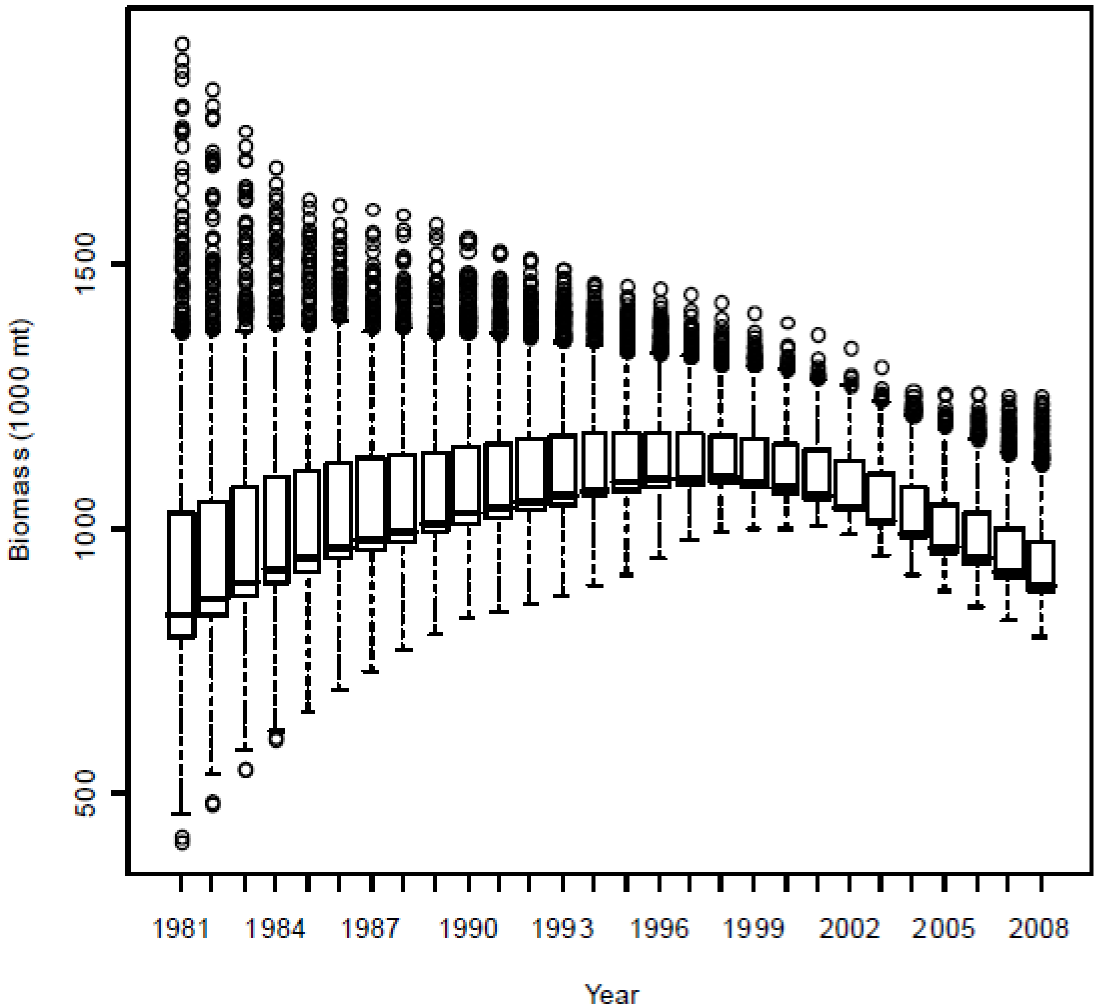

Figure 3: Boxplots with bootstrap biomass estimates for KLAMZ model (NEFSC 2009)[15]