ISSN: 1525-3066

Center for Policy Research
Working Paper No. 78

IDENTIFYING TECHNICALLY EFFICIENT FISHING VESSELS: A NON-EMPTY, MINIMAL SUBSET APPROACH

Alfonso Flores-Lagunes,
William C. Horrace, and Kurt E. Schnier

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

March 2006

$5.00

CENTER FOR POLICY RESEARCH – Spring 2006

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi Baltagi .... Economics
Pablo Beramendi .... Political Science
Dan Black .... Economics
Lloyd Blanchard .... Public Administration
William Duncombe .... Public Administration
Gary Engelhardt .... Economics
Deborah Freund .... Public Administration
Madonna Harrington Meyer .... Sociology
Christine Himes .... Sociology
William C. Horrace .... Economics
Bernard Jump .... Public Administration
Duke Kao .... Economics
Eric Kingson .... Social Work
Thomas Kniesner .... Economics
Tom Krebs .... Economics
Jeffrey Kubik .... Economics
Andrew London .... Sociology
Len Lopoo .... Public Administration
Jerry Miner .... Economics
Jan Ondrich .... Economics
John Palmer .... Public Administration
Lori Ploutz-Snyder .... Health and Physical Education
David Popp .... Public Administration
Grant Reeher .... Political Science
Stuart Rosenthal .... Economics
Ross Rubenstein .... Public Administration
Margaret Usdansky .... Sociology
Michael Wasylenko .... Economics
Janet Wilmoth .... Sociology

GRADUATE ASSOCIATES

Javier Baez .... Economics
Sonali Ballal .... Public Administration
Dana Balter .... Public Administration
Jason Bladen .... Economics
Jesse Bricker .... Economics
Maria Brown .... Social Science
Yong Chen .... Economics
Ginnie Cronin .... Social Science
Ana Dammert .... Economics
Mike Eriksen .... Economics
Katie Fitzpatrick .... Economics
Nadia Greenhalgh-Stanley .... Economics
Alexandre Genest .... Public Administration
Julie Anna Golebiewski .... Economics
Yue Hu .... Economics
Becky Lafrancois .... Economics
Joseph Marchand .... Economics
Neelakshi Medhi .... Social Science
Larry Miller .... Public Administration
Wendy Parker .... Sociology
Emily Pas .... Economics
Shawn Rohlin .... Economics
Cynthia Searcy .... Public Administration
Talisha Searcy .... Public Administration
Jeff Thompson .... Public Administration
Wen Wang .... Economics
Yoshikun Yamamoto .... Public Administration

STAFF

Kelly Bogart .... Administrative Secretary
Martha Bonney .... Publications/Events Coordinator
Karen Cimilluca .... Administrative Secretary
Kati Foley .... Administrative Assistant, LIS
Kitty Nasto .... Administrative Secretary
Candi Patterson .... Computer Consultant
Mary Santy .... Administrative Secretary
Tammy Tanner .... Librarian/Office Coordinator

Abstract

There is a growing resource economics literature, concerning the estimation of the technical efficiency of fishing vessels utilizing the stochastic frontier model. In these models, vessel output is regressed on a linear function of vessel inputs and a random composed error. Using parametric assumptions on the regression residual, estimates of vessel technical efficiency are calculated as the mean of a truncated normal distribution and are often reported in a rank statistic as a measure of a captain's skill and used to estimate excess capacity within fisheries. We demonstrate analytically that these measures are potentially flawed, and extend the results of Horrace (2005) to estimate captain skill for thirty nine vessels in the Northeast Atlantic herring fleet, based on homogenous and heterogeneous production functions within the fleet. When homogenous production is assumed, we find inferential inconsistencies between our methods and the methods of ranking the means of the technical inefficiency distributions for each vessel. When production is allowed to be heterogeneous, these inconsistencies are mitigated.

Introduction

There is a growing resource economics literature concerned with the estimation of the technical efficiency of fishing vessels utilizing the stochastic frontier model. Attention to the stochastic frontier model has often been motivated for two reasons. The first is the need to address the excess capacity prevalent in many of our fisheries, and the second centers on estimating the determinants of “skipper skill” in order to determine the characteristics of “highliners,” vessel captains who consistently out-fish others in the fleet. Excess capacity is conventionally estimated by comparing the current short-run production possessed by a vessel to that which could be achieved at the maximum utilization of their production inputs (Felthoven 2002; Kirkley et al. 2002, 2004). Estimation of “skipper skill” is achieved by regressing vessel technical efficiencies on captain demographic variables, and it is often concluded that “skipper skill” is positively correlated with a captain’s experience and education (Kirkley et al. 1998; Sharma and Leung 1998; Pascoe and Coglan 2002). Given the twofold importance of the stochastic frontier model in the fisheries economics literature, the inferential properties of technical efficiency measures need to be further investigated to assess their reliability in advising policy. This is the purpose of this research effort.

In this context the parametric stochastic frontier model (Meeusen and van den Broeck, 1977; and Aigner et al., 1977) yields estimates of the parameters (mean and variance) of a normal distribution for each vessel, which (when truncated at zero) represent the distribution of a vessel's inefficiency (see Aigner et al., 1977; and Battesse and Coelli, 1988). Of course the truncated distribution of a vessel's technical inefficiency is not a point estimate of technical inefficiency *per se*, so the stochastic frontier literature recommends the mean of the truncated normal distribution as a point estimate for inefficiency. All the aforementioned papers use this point estimate in assessing the inefficiency. Recently, Horrace (2005) argues that the mean of the truncated normal distribution is a poor estimate of technical inefficiency, since the mean is only one characterization of an otherwise complicated distribution. Horrace recommends calculating "the probability that a vessel is efficient or inefficient within the sample" using a multivariate truncated normal distribution, based on the estimated mean and variance of the underlying truncated normal distribution for each vessel. These probabilities can then be used to identify vessels that are efficient or inefficient with high or low probabilities. In particular the procedures can be used to identify a *single* vessel with a high probability of being efficient or inefficient. However, it can be shown that with reasonably large, homogeneous samples of fishing vessels, a *single* efficient or inefficient vessel is not forthcoming. In this case there is no inference on vessel technical efficiency.

The purpose of this paper is to extend the probabilistic results of Horrace (2005) to construct non-empty *subsets* of minimal cardinality with high probability of being efficient or inefficient. (This is a new construction, at least outside the Bayesian part of this literature.) That is, given a sample of fishing vessels and estimates of each vessel's probability of being efficient or inefficient, the goal is to identify the smallest, non-empty subsets of vessels that will contain the most and least efficient vessels with high probability (low error rate). This is markedly different from the results of Horrace (2005) which can only identify a *single* vessel as being efficient or inefficient with high probability, and which may often result in "no inference." Consequently, the inference results developed here are more practical, since the "non-empty" feature ensures that inference can always be performed. This notion of a non-empty subset is in keeping with the ranking and selection literature developed by Gupta (1965), and was first applied to the stochastic frontier model by Horrace and Schmidt (2000).

This paper also provides a practical application of the Horrace (2005) results to the North Atlantic herring fleet with an extension to the case of non-constant variance of the distributions of technical inefficiency. We demonstrate empirically that under a non-constant variance assumption, the usual point estimate of technical inefficiency (based strictly on the mean of its distribution) may be an inappropriate measure with regard to any particular vessel being efficient (inefficient) with high probability. It is shown that while a large (small) point estimate of the mean of technical inefficiency implies a small (large) probability that a vessel is efficient when the variance of the underling normal distribution is constant across vessels, this may not be the case when the variance is allowed to vary across vessels. In particular, unbalanced panels will produce erroneous inferences of relative efficiencies, because they necessarily imply heterogeneity in the variance estimates. The paper demonstrates this empirically by estimating a production function for an unbalanced panel of 39 vessels in the North Atlantic Herring fleet from 2000 to 2003. We show that the rankings of the usual point estimates of technical efficiency do not correspond to Horrace's "probability of being most efficient."

Another empirical contribution of the paper is that the production function is estimated two ways. Following Schnier et al. (2006), we estimate and perform inference on technical efficiency based on a homogenous production function across all 39 vessels and based on a heterogeneous production function across different classes of vessels using a latent class model. Inconsistencies between our methods and the usual method of ranking the mean of technical efficiency of each vessel exist under the homogenous production function assumption. In general, the technical inefficiency rank statistics do not correspond to the ranks of the probabilities of vessels being most or least efficient. The inconsistency across methodologies may imply that the homogenous production assumption is wrong (although a variety of other misspecifications may be driving this empirical result). Once the assumption is relaxed, and the production function is allowed to vary across classes of vessels, our methods and the ranking of the mean efficiencies produce more consistent results. This result underscores the importance of assessing heterogeneity in resource production functions when the mean of the inefficiency distribution is used as a point estimate for inefficiency.

The following section of the paper describes the stochastic frontier model estimated under the homogeneous and heterogeneous production function assumptions (see Schnier et al., 2006) as well as the rank probability methods outlined in Horrace (2005); it also deatils construction of the non-empty subsets. Section three describes the data and discusses the heterogeneous estimation methodology of El-Gamal and Grether (1995, 2000) used in our analysis. The empirical results are outlined in sections four and five for the homogenous and heterogeneous production functions, respectively. The final section summarizes our findings and proposes areas for additional research.

Parametric Frontier Models

Battese et al. (1989) and Battese and Coelli (1995) formulate a stochastic production function for unbalanced panel data as follows,

$$Y_{it} = f(X_{it};\beta)\exp\{\varepsilon_{it}\} \tag{1}$$

where Y_{it} is the dependent variable measured for vessel $i = 1,...,n$ in time period $t = 1,...,T_i$, X_{it} is a matrix of input variables for vessel i in time period t, β is a parameter vector, and ε_{it} is the error term. The error term is specified as follows,

$$\varepsilon_{it} = v_{it} - \eta_i \tag{2}$$

where v_{it} is an independently and identically distributed $\mathrm{N}(0,\sigma_V^2)$ random variable, and η_i is a one-sided, non-negative, vessel specific error term distributed as the truncation below zero of a $\mathrm{N}(\mu,\sigma_U^2)$ random variable, representing the technical inefficiency of vessel *i*. The likelihood function utilized is (Battese et al. 1989; Battese and Coelli 1995),

$$L(\theta;y) = -\frac{1}{2}\left(\sum_{i=1}^{n}T_i\right)\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(T_i-1)\log[(1-\gamma)\sigma_S^2] - \frac{1}{2}\sum_{i=1}^{n}\log\{\sigma_S^2[1+(T_i-1)\gamma]\}$$
$$-n\log[1-\Phi(-z)] + \sum_{i=1}^{n}\log[1-\Phi(-z_i^*)] - \frac{1}{2}nz^2 - \frac{1}{2}(y_i - x_i\beta)'(y_i - x_i\beta)[(1-\gamma)\sigma_S^2] + \frac{1}{2}\sum_{i=1}^{n}z_i^{*2}$$

where,

$$z = \frac{\mu}{(\sigma_S^2\gamma)^{0.5}}, \qquad z_i^* = \frac{\mu(1-\gamma) - T_i\gamma(\bar{y}_i - \bar{x}_i\beta)}{(\gamma(1-\gamma)\sigma_S^2[1+(T_i-1)\gamma])^{0.5}}, \qquad \bar{y}_i = \frac{y_i}{T_i}, \qquad \text{and } \bar{x}_i = \frac{x_i}{T_i}. \tag{3}$$

The parameters for estimation are β, $\gamma = \sigma_V^2/\sigma_S^2$, μ and $\sigma_S^2 = (\sigma_U^2 + \sigma_V^2)$.[1]

The conditional mean of the vessel specific error term, η_i, can be used to determine the level of technical efficiency possessed by each vessel. The traditional technical efficiency of each vessel is as follows (Battese et. al, 1989; Battese and Coelli 1992,1993),

[1] In the case of heterogeneous production each class within the latent class regressions possesses a separate parameter vector, β_h.

$$TE_i = E[\exp\{-\eta_i \mid \varepsilon_{it}\}] = \{\frac{1-\Phi\left[\sigma_i^* - \left(\frac{\mu_i^*}{\sigma_i^*}\right)\right]}{1-\Phi\left(\frac{-\mu_i^*}{\sigma_i^*}\right)} \exp\left\{-\mu_i^* + \frac{1}{2}\sigma_i^{*2}\right\} \tag{4}$$

where,

$$\mu_i^* = \frac{\mu\sigma_V^2 - \sigma_U^2 T_i \overline{E}_i}{\sigma_V^2 + T_i\sigma_U^2} \tag{5}$$

$$\overline{E}_i = \frac{1}{T_i}\sum_{t=1}^{T_i}\hat{\varepsilon}_{it} \tag{6}$$

$$\sigma_i^{*2} = \frac{\sigma_U^2\sigma_V^2}{\sigma_V^2 + T_i\sigma_U^2}. \tag{7}$$

The $\hat{\varepsilon}_{it}$ is the residual for the maximum likelihood estimate. Here equations (5) and (7) are the mean and variance of a normal distribution, which when truncated a zero, represents the conditional distribution of technical inefficiency of vessel *i*. That is, the truncated normal distribution of η_i conditional on ε_i is given by the truncation at zero of a N(μ_i^*, σ_i^{*2}). Equation (4) is the mean of a monotonic transformation of this truncated distribution. It has been argued in Horrace (2005) that equation (4) is a misleading measure of technical efficiency and recommends the empirical probabilities

$$F_k(\mu_i^*, \sigma_i^{*2}) = F_k = \Pr\{\eta_i \le \eta_k, \forall i \ne k\} \quad \text{and} \tag{8}$$

$$F_k^*(\mu_i^*, \sigma_i^{*2}) = F_k^* = \Pr\{\eta_i \ge \eta_k, \forall i \ne k\}. \tag{9}$$

(See Horrace, 2005, equations 4 and 5 for explicit formulae for these probabilities based on the normal-half-normal specification.) The probabilities, F_k and F_k^*, are readily calculable based on the estimates in equation (5) and (7), and correspond to "the probability that vessel *k* is most inefficient (least efficient)" and "the probability that vessel *k* is least inefficient (most efficient)," respectively. In particular, if F_k^* is relatively large, then one can say that vessel *k* is least inefficient (most efficient) with high probability. A shortcoming of this concept is that for large values of *n* and vessels of roughly the same efficiency, this probability, F_k^*, will necessarily be small. That is, it will typically be difficult for any one vessel to be least efficient with high probability. To see this, one need only recognize that the events in equations (8) and (9) are partitions of the probability space (see Horrace, 2005, p. 341). That is, the events do not overlap, and the sum of the F_k over *k* is equal to 1. For instance, if there are three vessels (of many) that are technically efficient (on the efficient frontier), then the probability that any one of these three vessels is least inefficient is necessarily less than 1/3, a fairly low probability, indeed.

Consider an alternative inferential approach developed here. The goal is not to identify a *single* vessel as least inefficient or most inefficient with high probability, but to identify a non-empty subset of vessels of minimum cardinality that contains the least and most inefficient vessel with high probability. This style of inference is more typical of the ranking and selection literature (e.g., Gupta, 1965) than is the inference of Horrace (2005). Let the ranked probabilities be:

$$F_{(n)} > F_{(n-1)} > ... > F_{(1)} \qquad \text{and} \qquad F^*_{\langle n\rangle} > F^*_{\langle n-1\rangle} > ... > F^*_{\langle 1\rangle}.$$

Notice the notational subtlety (k) is distinct from $\langle k\rangle$ in the rankings. Due to the multivariate nature of the probabilities, it should be noted that $F_k \neq 1 - F_k^*$, in general. Consequently, $\langle 1\rangle$ does not necessarily equal (n), and (1) does not necessarily equal $\langle n\rangle$. If there are enough significant digits in the calculation of μ_i^*, σ_i^{*2}, F_k, and F_k^*, ties in the rankings should not be a factor. However, any tie-breaking rule (including randomization) will not compromise the inference that follows. It should also be noted that these ranked probabilities may not correspond to the ranked technical efficiencies of equation (4). This feature of the calculations is made clear in the empirical analyses that follow, and it underscores the importance of not confusing the probabilities in equations 8 and 9 with technical efficiency scores. Define the subsets:

$\zeta_{\max} = \{(n),(n-1),...,(k)\}$ where (k) is the largest value of (j) such that $\sum_{i=(j)}^{(n)} F_i \geq 1-\gamma$,

and

$\zeta_{\min} = \{\langle n\rangle,\langle n-1\rangle,...,\langle k\rangle\}$ where $\langle k\rangle$ is the largest value of $\langle j\rangle$ such that $\sum_{i=\langle j\rangle}^{\langle n\rangle} F_i^* \geq 1-\alpha$,

and where $\gamma \in (0,0.5)$ and $\alpha \in (0,0.5)$ are error rates that are important in the sequel. The subset $\zeta_{\max} \subset \{1,...,n\}$ contain the indices associated with the largest probabilities, F_j, while the subset $\zeta_{\min} \subset \{1,...,n\}$ contain the indices associated with the largest probabilities F_j^*. Since the subsets $\zeta_{\max}$ and $\zeta_{\min}$ contain at least the index (n) and $\langle n\rangle$, respectively, inference on the minimal and maximal η_i (the least and most efficient vessel) is assured ($\zeta_{\max}$ and $\zeta_{\min}$ are non-empty) .

Let $\Psi_{\max}$ be the selection rule associated with the set $\zeta_{\max}$, and let $\Psi_{\min}$ be the selection rule associated with the set $\zeta_{\min}$. Let the ranked inefficiency terms be: $\eta_{[n]} > \eta_{[n-1]} > ... > \eta_{[1]}$, so that $[n]$ corresponds to the index of the most inefficient vessel in the sample and [1] corresponds to the index of the least inefficient vessel in the sample. Further, let correct selections CS and CS^* be defined as events: $CS = \{[n] \in \zeta_{\max}\}$ given $\Psi_{\max}$ and $CS^* = \{[1] \in \zeta_{\min}\}$ given $\Psi_{\min}$,respectively. That is, a correct selection occurs when $\zeta_{\max}$ and $\zeta_{\min}$ contain the indices associated with the most $[n]$ and least [1] inefficient vessels, respectively. Then the following result is true.

Result 1:

Given the selection rules $\Psi_{\max}$ *and* $\Psi_{\min}$, *the* $\zeta_{\max}$ *and* $\zeta_{\min}$ *are non-empty subsets of minimal cardinality such that:*

$\Pr\{CS\} = \Pr\{[n] \in \zeta_{\max}\} \geq 1-\gamma$ and $\Pr\{CS^*\} = \Pr\{[1] \in \zeta_{\min}\} \geq 1-\alpha$. □

The result follows directly from the way that the subsets are defined, but a formal proof is provided in the appendix. The result implies that, $\zeta_{\max}$ and $\zeta_{\min}$ contain the indices associated with the most and least inefficient vessels, respectively, with pre-specified confidence levels $1-\gamma$ and $1-\alpha$, respectively.[2] The selection rules $\Psi_{\max}$ and $\Psi_{\min}$ nest the selection rules in Horrace (2005). That is, the Horrace (2005) subsets are contained in $\zeta_{\max}$ and $\zeta_{\min}$. The inference hinges on independence of the technical inefficiency distributions. In principle, this independence assumption can be relaxed, however, calculation of the probabilities F_k and F_k^* becomes cumbersome even for relatively small values of *n*. It should also be noted that these results do not require homogeneity of the production function implied by equation (1). In fact, *any* specification of a conditional mean production function (homogenous or not) that produces estimates of μ_i^* and σ_i^* is acceptable for use with Result 1.

Notice that the *most inefficient* vessel in the sample is equivalent to the *least efficient* vessel in the sample, so it can be said that $\zeta_{\max}$ contains the *least* efficient vessel with high probability. It may seem counter-intuitive to associate the 'max' subscript of $\zeta_{\max}$ with the "*minimally* efficient vessel," but one only need remember that the minimal efficiency is equivalent to *maximal* inefficiency. In what follows we attempt to be consistent with the notation and couch all discussion on $\zeta_{\max}$ in terms of maximal *inefficiency* and all discussion on $\zeta_{\min}$ in terms of minimal *inefficiency*.

We should also mention that there is a Bayesian inference literature that has grown out of the stochastic frontier literature, as well. These techniques either directly or indirectly provide inference on relative ranks using Bayesian sampling techniques and are a viable alternative to the results presented here. For example, see Atkinson and Dorfman (2005), Fernandez et al. (2002), Tsionas (2002), Kim and Schmidt (2000), and Koop et al. (1997).

---

[2] It should be noted that the subsets are not necessarily disjoint. A referee pointed out that it would be useful to modify the selection rules so that the subsets are guaranteed to be disjoint; this is left for future research.

Data and Methodology

We exploit the selection rules in a new application to the North Atlantic herring fleet. To estimate the stochastic frontier model we use logbook data for the Northeast Atlantic herring fleet provided by the National Marine Fisheries Service (NMFS). Logbook data contains information on where a vessel fishes, what gear they utilize, the time and date of departure and return, their home port, the number of crew members on board and a list of vessel characteristics (length, gross-registered-tonnage, horsepower, and hold capacity). The data set utilized is identical to that used by Schnier et al. (2006) and it contains 2894 observations for 39 vessels over the years 2000 through 2003. Within the data set there are two degrees of spatial resolution: the centroid of the latitude and longitude coordinate of the statistical reporting area a vessel fishes and the macro-region within which this statistical reporting is contained. The first degree of spatial resolution is used to determine the distance that vessels travel to fish to calculate the hours that they fished. The second degree of spatial resolution is used to control for the unobserved stock density within the area fished and the seasonal migration patterns of herring. In particular we create dummy variables to indicate whether vessels fished inshore or offshore during particular seasons. Each observation in the logbook data set represents a single trip and is the most accurate data available on the Atlantic herring fleet.

The econometric specifications used to illustrate our inference methods are identical to those in Schnier et al. (2006). The dependent variable used within the regressions is the total catch, C_{it}, in metric tons of fish for each of the logbook data entries. The logbook data contains the latitude and longitude coordinates of the vessel's home port as well as the statistical reporting area fished. Using these coordinates Schnier et al. were able to estimate the distance traveled by each vessel. In addition, the temporal and spatial resolution of the logbook data allows them to determine the distance traveled and, hence, the hours each vessel spent steaming to the fishing locations (assuming an average vessel speed of

12 knots).[3] They were than able to back out an estimate of total hours spent fishing (gear deployed) on each trip for each vessel. They estimate the following heterogeneous production function,

$$\begin{aligned}\ln(C_{it}) = \beta_{0|h} + \beta_{1|h} * \ln(GRT_i) + \beta_{2|h} * \ln(HP_i) + \beta_{3|h} * \ln(Crew_{it}) + \beta_{4|h} * \ln(Hours_{it}) + \\ \beta_{5|h} * \ln(GRT_i) * \ln(Crew_{it}) + \beta_{6|h} * \ln(Crew_{it}) * \ln(Hours_{it}) + \beta_{7|h} * DumNoCrew_{it} \\ + \beta_{8|h} * DumSpWntInshore_{it} + \beta_{9|h} * DumSpWntOffshore_{it} \\ + \beta_{10|h} * DumSumFallOffshore_{it} + v_{it} - \eta_i\end{aligned} \tag{10}$$

where $h = 1,...,H$ indexes production tiers, *GRT* is the vessel's gross registered tonnage, *HP* is the vessel's horse power, *Crew* is the number of crew members utilized, *Hours* is the total hours that the gear was deployed during the trip, and $DumNoCrew_{it}$ is a dummy variable indicating whether or not the number of crew members on board the vessel *i* was observed in time period *t*. In the case that no crew members were observed, we substitute the mean number of crew members in the data set for the missing value. The variables, $DumSpWntInshore_{it}$, $DumSpWntOffshore_{it}$, and $DumSumFlOffshore_{it}$ are dummy variables indicating whether vessel *i* fished inshore during the Spring or Winter, offshore during the Spring or Winter or offshore during the Summer or Fall in time period *t* respectively. The dummy variables control for the respective inshore and offshore seasons and for the stock abundances present during these time periods.[4] Equation 10 is estimated assuming both homogeneous and heterogeneous production technologies. In the homogeneous case, this means that all vessels utilize the same short-run production functions (e.g., $\beta_{4|h} = \beta_4$ for all *h*). Under heterogeneous production, each tier, *h*, within the fleet possess a tier-specific production function captured by the variability of the coefficients over *h*.

To estimate the heterogeneous production function the El-Gamal and Grether estimation classification (EC) algorithm is used (El-Gamal and Grether 1995, 2000). In addition to Schnier et al.'s use of the EC algorithm, this method has recently been used by El-Gamal and Inanoglu (2005) to

[3] Average vessel speed was provided by contacts at Woods Hole Oceanographic Institute. For a more detailed description of the data utilized and the methodology used to obtain the hours fished see Schnier et al. (2006).
[4] It is acknowledged that the decision of "where to fish" is part of "skipper skill." Since we are controlling for this in the regression (inshore or offshore), the remaining technical inefficiency will not include this location component. However, inshore and offshore locations are very large areas, so the inefficiency will still capture the important decision of where to fish *within* these the inshore or offshore areas.

investigate heterogeneity in Turkish banking sector. The EC algorithm groups vessels into a pre-specified number of tiers, *H*, and estimates the parameter vector, $\theta_h = [\beta_{0|h},...,\beta_{11|h},\gamma,\mu,\sigma_s^2]$, for each tier assuming γ, μ, and σ_s^2 are held constant across all tiers.[5] To determine which vessels belong to each of the *H* tiers, each vessel's contribution to the likelihood function is the maximum of the joint likelihood of all their observations, given $\Theta = [\theta_1,...,\theta_H]$.[6] The econometric results for the homogeneous production technology are presented in Table 1. The heterogeneous production results are also in Table 1.[7] Schnier et al. (2006) proposed the use of the heterogeneous production functions to investigate whether or not "highliners" can be better represented by productivity measures and to illustrate that policy recommendations regarding excess capacity may be erroneous if they are based on the traditional homogeneous production function assumption. For a complete discussion of the results and the implied elasticities (given the interactions in the model), see Schnier et al. (2006).

Homogenous Production Function Results

The efficiency estimation results, assuming a homogeneous production technology, are contained in Table 2 for each vessel. The first column of the table contains the vessel number, *i*. The second column contains the variance of the technical efficiency distribution of vessel *i* before truncation at zero (equation 7). The third column contains the mean of the technical efficiency distribution for vessel *i* before truncation at zero (equation 5). The fourth column contains the technical efficiency estimate for vessel *i* (equation 4) based on the mean of the efficiency distribution after truncation; the vessels are ranked on

[5] Holding γ, μ and σ_S^2 constant across the tiers is a necessary assumption because the econometric model is "ill-posed" if these variables to vary across the tiers (El-Gamal and Inanoglu 2005).

[6] The joint likelihood is expressed as (Schnier et al. 2006):

$$\ln[L(Y_{it};X_{it} \mid \Theta;H] = \sum_{i=1}^{N} \arg\max_h \sum_{t=1}^{T_i} \ln(L(Y_{it};X_{it} \mid \theta_h)).$$

where $L(.)$ is the likelihood function for the stochastic frontier model outlined in equation 3. This likelihood function may possess many local maxima. Therefore, 500 random starting points were used to obtain the maximum value using the constrained maximum likelihood (CML) algorithm in GAUSS for each of iterations.

[7] Given that the number of segments, *H*, is predetermined, likelihood ratio tests, Bayesian information criterion (BIC) and Akiake information criterion (AIC) were used to determine the optimal number of segments to use within the estimation. For a more detailed description of the estimation procedure see Schnier et al. (2006).

this technical efficiency estimate. Notice that the results imply that vessels 3 and 33 are most efficient (TE equal 0.956 and 0.933, respectively), and vessels 14 and 38 are least efficient (TE equal 0.016 and 0.007, respectively). This is technical efficiency in an *absolute* sense, relative to some out-of-sample, absolute standard. These are the parametric stochastic frontier efficiency results that one typically sees. Notice that for the least efficient vessels the technical efficiencies are surprisingly small (this improves with the heterogeneous estimation, so this is, in part, why we prefer the heterogeneous model). Columns 5 and 6 of the table are the new results from Horrace (2005), with F_i equal to the probability that vessel *i* is most inefficient and F_i^* equal to the probability that vessel *i* is least inefficient.[8] These efficiency results are *relative* to other vessels in the sample (relative inefficiency as opposed to absolute inefficiency).[9]

The results in Table 2 are compelling. Even though vessels 30, 33, 34, and 35 possesses relatively high values of ranked TE (0.864, 0.933, 0.795, and 0.809, respectively), they are least inefficient with probability (approximately) zero, based on F_i^*. Why this contradictory result? The answer lies in the distribution of technical inefficiency for vessels 12 and 27, which are least inefficient with positive probability (0.3034 and 0.0279, respectively). Notice that prior to truncation the distributions of technical inefficiency for vessels 12 and 27 possesses relatively high variance (σ_{12}^* = 0.137 and σ_{27}^* = 0.400, respectively). Due to the high variances, we cannot reject the hypothesis that these two vessels are efficient, relative to vessels 30, 33, 34, and 35. For these results, the variance of the distribution after truncation, the variance of interest in these comparisons, is an increasing function of the variance before truncation.[10] Therefore, after truncation the distributions of vessels 12 and 27 also have high variance. This distinction underscores the fallacy of using the mean of the distribution of technical

[8] These probabilities are calculated with the "rectangle" algorithm in Horrace (2005), footnote 6, p.349.

[9] It is, indeed, the relative nature of these efficiency measures that provide the extra information to identify variance differences across vessels (heterogeneity) in the sequel.

[10] For the formula for the variance after truncation see Horrace (2005) equation 2. It is not clear whether this a general theoretical result or an artifact of the data. Investigation of this results is left for future research.

efficiency as an estimate of technical efficiency in the sample. The multivariate probability F_i^* is simply a more appropriate measure. It also underscores the importance of accounting for heterogeneity (different σ_i^*) when this is a possibility.[11]

Using Result 1, we conclude that at the 95% level: $\zeta_{\max}$ = {38} and $\zeta_{\min}$ = {3, 12, 27}. Therefore with probability 0.95 the most inefficient vessel is 38, and the least inefficient vessel is either vessel 3, vessel 12, or vessel 27. Notice, again, that vessel 33 is not even in the subset of the best (least inefficiency), while vessels 12 is. However, had we focused attention only on the TE_i, we would have come to the fallacious conclusion that vessel 33 is relatively efficient, while in a probabilistic sense it is not likely to be, when its distribution of technical efficiency is compared to those of other vessels in the sample (i.e., when F_i^* is used as a means of comparison). A valid question is whether or not the fallacious conclusion is due to the overly simplistic estimate, TE_i, or due to the fact that a homogenous production function has been assumed for all 39 vessels? Of course for this empirical exercise we cannot answer this definitively, but the empirical results that follow suggest that it may be due to the homogenous production specification.

Using likelihood ratio tests, Bayesian information criterion (BIC) and Akaike information criterion (AIC) in the context of the EC algorithm, Schnier et al. (2006) concluded that three different (heterogeneous) production functions fit the data better than a single homogenous function. Hence, we divide the sample of vessels into the three segments of Schnier et al., which correspond to the three different production functions based on the latent class regression model and estimated using the EC algorithm. Then, using the results on μ_i^* and σ_i^* in Table 2 (based on a homogenous production function), we recalculate relative probabilities F_i and F_i^* based on these subsets. As such, the technical efficiency results are relative to the sample segments proposed by Schnier et al.. These are contained in

[11] Heterogeneity is generally not identified in cross-section parametric stochastic frontier models, only in panel data models.

Tables 3 – 5 for each subset. These tables are read identically to Table 2, and the only columns that have changed are the relative probabilities, F_i and F_i^*. The importance of the variance is underscored in these tables, as well. For example, in Table 3, vessel 27 has the lowest value of μ_i^* = 0.273, but because of its relatively high variance prior to truncation, it is not the least inefficient vessel based on TE_i; vessel 10 is. However, we cannot reject the hypothesis vessel 17 might be least inefficient when we examine the probability, F_i^*. This is echoed in Table 5, the high variability of vessel 23 does not preclude it from being the most inefficient vessel based on the F_i, which imply the singleton, $\zeta_{\max}$ = {23}. The usual TE_i estimates do not capture these heterogeneity effects, and one would have erroneously concluded that vessel 22 is the most inefficient. Finally in all three tables, there are still very few efficient vessels, and the inefficient vessels have extremely low efficiency scores (e.g., TE_i = 0.007 for vessel 38 in Table 3). One could view these low scores as counterintuitive, but this was merely a thought experiment to see if forcing the homogenous results into the tiers implied by the heterogeneous results improves inference over the purely homogenous case. In general, it does not. However, things are improved when we moved to the heterogeneous production function, the empirical focus of the next exercise.

Heterogeneous Production Function Results

The heterogeneous production results of Schnier et al. (2006) are reproduced in the last three columns of Table 1. Again, the EC algorithm divides the sample of vessels into the three segments, which correspond to three different production functions, based on the latent class regression. We reproduce the latent class inefficiency results of Schnier et al. for each segment in Tables 6-8. In Table 6, vessels 5, 10, 14, 27, and 39 are the relatively efficient (least inefficient) with probability at least 0.95. Now many vessels are operating on the frontier and the least efficient vessel (vessel 38) is now 29.8% efficient (versus 0.7% efficient in the homogenous results). These results underscore the importance of accounting for heterogeneity in the production functions of fishing vessels. Oddly, vessel 14 is now in contention for high efficiency (least inefficiency) in Table 6, even though it was close to the bottom in terms of its probability of being most inefficient in the homogenous case (Tables 2 and 3). The main difference is that with the assumption of a homogenous production function in Table 2 and 3, the mean of the distribution of technical inefficiency for vessel 8 (prior to truncation) is large (4.188), while under a heterogeneous production function assumption it is small (-0.077). This makes sense, since differences in the specification of the conditional mean production function are more likely to manifest themselves as differences in the conditional mean of technical efficiency than as differences in the conditional variance. Notice that, when the production function (conditional mean) changes, the conditional variance of vessel 8 doesn't change all that much: 0.069 (in Tables 2 and 3) versus 0.064 (In Table 6).

Generally speaking, the TE_i results in Tables 6-8 correspond to the inferences implied by the $\zeta_{\min}$ and $\zeta_{\max}$ results (with a few exceptions). In particular, the rank ordering at the top and bottom of the tables correspond, whether we focus on TE_i or F_i^*. For instance, in Table 6 and 7, the subsets of least inefficiency, $\zeta_{\min}$ = { 5, 10, 14, 27, 39} and $\zeta_{\min}$ = {7, 8, 12, 19} (respectively), correspond to vessels with high values of technical efficiency, TE_i, in proper rank order. Interestingly, in neither case was the vessel with the *highest* value for technical efficiency included in the subset (i.e., vessel 2 with TE_i = 0.928 in Table 6 and vessel 21 with TE_i = 0.907 in Table 7 were not in the subsets of least inefficiency, $\zeta_{\min}$). These two vessels were "beaten" in terms of their μ_i^* being larger than those in the subsets, and (perhaps, more importantly) they both had relatively small variance prior to truncation than the vessels in the subsets. The implication being that they were "*more precisely, more inefficient*" than the vessels in the subsets. Had their variances been larger, we may not have been able to reject the hypothesis that they were least inefficient. Things are admittedly more consistent in the heterogeneous results of Tables 6 – 8 than in the homogenous results of Table 2 – 5, but the erroneous "rank conclusions" of the usual technical efficiency estimates still cannot be ignored.

According to Schnier et al. (2006), the vessels in tiers one and two used similar fishing technologies: mid-water trawling, where a large net is dragged behind the vessel to catch fish. Tier three, however, possessed a mix of two fishing techniques. Some of the vessels were mid-water trawlers, but most of them used purse seine gear, which is not dragged, but used to encircle pods of fish. Not surprising, the heterogeneous efficiency results of this tier are different than those of tiers one and two, where the fishing techniques are similar across vessels. The first column of Table 8 contains the vessel number and the type of net employed: P – purse-seine gear and T – Trawler. It seems that trawlers tend to be *probabilistically* most and least inefficient, when compared to the purse seine gear vessels, as they tend to be grouped at the top and bottom of the rank statistic in the table. Also, there is only a single vessel on the 95% frontier in Table 8. Why this is the case is unclear, but we speculate that the mixed fishing techniques may have something to do with it.

Conclusions

This paper extends the probability results of Horrace (2005) to produce non-empty subsets of minimal cardinality that contain the least and most inefficient vessels with predetermined probability. The new results are useful, because the results of Horrace (2005) may not always exist, while our inference will always exist. The application to fishing vessels was particularly relevant because it represents a case where the results imply that heterogeneous production functions are preferred to a homogenous production function. Under homogeneity the inference implied by the usual TE_i estimate, and that implied by the non-empty subsets, were not consistent. Under heterogeneity of production, correspondence of the results was markedly improved. That is, the vessels at the top and bottom of the ranking on the probability F_i^* tended to correspond to vessels at the extreme end of the ranking on the TE_i score.

One referee pointed out that an interesting extension would be to rank the vessels on F_i^* and F_i, trimming the vessels in the subsets $\zeta_{\min}$ and $\zeta_{\max}$ from the sample. The inference could then be re-performed on the remaining subset of vessels, producing second-best (least inefficient) and second-worst (most inefficient) subsets. The trimming procedure could be repeated until the original sample was exhausted. A similar experiment was performed in Horrace, Marchand, and Smeeding (2006) in the ranking of income inequality and poverty measures across countries. But as pointed out in that paper, the procedure does not correctly control for the overall error rate of the problem. Figuring out how one might adequately control for this error rate might prove useful in subsequent research.

Finally, the results extend beyond the model illustrated within this research. Global fisheries management has been plagued by the concern of excess capacity and overcapitalization as a result of the "race to fish." One policy solution to this phenomenon has been the support of vessel buybacks when fisheries policy is restructured. These buybacks are often targeted at removing the most inefficient vessels within a fleet to reduce the degree of excess capacity (Guyader et al. 2004). The methodologies illustrated in this paper will allow policy makers to more concretely estimate which vessels are the most inefficient within a fleet and facilitate the development of more efficient fisheries policy, which in turn should help alleviate the pressures on the resource.

Appendix

Proof of Result 1:

Suppose that $\zeta_{\max} = \{(n),(n-1),...,(j+1),(j)\}$ associated with the largest probabilities:

$$F_{(n)} > F_{(n-1)} > ... > F_{(j+1)} > F_{(j)}.$$

In particular it is supposed that the index $(j-1)$*, associated with* $F_{(j-1)} < F_{(j)}$ *is not an element of* $\zeta_{\max}$*. Then by the continuity of* η *and by definition of* $\zeta_{\max}$*, it is true that:*

$$F_{(n)} > F_{(n-1)} > ... > F_{(j+1)} < 1-\gamma.$$

Therefore it must be true that:

$$F_{(n)} > F_{(n-1)} > ... > F_{(j+1)} > F_{(j)} \geq 1-\gamma, \tag{A.1}$$

for if it were not true, then index $(j-1)$ *would have satisfied* $\Psi_{\max}$ *and would have been an element of* $\zeta_{\max}$*, which contradicts the initial supposition that it is not an element* $\zeta_{\max}$*. Then by equation* (7) *in* Horrace (2005)*,* $\Pr\{CS\} = \Pr\{[n] \in \zeta_{\max} \geq 1-\gamma$*. Furthermore,* $\zeta_{\max}$ *has minimum cardinality, since replacing any index* $(s) \in \zeta_{\max}$ *with any index* $(t) \notin \zeta_{\max}$ *will clearly violate* (A.1)*, while augmenting* $\zeta_{\max}$ *with any* $(t) \notin \zeta_{\max}$ *will satisfy* (10)*, but the cardinality of* $\zeta_{\max}$ *will have been increased by one. Similar arguments can be made to prove the results for* $\zeta_{\min}$. □

References

Aigner D.J., Lovell A.K. and P. Schmidt. 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics* 6:21-37.

Atkinson, S.E. and J.H. Dorfman. 2005. Multiple comparisons with the best: Bayesian precision measures of efficiency rankings. *Journal of Productivity Analysis*, 23, 359-382.

Battese G.E., Coelli T.J. and T.C. Colby. 1989. Estimation of Frontier Production Functions and the Efficiencies of Indian Farms Using Panel Data From ICRISAT's Village Level Studies. *Journal of Quantitative Economics* 5(2):327-348.

Battese G.E. and T.J. Coelli. 1988. Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. *Journal of Econometrics* 38:387-399.

Battese G.E. and T.J. Coelli. 1992. Frontier Production Functions, Technical Efficiency, ad Panel Data. *Journal of Productivity Analysis* 3:153-169.

Battese G.E. and T.J. Coelli. 1993. A Stochastic Frontier Production Incorporating a Model for Technical Inefficiency Effects. *Working Papers in Econometrics and Applied Statistics* No.69, Department of Econometrics, University of New England, Armidale.

Battese G.E. and T.J. Coelli. 1995. A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics* 20:325-332.

Bera, A.K. and S.C. Sharma. 1999. Estimating production uncertainty in stochastic frontier models, *Journal of Productivity Analysis*, 12, 187-210.

Bjorndal T. 1989. Production in a Schooling Fishery: The Case of the North Sea Herring Fishery. *Land Economics* 65:49-56.

El-Gamal M. and D. Grether. 1995. Are People Bayesian? Uncovering Behavior Strategies. *Journal of American Statistical Association* 90:1137-1145.

El-Gamal M. and D. Grether. 2000. Changing Decision Rules: Uncovering Behavioral Strategies Using Estimation Classification (EC). In *Preferences, Beliefs, and Attributes in Decision Making*, ed. M. Machina et al., New York:Kluwer.

El-Gamal, M. and H. Inanoglu. 2005. Inefficiency and Heterogeneity in Turkish Banking: 1990-2000. *Journal of Applied Econometrics* 20:641-664.

Felthoven R.G. 2002. Effects of the American Fisheries Act. *Marine Resource Economics* 17:181-206.

Fernandez, C., Koop, G., Steel, M.F.J., 2002. Multiple output production with undesirable outputs: an application to nitrogen surplus in agriculture. *Journal of the American Statistical Association* 97, 432–442.

Garcia del Hoyo J.J., Espino D.C. and R.J. Toribio. Determination of Technical Efficiency of Fisheries by Stochastic Frontier Models: A Case on the Gulf of Cadiz (Spain). *ICES Journal of Marine Science* 61:417-423.

Gupta, S.S., 1965, On some multiple decision (selection and ranking) rules, *Technometrics*, 7, 225-245.

Guyader, O., Daures, F. and S. Fifas. 2004. A Bioeconomic Analysis of the Impact of Decommissioning Programs: Application to a Limited-Entry French Scallop Fishery. *Marine Resource Economics* 19:225-242.

Horrace, W.C. and P. Schmidt. 1996. Confidence statements for efficiency estimates from stochastic frontier models, *Journal of Productivity Analysis*, 7, 257-82.

Horrace, W.C. and P. Schmidt. 2000. Multiple comparisons with the best, with economic applications, *Journal of Applied Econometrics*, 15, 1-26.

Horrace, W.C. 2005. On ranking and selection from independent truncated normal distributions, J*ournal of Econometrics*, 126, 335-354.

Horrace, W.C., J.T. Marchand, and T.M. Smeeding. 2006. Ranking Inequality: Applications of Multivariate Subset Selection, Unpublished Manuscript, Center for Policy Research, The Maxwell School of Syracuse University.

Kim, Y., Schmidt, P., 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, 91–118.

Kirkley J., Squires D. and I.E. Strand. 1998. Characterizing Managerial Skill and Technical Efficiency in a Fishery. *Journal of Productivity Analysis* 9:145-160.

Kirkley, J., Morrison Paul, C.J. and D. Squires. 2002. Capacity and Capacity Utilization in Common-Pool Resource Industries. *Environmental and Resource Economics*. 22:71-97.

Kirkley, J., Morrison Paul, C.J. and D. Squires. 2004. Deterministic and stochastic capacity estimation for fishery capacity reduction. *Marine Resource Economics* 19:271-294.

Koop, G., Osiewalski, J., Steel, M.F.J., 1997. Bayesian efficiency analysis through individual effects: hospital cost frontiers. *Journal of Econometrics*, 77–105.

Meeusen W. and J. van den Broeck. 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review* 18:435-444.

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.

Schnier K.E., C.M. Anderson, and W.C. Horrace. 2006. Estimating Heterogeneous Production in Fisheries. *Working Paper, University of Rhode Island – Department of Environmental and Natural Resource Economics.*

Sharma K.R. and P.S. Leung. 1998. Technical Efficiency of Hawaii's Longline Fishery. *Marine Resource Economics* 13(4):259-274.

Squires D. and J. Kirkley. 1999. Skipper Skill and Panel Data in Fishing Industries. *Can. J. Fish. Aquat. Sci.* 56:2011-2018.

Thorlindsson T. 1988. The Skipper Effect in the Icelandic Herring Fishery. *Human Organization* 47:199-213.

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

Table 1: Stochastic Frontier Results (Schnier et al., 2006)[a]

Variable	No Segments	Tier 1	Tier 2	Tier 3
Constant	-2.2913	-16.0335**	-10.1388**	1.8865*
	(-1.62)	(-11.03)	(-3.88)	(1.67)
GRT	0.2653*	-0.5028**	2.7700**	0.1658
	(1.71)	(-3.26)	(6.91)	(1.42)
HP	0.7781**	2.8326**	0.1254	0.3249**
	(3.57)	(10.85)	(0.38)	(2.16)
Crew	0.2679	0.4373	9.8811**	-0.4213
	(0.62)	(1.06)	(4.06)	(-1.05)
Hours	0.1130	0.4405**	-0.2826**	-0.3380**
	(1.20)	(2.99)	(-2.28)	(-2.24)
GRT*Crew	0.0184	-0.0146	-2.0462**	0.0845
	(0.17)	(-0.13)	(-4.36)	(1.10)
Crew*Hours	-0.0668	-0.1457	0.2823**	0.1414
	(-1.04)	(-1.32)	(3.08)	(1.49)
No-Crew	-0.1314**	0.0081	-0.1752**	-0.0605
	(-2.95)	(0.08)	(-2.86)	(-0.79)
Sp. Wint. Insh.	-0.3221**	-0.0486	-0.0183	-0.5759**
	(-6.93)	(-0.53)	(-0.22)	(-9.25)
Sp. Wint. Off.	0.3829**	0.6814**	0.3106**	-0.1518*
	(7.70)	(7.24)	(3.94)	(-1.66)
Sum. Fall Off.	0.2094**	-0.0314	0.3870**	-0.2274*
	(4.51)	(-0.31)	(6.88)	(-1.83)
γ	0.9597**	-------	-------	0.8807**
	(25.19)			(13.88)
σ^2_S	10.3535	-------	-------	3.2835*
	(1.06)			(1.88)
μ	-3.9023	-------	-------	-6.9310
	(-0.61)			(-1.34)
Number of Vessels	39	13	12	14
Mean Log-Likelihood	-1.00957	-------	-------	-0.96396

**indicates significant at the 95% level;*indicates significant at the 95% level; t-stats in parentheses.
[a]for a more detailed discussion of the implied elasticities and marginal products see Schnier et al. (2006).

Table 2. Homogeneous Vessel Efficiency Results Sorted on F_i^*

Vessel i	σ_i^*	μ_i^*	TE_i	F_i	F_i^*
3	**0.001**	**0.032**	**0.956**	**0**	**0.6375**
12	**0.137**	**-0.142**	**0.794**	**0**	**0.3043**
27	**0.400**	**0.273**	**0.586**	**0**	**0.0279**
8	0.204	0.278	0.648	0	0.0240
10	0.137	0.335	0.660	0	0.0049
6	0.103	0.316	0.684	0	0.0014
33	0.002	0.065	0.933	0	0
30	0.001	0.147	0.864	0	0
35	0.007	0.215	0.809	0	0
34	0.025	0.210	0.795	0	0
19	0.042	0.272	0.746	0	0
31	0.002	0.301	0.741	0	0
26	0.007	0.599	0.551	0	0
29	0.007	0.670	0.514	0	0
21	0.012	0.674	0.513	0	0
13	0.010	0.680	0.509	0	0
16	0.008	0.761	0.469	0	0
36	0.010	0.773	0.464	0	0
15	0.001	0.965	0.381	0	0
18	0.059	1.128	0.334	0	0
20	0.052	1.264	0.290	0	0
7	0.400	1.508	0.262	0	0
25	0.007	1.395	0.249	0	0
2	0.001	1.423	0.241	0	0
24	0.009	1.549	0.213	0	0
22	0.006	1.555	0.212	0	0
9	0.003	1.845	0.158	0	0
23	0.137	2.103	0.131	0	0
17	0.042	2.135	0.121	0	0
39	0.014	2.598	0.075	0	0
5	0.035	2.623	0.074	0	0
1	0.204	2.863	0.063	0	0
4	0.022	2.907	0.055	0	0
28	0.028	2.975	0.052	0	0
11	0.012	3.566	0.028	0	0
32	0.103	3.849	0.022	0	0
37	0.042	3.887	0.021	0	0
14	0.069	4.188	0.016	0	0
38	**0.042**	**4.928**	**0.007**	**1.0000**	**0**

TE_i is the Technical Efficiency of vessel *i*.

F_i is the probability that vessel *i* is least efficient.

F_i^* is the probability that vessel *i* is most efficient.

$\zeta_{\max}$ = {38} at 95% level.

$\zeta_{\min}$ = {3, 12, 27} at 95% level.

Table 3. Homogeneous Vessel Efficiency Results, Sorted on F_i^*

Vessel i	σ_i^*	μ_i^*	TE_i	F_i	F_i^*
10	**0.137**	**0.335**	**0.660**	**0**	**0.5837**
27	**0.400**	**0.273**	**0.586**	**0**	**0.4163**
2	0.001	1.423	0.241	0	0
1	0.204	2.863	0.063	0	0
17	0.042	2.135	0.121	0	0
39	0.014	2.598	0.075	0	0
5	0.035	2.623	0.074	0	0
4	0.022	2.907	0.055	0	0
28	0.028	2.975	0.052	0	0
11	0.012	3.566	0.028	0	0
37	0.042	3.887	0.021	0	0
14	0.069	4.188	0.016	0	0
38	**0.042**	**4.928**	**0.007**	**1**	**0**

F_i is the probability that vessel *i* is least efficient.

F_i^* is the probability that vessel *i* is most efficient.

$\zeta_{\max}$ = {38} at 95% level.

$\zeta_{\min}$ = {10, 27} at 95% level.

Table 4. Homogeneous Vessel Efficiency Results, Sorted on F_i^*

Vessel i	σ_i^*	μ_i^*	TE_i	F_i	F_i^*
12	**0.137**	**-0.142**	**0.794**	**0**	**0.8919**
8	**0.204**	**0.278**	**0.648**	**0**	**0.0831**
6	0.103	0.316	0.684	0	0.0184
19	0.042	0.272	0.746	0	0.0066
7	0.400	1.508	0.262	0	0.0001
21	0.012	0.674	0.513	0	0
13	0.010	0.680	0.509	0	0
18	0.059	1.128	0.334	0	0
15	0.001	0.965	0.381	0	0
24	0.009	1.549	0.213	0	0
9	0.003	1.845	0.158	0	0
32	**0.103**	**3.849**	**0.022**	**1.0000**	**0**

F_i is the probability that vessel *i* is least efficient.

F_i^* is the probability that vessel *i* is most efficient.

$\zeta_{\max}$ = {32} at 95% level.

$\zeta_{\min}$ = {8, 12} at 95% level.

Table 5. Homogeneous Vessel Efficiency Results, Sorted on F_i^*

Vessel i	σ_i^*	μ_i^*	TE_i	F_i	F_i^*
3	**0.001**	**0.032**	**0.956**	**0**	**1.0000**
34	0.025	0.210	0.795	0	0
33	0.002	0.065	0.933	0	0
30	0.001	0.147	0.864	0	0
35	0.007	0.215	0.809	0	0
23	**0.137**	**2.103**	**0.131**	**1.0000**	**0**
31	0.002	0.301	0.741	0	0
26	0.007	0.599	0.551	0	0
29	0.007	0.670	0.514	0	0
20	0.052	1.264	0.290	0	0
36	0.010	0.773	0.464	0	0
16	0.008	0.761	0.469	0	0
25	0.007	1.395	0.249	0	0
22	0.006	1.555	0.212	0	0

F_i is the probability that vessel *i* is least efficient.

F_i^* is the probability that vessel *i* is most efficient.

$\zeta_{\max}$ = {23} at 95% level.

$\zeta_{\min}$ = {3} at 95% level.

Table 6. Heterogeneous Vessel Efficiency Results, Sorted on F_i^*

Vessel i	σ_i^*	μ_i^*	TE_i	F_i	F_i^*
14	**0.064**	**-0.077**	**0.846**	**0**	**0.3015**
10	**0.125**	**-0.360**	**0.841**	**0**	**0.2756**
27	**0.345**	**-1.553**	**0.844**	**0**	**0.1455**
39	**0.013**	**0.030**	**0.905**	**0**	**0.1333**
5	**0.032**	**0.040**	**0.859**	**0**	**0.1017**
1	0.183	0.034	0.725	0	0.0414
37	0.039	0.126	0.817	0	0.0014
11	0.012	0.094	0.881	0	0
2	0.001	0.074	0.928	0	0
17	0.039	0.303	0.731	0	0
4	0.021	0.242	0.781	0	0
28	0.026	0.387	0.685	0	0
38	**0.039**	**1.230**	**0.298**	**1**	**0**

F_i is the probability that vessel *i* is least efficient.

F_i^* is the probability that vessel *i* is most efficient.

ζ_{max} = {38} at 95% level.

ζ_{min} = {5, 10, 14, 27, 39} at 95% level.

Table 7. Heterogeneous Vessel Efficiency Results, Sorted on F_i^*

Vessel i	σ_i^*	μ_i^*	TE_i	F_i	F_i^*
19	**0.039**	**-0.182**	**0.903**	**0**	**0.6969**
12	**0.125**	**-0.518**	**0.863**	**0**	**0.1885**
7	**0.345**	**-1.129**	**0.814**	**0**	**0.0550**
8	**0.183**	**-0.047**	**0.743**	**0**	**0.0287**
32	0.095	0.095	0.768	0	0.0173
6	0.095	0.126	0.758	0	0.0111
21	0.011	0.041	0.907	0	0.0035
18	0.055	0.332	0.705	0	0
13	0.009	0.135	0.864	0	0
15	0.001	0.692	0.501	0	0
24	0.009	0.967	0.382	0	0
9	**0.002**	**1.032**	**0.357**	**1.000**	**0**

F_i is the probability that vessel *i* is least efficient.

F_i^* is the probability that vessel *i* is most efficient.

ζ_{max} = {9} at 95% level.

ζ_{min} = {7, 8, 12, 19} at 95% level.

Table 8. Heterogeneous Vessel Efficiency Results, Sorted on F_i^*

Vessel *i/Type*	σ_i^*	μ_i^*	TE_i	F_i	F_i^*
3/P	**0.001**	**0.001**	**0.971**	**0**	**0.9630**
16T	0.008	0.011	0.930	0	0.0386
20/T	0.048	0.179	0.783	0	0
23/T	0.125	0.604	0.555	0	0
34/P	0.023	0.173	0.817	0	0
33/P	0.002	0.065	0.934	0	0
30/P	0.001	0.193	0.825	0	0
35/P	0.006	0.215	0.808	0	0
31/P	0.001	0.494	0.611	0	0
29/P	0.007	0.504	0.606	0	0
26/P	0.006	0.602	0.550	0	0
22/T	0.006	0.836	0.435	0	0
36/P	0.010	0.913	0.403	0	0
25/T	**0.006**	**1.230**	**0.293**	**1.000**	**0**

TE_i reported in Schnier et al. (2006).

F_i is the probability that vessel *i* is least efficient.

F_i^* is the probability that vessel *i* is most efficient.

$\zeta_{\max}$ = {25} at 95% level.

$\zeta_{\min}$ = {3} at 95% level.